Local genetic correlation gives insights into the shared genetic architecture of complex traits

Huwenbo Shi[1], Nicholas Mancuso[2], Sarah Spendlove[4], and Bogdan Pasaniuc[1,2,3]

[1]Bioinformatics Interdepartmental Program, University of California, Los Angeles, 90024
[2]Dept of Pathology and Laboratory Medicine,
[3]Dept of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, 90024
[4]Dept of Biology, Brigham Young University, 84602

Abstract

Although genetic correlations between complex traits provide valuable insights into epidemiological and etiological studies, a precise quantification of which genomic regions contribute to the genome-wide genetic correlation is currently lacking. Here, we introduce ρ-HESS, a technique to quantify the correlation between pairs of traits due to genetic variation at a small region in the genome. Our approach only requires GWAS summary data and makes no distributional assumption on the causal variant effects sizes while accounting for linkage disequilibrium (LD) and overlapping GWAS samples. We analyzed large-scale GWAS summary data across 35 complex traits, and identified 27 genomic regions that contribute significantly to the genetic correlation among these traits. Notably, we find 7 genomic regions that contribute to the genetic correlation of 12 pairs of traits that show negligible genome-wide correlation, further showcasing the power of local genetic correlation analyses. Finally, we leverage the distribution of local genetic correlations across the genome to assign putative direction of causality for 15 pairs of traits.

Introduction

Genomic regions that harbor variants contributing to multiple traits provide valuable insights into the underlying biological mechanisms with which genetic variation impacts complex traits[1,2,3,4,5,6,7]. Therefore, both the identification of new such regions as well as quantifying the correlation in causal effects at known shared regions, is of great importance in epidemiological and etiological studies. For example, genetic variants associated with multiple traits in genome-wide associations studies (GWAS), can be used as instrumental variables in Mendelian randomization analyses to identify causal relationships among complex traits[7,8,9,10]. Unfortunately many risk variants are left undetected by existing GWAS due to a combination of high polygenicity (i.e. many variants of small effects) and sample sizes which limits the power to detect genetic variants of small effect[11]. To improve accuracy at sub-GWAS significant loci, recent works[1,2] proposed to utilize the posterior probability of two traits sharing a causal variant at each locus to detect genetic overlap. Although powerful in detecting shared genetic risk variants, the posterior probability does not convey the direction or magnitude of the genetic effect at the overlapped loci[1,2]. Alternative approaches have used genetic correlation (i.e. the correlation between the causal effects), that summarizes both direction and magnitude of effects, to gain insights into genetic overlap of complex traits[12,13]. Due to polygenicity assumptions, genetic correlation has been only investigated in genome-wide context by aggregating information across all variants in the genome[14,15]. In this work, we investigate local genetic correlation (i.e. correlation between a pair of traits only due to genetic variants from a small region in the genome) as means to dissect the genetic sharing between pairs of traits.

Traditional methods to estimate genetic correlation between a pair of traits rely on pedigree or family data, requiring phenotype measurements of the traits on the same set of individuals[13,14]. While more recent bivariate-REML[15] and HE-regression[16] methods provide the convenience to estimate genetic correlation from genotype and phenotype data of unrelated individuals, they are hindered by the lack of availability of large-scale individual-level data due to privacy concerns[12,14,17,18]. A recently proposed method, cross-trait LD score regression (LDSC) circumvented this hindrance by estimating genetic correlation from GWAS

summary data, which are publicly available from most large GWAS consortia[19,20], and created an atlas of genetic correlation across multiple human complex traits and diseases[14]. However, due to model assumptions on trait polygenicity, cross-trait LDSC has been only applied for estimating genome-wide genetic correlation[12,14], which quantifies the direction and magnitude of genetic sharing across the entire genome[14].

In this work we introduce ρ-HESS, a method to estimate the local genetic correlation between a pair of traits at each region in the genome from GWAS summary data, while accounting for overlapping GWAS samples and linkage disequilibrium (LD). Our method estimates the contribution to the genetic correlation of typed variants for each LD block in the genome; we utilize approximately independent LD blocks of roughly 1.5Mb in size on the average[21]. We make no distributional assumption on the causal effect sizes by treating them as fixed quantities, which allows for a broad range of causal genetic architectures at small regions in the genome. Our approach can be viewed as a natural extension to pairs of traits of recently proposed methods that quantify local SNP-heritability from GWAS summary data under a fixed effects model[18]. Through extensive simulations, we demonstrate that given in-sample LD, ρ-HESS yields unbiased estimates of local genetic covariance, and approximately unbiased estimates of local genetic covariance and correlation when LD is estimated from external population reference panels such as the 1000 Genomes data[22].

We analyzed GWAS summary data across 35 complex traits and identified 234 pairs of traits with significant genome-wide genetic correlation; these include previously reported correlations (e.g., lipid traits) as well as correlations between traits not reported before (e.g., blood related traits). Second, we identify 27 genomic regions that show significant local genetic covariance as well as local SNP-heritability across 29 pairs of traits. For example, ρ-HESS estimates a local genetic correlation of -0.95 (s.e. 0.16) for the region chr2:21-23M, which harbors the APOB gene for HDL and TG. Notably, 7 (out of the 27) significantly correlated genomic regions across 12 pairs of traits are significant in local analyses, although the genome-wide genetic correlation is not significantly different from 0. For example, when considering genetic correlation between mean cell volume (MCV) and platelet count (PLT), region chr6:134-136M shows a local genetic correlation of 1.0 (s.e. 0.16), whereas the genome-wide genetic correlation is negligible (0.01 s.e. 0.04). This shows that these traits harbor

genetic overlap at a local level (e.g., due to pleiotropy and/or shared pathways) and emphasize the power of local correlation analysis.

Having estimates of local genetic correlation at every region in the genome allows us to perform bi-directional analyses to assign putative direction of causality for a pair of traits. Intuitively, all loci that harbor variants that contribute to one trait will have a consistent direction of effect on the second trait under a causal model, whereas under no causal model we expect no consistency in the direction of genetic effects (see Figure 1). This extends previously proposed works[2] from individual variant effects to local genetic covariances. We assign putative direction of causality for 15 pairs of traits. Reassuringly, our analyses show that height causally decreases body mass index (BMI), and that BMI causally increases triglyceride (TG), consistent with expectation and previous works[2,3]. We also identify putative causality between previously unreported traits. For example, our analyses suggest that schizophrenia (SCZ) causally increases the risk for ulcerative colitis (UC), providing possible explanation to why prevalence of UC is higher in SCZ patients but not vice versa[23]. Interestingly we also find evidence of age at menarche (AM) causally decreases the level of triglyceride (TG). Overall, our results motivate further work in confirming putative causality direction among these traits.

Results

Overview of methods

Genetic correlation measures the similarity between a pair of traits driven by genetic variation (e.g., correlation on causal effect sizes), and enjoys wide applications in understanding relations between complex traits[12,24,25]. Genetic correlation is traditionally estimated as a single measure across the entire genome to capture the contribution of genetic variation across the entire genome to the correlation between phenotypes. Here, we introduce local genetic correlation, the similarity between pairs of traits driven by genetic variation localized at a specific region in the genome (e.g., one LD block), as a principled way to dissect the genome-wide genetic correlation between traits. For example, a high genome-wide genetic

covariance can be driven by one or a few genomic regions containing a shared risk variant, or by a large number of regions each with a small contribution (see Figure 1). The distribution of local genetic covariances reflects causality relations (where all risk variants for one trait are risk variants for the other trait) and/or pleiotropic regions (risk variants contributing to both traits through different pathways).

Under the assumption that true causal effects of genetic variants on the two traits are known, the local genetic covariance between the two traits is simply $\boldsymbol{\beta}^{\intercal}\boldsymbol{V}\boldsymbol{\gamma}$, where $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are the causal effect vectors and $\boldsymbol{V}$ is the pairwise correlation between variants (i.e. linkage disequilibrium, LD) (see Methods). A traditional GWAS estimates marginal effect sizes at each variant that contain statistical noise due to finite sample size and are confounded by LD[19]. We deconvolute LD and account for the statistical noise to estimate the local genetic covariance as:

$$\hat{\rho}_{g,local} = \frac{n_1 n_2 \hat{\boldsymbol{\beta}}_{gwas}^{\intercal} \boldsymbol{V}^{\dagger} \hat{\boldsymbol{\gamma}}_{gwas} - n_s q \rho}{n_1 n_2 - n_s q}. \tag{1}$$

Here, $\hat{\boldsymbol{\beta}}_{gwas}$ and $\hat{\boldsymbol{\gamma}}_{gwas}$ are the vectors of standardized marginal GWAS effect sizes, n_1 and n_2 the sample size of each GWAS study, n_s the number of samples shared across the two GWASs, q the rank of the LD matrix $\boldsymbol{V}$, ρ the total phenotypic correlation between the two traits and $\boldsymbol{V}^{\dagger}$ is Moore-Penrose pseudoinverse of the LD matrix ($\boldsymbol{V}$). The variance of the estimator follows from bilinear form theory, and has magnitude approximately on the order of $\frac{q}{n_1 n_2}$ (see Methods). To obtain local genetic correlation, we standardize the covariance as $\hat{\rho}_{g,local} / \sqrt{\hat{h}^2_{\phi,g,local} \hat{h}^2_{\psi,g,local}}$, where $\hat{h}^2_{*,g,local}$ are the local SNP-heritability[18] estimates for the two traits.

When in-sample LD matrix ($\boldsymbol{V}$) is available, Equation (1) yields an unbiased estimator of local genetic covariance (see Results). When individual-level genotype data is not available, we use LD matrix ($\hat{\boldsymbol{V}}$) estimated from external reference panels (e.g., 1000 Genomes Project). The small sample size of external reference panels creates statistical noise in the estimated external reference LD matrix $\hat{\boldsymbol{V}}$ and results in biased estimates of local genetic covariance and correlation. To account for the statistical noise, we apply truncated-SVD regularization (see Methods).

Accuracy of local correlation estimation in simulations

We evaluated the performance of ρ-HESS through extensive simulations across various disease architectures. First, we assessed the performance of our approach across varying simulated local genetic covariances. As expected, when in-sample LD matrix is available, ρ-HESS provides an unbiased estimate of local genetic covariance (see Figure 2). For completeness, we also adapted cross-trait LDSC[14] for local estimation and observed biased results; this is expected as cross-trait LDSC is not designed for local estimation under fixed effects at causal variants. For example, for simulated local genetic covariance of 1.7×10^{-4}, ρ-HESS yields an estimate of 2.0×10^{-4} ($s.e.$ 4.3×10^{-5}), whereas cross-trait LDSC yields an estimate of 6.0×10^{-4}($s.e.$ 3.4×10^{-5}). Second, we assessed the performance of local genetic correlation (i.e. the local genetic covariance divided by the square root of their respective local heritability estimates). Overall, the local genetic correlation estimates are approximately unbiased; e.g., for a simulated local genetic correlation of 0.35 ρ-HESS yielded an estimate of 0.22 (s.e. 0.19) (see Figure 2). Third, we quantified the performance of our approach in the case when in-sample LD is unavailable and needs to be estimated from external reference panels. We observe that the ρ-HESS estimates of local genetic covariance and correlation are still approximately unbiased (see Figure 2) when truncated-SVD is used for regularizing the LD matrix (see Methods). Noticeably, the standard errors of ρ-HESS estimates also decrease due to the truncated-SVD regularization. Finally, we compared the performance of local genetic correlation estimation in simulations with different degrees of polygenecity. Overall, we observe that ρ-HESS is not sensitive to the underlying polygenicity of the trait, and provides nearly unbiased and consistent estimates of local genetic covariance and correlation (see Figure 2).

Partitioning the genetic correlation across local genomic regions

We analyzed GWAS summary data of 35 complex traits to obtain local genetic covariances/correlations at 1,703 approximately LD-independent regions in the genome (∼1.5Mb on the average)[21]. First, we aggregated the local estimates into genome-wide estimates of genetic correlation (see Methods, Supplementary Figure 2-8) and found a high degree of

concordance with genetic correlation estimated by cross-trait LDSC regression ($R = 0.74$, see Figure 3, Supplementary Figure 1). We note that our estimator provides consistently lower estimates than LDSC for pairs of traits from the same consortium where our approach assumes full sample overlap and therefore is more conservative (see Discussion). We report 234 pairs of traits with significant genome-wide genetic correlation after correcting for 595 pairs investigated ($p < 0.05/595$). These include previously reported genetic correlations, e.g. body mass index (BMI) and triglyceride (TG), as well as complex traits that have not been studied before using genetic correlation, e.g. type 2 diabetes (T2D) and ulcerative colitis (UC).

Next, we searched for genomic regions that disproportionately contribute to the genetic correlation of the 35 analyzed traits; we excluded the HLA region due to complex LD patterns. We identify 27 genomic regions that show significant local genetic correlation (two-tailed $p < 0.05/1703/595$) as well as significant local SNP-heritability (one-tailed $p < 0.05/1703/35$) (see Figures 4, 5, Supplementary Figure 9, Supplementary Table 1 and 2). 22 out of 27 regions contain GWAS significant SNPs for both traits, whereas 5 of these loci contain GWAS significant hits for only one or none of traits and can be viewed as new risk regions for these traits. For example, the estimate of local genetic correlation between HDL and TG at chr11:116-117Mb is -0.82 (s.e. 0.07), suggesting highly shared genetic architecture at this locus for HDL and TG. Indeed, the locus chr11:116-117M harbors the APOA gene, which is known to be associated with multiple lipid traits[26].

Since genetic correlation is an aggregated manifestation of local genetic covariance, for pairs of traits with highly positive or negative genetic correlation, we expect the distribution of local genetic covariances to be shifted towards the positive or negative side; whereas for pairs of traits with low genetic correlation, we expect the distribution of local genetic covariances to be centered around zero (see Figure 6,7,8). Indeed, pairs of traits with higher genome-wide genetic correlation tend to harbor more loci with significant local genetic covariance (see Figure 4). For instance, only one locus exhibits significant local genetic covariance for the pair of traits age at menarche (AM) and height ($r_g = 0.13$, s.e. 0.01), whereas four loci show significant local genetic covariance for the pair of traits LDL and TG ($r_g = 0.44$, s.e. 0.02).

Local correlations for pairs of traits with negligible genome-wide correlation

Although many complex traits are known to share risk loci, some pairs of traits show negligible genome-wide genetic correlation. For example HDL and LDL share several GWAS risk loci[26] but the genome-wide genetic correlation is negligible (-0.05, s.e. 0.02)[14](see Figure 3). The absence of significant genome-wide genetic correlation between these pairs of traits can be attributed to either symmetric distribution of local genetic covariance (positive local genetic covariance cancels out negative local genetic covariance, see Figure 1) and/or lack of power to declare significance for genome-wide genetic correlation. Thus, we hypothesize that at the locus-specific level, many loci may manifest significant local genetic covariance even if the genome-wide genetic correlation between a pair of traits is not significant. Indeed, 9 genomic regions show significant local genetic correlation (two-tailed $p < 0.05/1703$) for HDL and LDL (see Figure 7). Some of these loci, e.g. chr2:21M-23M, chr11:116M-117M, and chr19:44M-46M, harbor genes (APOB, APOA, and APOE, respectively) that are well known to be involved in lipid genetics[26,27,28]. Across all pairs of traits with non-significant genome-wide correlation, we identify 7 regions across 12 pairs of traits with significant local genetic covariance (two-tailed $p < 0.05/1703/595$) as well as significant local SNP-heritability (one-tailed $p < 0.05/1703/35$) (see Figure 5, Supplementary Table 2). Most of these loci also show significant local genetic correlation, suggesting high similarity in the effects of genetics on traits at these loci (see Figure 5, Supplementary Table 2). For example the region chr6:134-136M harbors the blood-trait-associated gene HBS1L[29,30], and contributes to local genetic covariance across many blood traits (MCH, MCV, RBC, and PLT).

Putative causal relationships between complex traits

The distribution of local genetic correlations across the genome can be leveraged to infer putative direction of causality across pairs of traits. We illustrate this approach through 4 examples of putative causality directions (see Figure 9). First, consider the example of BMI and TG, genetically correlated genome-wide ($\hat{r}_g$=0.53, s.e. 0.02). The local genetic correlation restricted only to loci that harbor a GWAS hit for BMI (and no GWAS hit for

TG) is significant $\hat{r}_{g,local,BMI}$=0.52, s.e. 0.05 and consistent with genome-wide correlation. In contrast, the local genetic correlation estimated at loci that harbor GWAS hits for TG (and no GWAS hits for BMI) is not significantly different from 0 ($\hat{r}_{g,local,TG}$=-0.020, s.e. 0.053). This shows that genetic variants that contribute to BMI have a consistent effect on TG whereas genetic variants associated to TG do not have a consistent effect on BMI. These two observations provide evidence favoring the model in which BMI causally increases TG (see Figure 9). Second, the pattern where one of the conditional correlations is not significant needs not be just in the positive direction. For example, BMI and height are genome-wide genetically correlated at ($\hat{r}$ = -0.08, s.e. 0.01), whereas $\hat{r}_{g,local,height}$ is significantly negative (-0.15, s.e. 0.02), and $\hat{r}_{g,local,height}$ close to zero (-0.01, s.e. 0.04). This supports a model where height causally decreases BMI (as expected, see Figure 9). We note, however, that local genetic correlations between can also be induced by shared biological pathways driven by genetic variations at these loci and/or an unobserved confounder; for example, shared genetic variants impacting a sub-phenotype that is pleiotropic for both traits could induce local genetic correlations consistent with a causal model. Thus, we exercise caution when utilizing the local genetic correlations to identify putative causality. Finally, the conditional correlation approach cannot distinguish between putative causality (or other complex model that includes unobserved phenotypes) when both conditional correlations are significant or when both correlations are not significantly different from 0 (e.g., AM-BMI and MCH-SCZ, see Figure 9)

Through our bi-directional analyses (see Methods), we identified a total of 15 pairs of traits that support a causal directional effect (see Figure 10, Supplementary Figure 10-13). As an example, our bi-direction analyses provide evidence favoring the model, in which age at menarche (AM) causally decreases triglyceride (TG) level ($\hat{r}_{g,local,AM} = -0.28$ $s.e.$ 0.06, $\hat{r}_{g,local,TG} = 0.04$ $s.e.$ 0.05), suggesting that the association observed between AM and TG is likely induced by the effect of AM on TG[31,32]. Previous work also shows that association exists between schizophrenia (SCZ) and ulcerative colitis (UC), however, the direction of causality is uncertain[23]. Our bi-directional analyses favors a model in which SCZ causally increases the risk of UC ($\hat{r}_{g,local,SCZ} = 0.30$ $s.e.$ 0.05, $\hat{r}_{g,local,UC} = -0.02$ $s.e.$ 0.07). This result may provide insights into the previous observations of no increase in prevalence of SCZ in UC

patients[33]. However, our result does not rule out other causal models, for example, it could be that shared genetic variants drive an immune system phenotype that is pleiotropic for both SCZ and UC, inducing the relationship between $\hat{r}_{g,local,SCZ}$ and $\hat{r}_{g,local,UC}$[34,35]. Interestingly, we also observe that our bi-directional analyses support a model in which years of education (EY) causally decreases hemoglobin level (HB), LDL, TG, and the risk for rheumatoid arthritis (RA) (see Figure 10). We note that these results are consistent with previous conclusions on the causal effect of education on health[36,37]. However, we emphasize that education attainment (or other studied traits) may be confounded by other factors such as social status and that one should exercise caution when interpreting these results. Finally, our bi-directional analyses over local genetic correlation also suggest that BMI likely causally increases triglyceride (TG) and that height causally decreases BMI, consistent with our expectations and previous findings[2,3].

Discussion

We have described ρ-HESS, a method to estimate local genetic covariance and correlation from GWAS summary association data. Through extensive simulations, we demonstrated that our method is approximately unbiased and provides consistent results irrespective of causal architecture. We then applied our method to large-scale GWAS summary association data of 35 complex traits. Compared with cross-trait LDSC, our methods identified considerably more pairs of traits displaying significant genome-wide genetic correlation likely because of the truncated-SVD regularization of the LD matrix, which decreases the standard error of the estimates. We identify genomic regions that are significantly correlated across pairs of traits regardless of the significance of genome-wide correlation. Finally, we performed bi-directional analyses over the local genetic covariances, and report plausible causal relationships for 15 pairs of traits.

We conclude with several limitations of our methods highlighting areas for future work. First, our estimator requires phenotype correlation between two traits, as well as the number of shared individuals between the two GWASs. We estimate the phenotype correlation through cross-trait LDSC assuming full sample overlap between GWAS within the same

consortium and no sample overlap between GWAS across two consortium. Although we believe this is supported by the data we analyzed in this study, mis-specifications in the overlapping sample sizes could introduce biases. Second, we emphasize that our bi-directional analyses only identify putative causality and is not proof of causality relations; exact inference of causal relations is complicated by unobserved confounders such as socioeconomic status and/or biological pathways. Furthermore, most of the GWAS summary association data are adjusted for covariates such as age, gender, to increase statistical power[38], and previous works have shown that adjusting for covariates can potentially lead to false positives[39]. In the bi-directional analyses over local genetic correlation, covariate-adjusted summary association data can potentially lead to false inference of causality. However, the extent to which covariate-adjusted summary association data lead to false discovery of causality remains unclear, and we leave detailed investigation of this issue as future work. Third, in our real data analyses, we made the assumption that the loci are independent of each other. In reality however correlations may exist across adjacent loci due to long range LD, and can lead to biased estimates. Nevertheless, we note that previous works have indicated the effect of LD leakage to be minimal, and we conjecture that this statement still hold in estimating local genetic covariance. Lastly, we uses truncated-SVD to regularize LD matrix and to reduce standard error in the estimates of local genetic covariance, at the cost of introducing bias. Currently, we use a fixed number of eigenvectors in the truncated-SVD regularization, across all the loci. However, this approach may not be optimal for genomic regions with different LD structure, and leave a principled approach of estimating the number of eigenvectors as future work.

Acknowledgement

We are grateful to Gleb Kichaev, Malika Kumar, Suraj Alva, and James Boocock for their helpful discussions that greatly improved the quality of this manuscript. We also thank Dr. Nicole Soranzo for kindly sharing summary data for the platelet traits. The authors declare no conflict of interest.

References

[1] Claudia Giambartolomei, Damjan Vukcevic, Eric E Schadt, Lude Franke, Aroon D Hingorani, Chris Wallace, and Vincent Plagnol. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet*, 10(5): e1004383, 2014.

[2] Joseph K Pickrell, Tomaz Berisa, Jimmy Z Liu, Laure Ségurel, Joyce Y Tung, and David A Hinds. Detection and interpretation of shared genetic influences on 42 human traits. *Nature genetics*, 2016.

[3] Nicholas Mancuso, Huwenbo Shi, Page Goddard, Gleb Kichaev, Alexander Gusev, and Bogdan Pasaniuc. Integrating gene expression with summary association statistics to identify susceptibility genes for 30 complex traits. *bioRxiv*, 2016. doi: 10.1101/072967. URL http://biorxiv.org/content/early/2016/09/01/072967.

[4] Alexander Gusev, Arthur Ko, Huwenbo Shi, Gaurav Bhatia, Wonil Chung, Brenda WJH Penninx, Rick Jansen, Eco JC De Geus, Dorret I Boomsma, Fred A Wright, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nature genetics*, 2016.

[5] Alkes L Price, Chris CA Spencer, and Peter Donnelly. Progress and promise in understanding the genetic basis of common diseases. In *Proc. R. Soc. B*, volume 282, page 20151684. The Royal Society, 2015.

[6] Nuala A Sheehan, Vanessa Didelez, Paul R Burton, and Martin D Tobin. Mendelian randomisation and causal inference in observational epidemiology. *PLoS Med*, 5(8):e177, 2008.

[7] Benjamin F Voight, Gina M Peloso, Marju Orho-Melander, Ruth Frikke-Schmidt, Maja Barbalic, Majken K Jensen, George Hindy, Hilma Hólm, Eric L Ding, Toby Johnson, et al. Plasma hdl cholesterol and risk of myocardial infarction: a mendelian randomisation study. *The Lancet*, 380(9841):572–580, 2012.

[8] Debbie A Lawlor, Roger M Harbord, Jonathan AC Sterne, Nic Timpson, and George Davey Smith. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Statistics in medicine*, 27(8):1133–1163, 2008.

[9] George Davey Smith and Gibran Hemani. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Human molecular genetics*, 23(R1):R89–R98, 2014.

[10] George Davey Smith, Shah Ebrahim, Sarah Lewis, Anna L Hansell, Lyle J Palmer, and Paul R Burton. mendelian randomization.

[11] Jian Yang, Beben Benyamin, Brian P McEvoy, Scott Gordon, Anjali K Henders, Dale R Nyholt, Pamela A Madden, Andrew C Heath, Nicholas G Martin, Grant W Montgomery, et al. Common snps explain a large proportion of the heritability for human height. *Nature genetics*, 42(7):565–569, 2010.

[12] Brendan K Bulik-Sullivan, Po-Ru Loh, Hilary K Finucane, Stephan Ripke, Jian Yang, Nick Patterson, Mark J Daly, Alkes L Price, Benjamin M Neale, Schizophrenia Working Group of the Psychiatric Genomics Consortium, et al. Ld score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics*, 47(3):291–295, 2015.

[13] Michael Neale and Lon Cardon. *Methodology for genetic studies of twins and families*, volume 67. Springer Science & Business Media, 1992.

[14] Brendan Bulik-Sullivan, Hilary K Finucane, Verneri Anttila, Alexander Gusev, Felix R Day, Po-Ru Loh, Laramie Duncan, John RB Perry, Nick Patterson, Elise B Robinson, et al. An atlas of genetic correlations across human diseases and traits. *Nature genetics*, 2015.

[15] Sang Hong Lee, Jian Yang, Michael E Goddard, Peter M Visscher, and Naomi R Wray. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics*, 28(19):2540–2542, 2012.

[16] JK Haseman and RC Elston. The investigation of linkage between a quantitative trait and a marker locus. *Behavior genetics*, 2(1):3–19, 1972.

[17] Hilary K Finucane, Brendan Bulik-Sullivan, Alexander Gusev, Gosia Trynka, Yakir Reshef, Po-Ru Loh, Verneri Anttila, Han Xu, Chongzhi Zang, Kyle Farh, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature genetics*, 47(11):1228–1235, 2015.

[18] Huwenbo Shi, Gleb Kichaev, and Bogdan Pasaniuc. Contrasting the genetic architecture of 30 complex traits from summary association data. *bioRxiv*, page 035907, 2016.

[19] Bogdan Pasaniuc and Alkes L. Price. Dissecting the genetics of complex traits using summary association statistics. *Nat Rev Genet*, advance online publication, Nov 2016. ISSN 1471-0064. URL http://dx.doi.org/10.1038/nrg.2016.142. Review.

[20] JIE ZHENG, Mesut Erzurumluoglu, Benjamin Elsworth, Laurence Howe, Philip Haycock, Gibran Hemani, Katherine Tansey, Charles Laurin, Beate St. Pourcain, Nicole Warrington, Hilary Finucane, Alkes Price, Brendan Bulik-Sullivan, Verneri Anttila, Lavinia Paternoster, Tom Gaunt, David Evans, and Benjamin Neale. Ld hub: a centralized database and web interface to perform ld score regression that maximizes the potential of summary level gwas data for snp heritability and genetic correlation analysis. *Bioinformatics*, September 2016. doi: 10.1093/bioinformatics/btw613. URL https://doi.org/10.1093/bioinformatics/btw613.

[21] Tomaz Berisa and Joseph K Pickrell. Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics (Oxford, England)*, 32(2):283, 2016.

[22] 1000 Genomes Project Consortium et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491(7422):56–65, 2012.

[23] FC Dohan. More on celiac disease as a model for schizophrenia. 1983.

[24] Joseph P Hegmann and Bernard Possidente. Estimating genetic correlations from inbred strains. *Behavior genetics*, 11(2):103–114, 1981.

[25] Gregory Carey. Inference about genetic correlations. *Behavior genetics*, 18(3):329–338, 1988.

[26] Global Lipids Genetics Consortium et al. Discovery and refinement of loci associated with lipid levels. *Nature genetics*, 45(11):1274–1283, 2013.

[27] Godfrey S Getz and Catherine A Reardon. Apoprotein e as a lipid transport and signaling protein in the blood, liver, and artery wall. *Journal of lipid research*, 50 (Supplement):S156–S161, 2009.

[28] C Pallaud, R Gueguen, C Sass, M Grow, S Cheng, G Siest, and S Visvikis. Genetic influences on lipid metabolism trait variability within the stanislas cohort. *Journal of lipid research*, 42(11):1879–1890, 2001.

[29] Nicole Soranzo, Tim D Spector, Massimo Mangino, Brigitte Kühnel, Augusto Rendon, Alexander Teumer, Christina Willenborg, Benjamin Wright, Li Chen, Mingyao Li, et al. A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the haemgen consortium. *Nature genetics*, 41(11):1182–1190, 2009.

[30] Pim Van Der Harst, Weihua Zhang, Irene Mateo Leach, Augusto Rendon, Niek Verweij, Joban Sehmi, Dirk S Paul, Ulrich Elling, Hooman Allayee, Xinzhong Li, et al. Seventy-five genetic loci influencing the human red blood cell. *Nature*, 492(7429):369–375, 2012.

[31] Noel T Mueller, Bruce B Duncan, Sandhi M Barreto, Dora Chor, Marina Bessel, Estela ML Aquino, Mark A Pereira, and Maria Inês Schmidt. Earlier age at menarche is associated with higher diabetes risk and cardiometabolic disease risk factors in brazilian adults: Brazilian longitudinal study of adult health (elsa-brasil). *Cardiovascular diabetology*, 13(1):1, 2014.

[32] Doris Stöckl, Christa Meisinger, Annette Peters, Barbara Thorand, Cornelia Huth, Margit Heier, Wolfgang Rathmann, Bernd Kowall, Heidi Stöckl, and Angela Döring. Age at menarche and its association with the metabolic syndrome and its components: results from the kora f4 study. *PloS one*, 6(10):e26076, 2011.

[33] J West, RF Logan, RB Hubbard, and TR Card. Risk of schizophrenia in people with coeliac disease, ulcerative colitis and crohn's disease: a general population-based study. *Alimentary pharmacology & therapeutics*, 23(1):71–74, 2006.

[34] Norbert Muller and Markus J Schwarz. The role of immune system in schizophrenia. *Current immunology reviews*, 6(3):213–220, 2010.

[35] Richard P Macdermott and William F Stenson. Alterations of the immune system in ulcerative colitis and crohn's disease. *Advances in immunology*, 42:285–328, 1988.

[36] Mary A Silles. The causal effect of education on health: Evidence from the united kingdom. *Economics of Education review*, 28(1):122–128, 2009.

[37] David P Baker, Juan Leon, Emily G Smith Greenaway, John Collins, and Marcela Movit. The education effect on population health: a reassessment. *Population and development review*, 37(2):307–332, 2011.

[38] Joel Mefford and John S Witte. The covariate's dilemma. *PLoS Genet*, 8(11):e1003096, 2012.

[39] Hugues Aschard, Bjarni J Vilhjálmsson, Amit D Joshi, Alkes L Price, and Peter Kraft. Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. *The American Journal of Human Genetics*, 96(2):329–339, 2015.

[40] Leon Isserlis. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12(1/2):134–139, 1918.

[41] Shayle R Searle. *Linear models*, page 65. John Wiley & Sons, Inc., 1971.

[42] Bradley Efron. Bayesian inference and the parametric bootstrap. *The annals of applied statistics*, 6(4):1971, 2012.

[43] Richard A Gibbs, John W Belmont, Paul Hardenbol, Thomas D Willis, Fuli Yu, Huanming Yang, Lan-Yang Ch'ang, Wei Huang, Bin Liu, Yan Shen, et al. The international hapmap project. *Nature*, 426(6968):789–796, 2003.

[44] Adam E Locke, Bratati Kahali, Sonja I Berndt, Anne E Justice, Tune H Pers, Felix R Day, Corey Powell, Sailaja Vedantam, Martin L Buchkovich, Jian Yang, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*, 518(7538): 197–206, 2015.

[45] Andrew R Wood, Tonu Esko, Jian Yang, Sailaja Vedantam, Tune H Pers, Stefan Gustafsson, Audrey Y Chu, Karol Estrada, Jian'an Luan, Zoltán Kutalik, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature genetics*, 46(11):1173–1186, 2014.

[46] Dmitry Shungin, Thomas W Winkler, Damien C Croteau-Chonka, Teresa Ferreira, Adam E Locke, Reedik Mägi, Rona J Strawbridge, Tune H Pers, Krista Fischer, Anne E Justice, et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature*, 518(7538):187–196, 2015.

[47] Christian Gieger, Aparna Radhakrishnan, Ana Cvejic, Weihong Tang, Eleonora Porcu, Giorgio Pistis, Jovana Serbanovic-Canic, Ulrich Elling, Alison H Goodall, Yann Labrune, et al. New gene functions in megakaryopoiesis and platelet formation. *Nature*, 480(7376):201–208, 2011.

[48] Josée Dupuis, Claudia Langenberg, Inga Prokopenko, Richa Saxena, Nicole Soranzo, Anne U Jackson, Eleanor Wheeler, Nicole L Glazer, Nabila Bouatia-Naji, Anna L Gloyn, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on

type 2 diabetes risk. *Nature genetics*, 42(2):105–116, 2010.

[49] Nicole Soranzo, Serena Sanna, Eleanor Wheeler, Christian Gieger, Dörte Radke, Josée Dupuis, Nabila Bouatia-Naji, Claudia Langenberg, Inga Prokopenko, Elliot Stolerman, et al. Common variants at 10 genomic loci influence hemoglobin a1c levels via glycemic and nonglycemic pathways. *Diabetes*, 59(12):3229–3239, 2010.

[50] Aysu Okbay, Jonathan P Beauchamp, Mark Alan Fontana, James J Lee, Tune H Pers, Cornelius A Rietveld, Patrick Turley, Guo-Bo Chen, Valur Emilsson, S Fleur W Meddens, et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, 533(7604):539–542, 2016.

[51] Aysu Okbay, Bart ML Baselmans, Jan-Emmanuel De Neve, Patrick Turley, Michel G Nivard, Mark Alan Fontana, S Fleur W Meddens, Richard Karlsson Linnér, Cornelius A Rietveld, Jaime Derringer, et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature genetics*, 2016.

[52] Hou-Feng Zheng, Vincenzo Forgetta, Yi-Hsiang Hsu, Karol Estrada, Alberto Rosello-Diez, Paul J Leo, Chitra L Dahia, Kyung Hyun Park-Min, Jonathan H Tobias, Charles Kooperberg, et al. Whole-genome sequencing identifies en1 as a determinant of bone density and fracture. *Nature*, 526(7571):112–117, 2015.

[53] John RB Perry, Felix Day, Cathy E Elks, Patrick Sulem, Deborah J Thompson, Teresa Ferreira, Chunyan He, Daniel I Chasman, Tõnu Esko, Gudmar Thorleifsson, et al. Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature*, 514(7520):92–97, 2014.

[54] Yukinori Okada, Di Wu, Gosia Trynka, Towfique Raj, Chikashi Terao, Katsunori Ikari, Yuta Kochi, Koichiro Ohmura, Akari Suzuki, Shinji Yoshida, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*, 506(7488):376–381, 2014.

[55] Schizophrenia Working Group of the Psychiatric Genomics Consortium et al. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, 511(7510):421–427, 2014.

[56] Jimmy Z Liu, Suzanne van Sommeren, Hailiang Huang, Siew C Ng, Rudi Alberts,

Atsushi Takahashi, Stephan Ripke, James C Lee, Luke Jostins, Tejas Shah, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nature genetics*, 47(9):979–986, 2015.

[57] Andrew P Morris, Benjamin F Voight, Tanya M Teslovich, Teresa Ferreira, Ayellet V Segre, Valgerdur Steinthorsdottir, Rona J Strawbridge, Hassan Khan, Harald Grallert, Anubha Mahajan, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature genetics*, 44(9): 981, 2012.

Methods

Local genetic correlation under fixed-effect model

Let $\phi = \boldsymbol{x}^\intercal\boldsymbol{\beta} + \epsilon$ and $\psi = \boldsymbol{x}^\intercal\boldsymbol{\gamma} + \delta$ be two traits measured at an individual, standardized so that $\mathrm{E}[\phi] = \mathrm{E}[\psi] = 0$ and $\mathrm{Var}[\phi] = \mathrm{Var}[\psi] = 1$, where $\boldsymbol{\beta}, \boldsymbol{\gamma} \in \mathbb{R}^p$ are the fixed effect size vectors for the two traits; $\boldsymbol{x} \in \mathbb{R}^p$, the genotype vector of the individual at p SNPs, standardized so that $\mathrm{E}[\boldsymbol{x}] = \boldsymbol{0}$, and $\mathrm{Var}[\boldsymbol{x}] = \boldsymbol{V}$, the LD matrix; and ϵ, δ, environmental effects independent of $\boldsymbol{x}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, with $\mathrm{E}[\epsilon] = \mathrm{E}[\delta] = 0$, $\mathrm{Var}[\epsilon] = \sigma_\epsilon^2$, $\mathrm{Var}[\delta] = \sigma_\delta^2$, and $\mathrm{Cov}[\epsilon, \delta] = \rho_e$. Under these assumptions, one can decompose the phenotypic covariance, ρ, between ϕ and ψ as

$$\begin{aligned}\rho &= \mathrm{Cor}[\phi, \psi] = \mathrm{Cov}[\phi, \psi] = \mathrm{E}[\phi\psi] - \mathrm{E}[\phi]\,\mathrm{E}[\psi] \\ &= \mathrm{E}[(\boldsymbol{x}^\intercal\boldsymbol{\beta} + \epsilon)(\boldsymbol{x}^\intercal\boldsymbol{\gamma} + \delta)^\intercal] = \boldsymbol{\beta}^\intercal\,\mathrm{E}[\boldsymbol{x}\boldsymbol{x}^\intercal]\boldsymbol{\gamma} + \mathrm{Cov}[\epsilon, \delta] = \boldsymbol{\beta}^\intercal\boldsymbol{V}\boldsymbol{\gamma} + \rho_e,\end{aligned} \tag{2}$$

where $\rho_g = \boldsymbol{\beta}^\intercal\boldsymbol{V}\boldsymbol{\gamma}$ is the genetic covariance between the two traits. For standardized traits, phenotypic covariance coincides with phenotypic correlation. Thus, given the true effect size vectors, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and the LD matrix $\boldsymbol{V}$, one can easily obtain ρ_g. We define genetic correlation, r_g, as $r_g = \frac{\rho_g}{\sqrt{h^2_{g\phi} h^2_{g\psi}}}$, i.e. genetic covariance standardized by square root of SNP-heritability of the two traits.

Estimating local genetic covariance from GWAS summary data

In two GWASs involving n_1 individuals for trait 1 ($\boldsymbol{\phi}$), n_2 individuals for trait 2 ($\boldsymbol{\psi}$), and n_s shared individuals, we assume

$$\begin{bmatrix} \boldsymbol{\phi} \\ \boldsymbol{\phi}_s \end{bmatrix} = \begin{bmatrix} \boldsymbol{Y} \\ \boldsymbol{X}_s \end{bmatrix}\boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{\epsilon} \\ \boldsymbol{\epsilon}_s \end{bmatrix}, \begin{bmatrix} \boldsymbol{\psi} \\ \boldsymbol{\psi}_s \end{bmatrix} = \begin{bmatrix} \boldsymbol{Z} \\ \boldsymbol{X}'_s \end{bmatrix}\boldsymbol{\gamma} + \begin{bmatrix} \boldsymbol{\delta} \\ \boldsymbol{\delta}_s \end{bmatrix}, \tag{3}$$

where $(\boldsymbol{\phi}, \boldsymbol{\phi}_s) \in \mathbb{R}^{n_1}$ and $(\boldsymbol{\psi}, \boldsymbol{\psi}_s) \in \mathbb{R}^{n_2}$ are the standardized trait values of all individuals in each GWAS; $(\boldsymbol{Y}, \boldsymbol{X}_s) \in \mathbb{R}^{n_1 \times p}$, $(\boldsymbol{Z}, \boldsymbol{X}'_s) \in \mathbb{R}^{n_2 \times p}$, column standardized genotype matrices of all individuals in each GWAS, where $\boldsymbol{X}_s$ and $\boldsymbol{X}'_s$ represent the genotype matrices for the same set of individuals and SNPs but standardized differently in each GWAS; $(\boldsymbol{\epsilon}, \boldsymbol{\epsilon}_s) \in \mathbb{R}^{n_1}$,

$(\boldsymbol{\delta}, \boldsymbol{\delta}_s) \in \mathbb{R}^{n_2}$, environmental effects of all individuals in each GWAS. We use the subscript 's' to represent individuals shared by both GWASs. We further assume that $\mathrm{E}[\boldsymbol{\epsilon}] = \mathrm{E}[\boldsymbol{\delta}] = \mathrm{E}[\boldsymbol{\epsilon}_s] = \mathrm{E}[\boldsymbol{\delta}_s] = \mathbf{0}$, $\mathrm{Var}[\boldsymbol{\epsilon}] = \mathrm{Var}[\boldsymbol{\epsilon}_s] = \sigma_\epsilon^2 \boldsymbol{I}$, $\mathrm{Var}[\boldsymbol{\delta}] = \mathrm{Var}[\boldsymbol{\delta}_s] = \sigma_\delta^2 \boldsymbol{I}$, $\mathrm{Cov}[\boldsymbol{\epsilon}, \boldsymbol{\delta}] = \mathbf{0}$, and $\mathrm{Cov}[\boldsymbol{\epsilon}_s, \boldsymbol{\delta}_s] = \rho_e \boldsymbol{I}$.

In a traditional GWAS, we obtain marginal effect size estimates, $\hat{\boldsymbol{\beta}}_{gwas}$ and $\hat{\boldsymbol{\gamma}}_{gwas}$, as

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{gwas} &= \frac{1}{n_1} \left[\boldsymbol{Y}^\intercal \ \boldsymbol{X}_s^\intercal \right] \begin{bmatrix} \boldsymbol{\phi} \\ \boldsymbol{\phi}_s \end{bmatrix} = \frac{1}{n_1} \left(\boldsymbol{Y}^\intercal \boldsymbol{Y} + \boldsymbol{X}_s^\intercal \boldsymbol{X}_s \right) \boldsymbol{\beta} + \frac{1}{n_1} \left(\boldsymbol{Y}^\intercal \boldsymbol{\epsilon} + \boldsymbol{X}_s^\intercal \boldsymbol{\epsilon}_s \right) \\
\hat{\boldsymbol{\gamma}}_{gwas} &= \frac{1}{n_2} \left[\boldsymbol{Z}^\intercal \ \boldsymbol{X'}_s^\intercal \right] \begin{bmatrix} \boldsymbol{\psi} \\ \boldsymbol{\psi}_s \end{bmatrix} = \frac{1}{n_2} \left(\boldsymbol{Z}^\intercal \boldsymbol{Z} + \boldsymbol{X'}_s^\intercal \boldsymbol{X'}_s \right) \boldsymbol{\gamma} + \frac{1}{n_2} \left(\boldsymbol{Z}^\intercal \boldsymbol{\delta} + \boldsymbol{X'}_s^\intercal \boldsymbol{\delta}_s \right).
\end{aligned} \tag{4}$$

Assuming individuals in both GWASs are drawn from the same population with LD matrix $\boldsymbol{V}$, we have $\hat{\boldsymbol{\beta}}_{gwas} \sim N\left(\boldsymbol{V}\boldsymbol{\beta}, \frac{\sigma_\epsilon^2}{n_1}\boldsymbol{V}\right)$, $\hat{\boldsymbol{\gamma}}_{gwas} \sim N\left(\boldsymbol{V}\boldsymbol{\gamma}, \frac{\sigma_\delta^2}{n_2}\boldsymbol{V}\right)$. We also find

$$\mathrm{Cov}[\hat{\boldsymbol{\beta}}_{gwas}, \hat{\boldsymbol{\gamma}}_{gwas}] = \mathrm{E}[\hat{\boldsymbol{\beta}}_{gwas}\hat{\boldsymbol{\gamma}}_{gwas}^\intercal] - (\boldsymbol{V}\boldsymbol{\beta})(\boldsymbol{V}\boldsymbol{\gamma})^\intercal = \frac{\rho_e}{n_1 n_2} \mathrm{E}[\boldsymbol{X}_s^\intercal \boldsymbol{X'}_s] = \frac{\rho_e n_s}{n_1 n_2} \boldsymbol{V}, \tag{5}$$

where the last equality follows from Isserlis' theorem[40].

Under infinite sample sizes, $\mathrm{Var}[\hat{\boldsymbol{\beta}}_{gwas}] = \mathrm{Var}[\hat{\boldsymbol{\gamma}}_{gwas}] = \mathrm{Cov}[\hat{\boldsymbol{\beta}}_{gwas}, \hat{\boldsymbol{\gamma}}_{gwas}] = \mathbf{0}$, and we have $\boldsymbol{\beta} = \boldsymbol{V}^{-1}\hat{\boldsymbol{\beta}}_{gwas}$, $\boldsymbol{\gamma} = \boldsymbol{V}^{-1}\hat{\boldsymbol{\gamma}}_{gwas}$. Thus, local genetic covariance, $\rho_{g,local}$, can be computed as

$$\rho_{g,local} = (\hat{\boldsymbol{\beta}}_{gwas}^\intercal \boldsymbol{V}^{-1}) \boldsymbol{V} (\boldsymbol{V}^{-1}\hat{\boldsymbol{\gamma}}_{gwas}) = \hat{\boldsymbol{\beta}}_{gwas}^\intercal \boldsymbol{V}^{-1} \hat{\boldsymbol{\gamma}}_{gwas}. \tag{6}$$

However, when sample sizes are finite, from bilinear form theory[41], the covariance between $\hat{\boldsymbol{\beta}}_{gwas}$ and $\hat{\boldsymbol{\gamma}}_{gwas}$ creates bias, resulting in

$$\mathrm{E}[\hat{\boldsymbol{\beta}}_{gwas}^\intercal \boldsymbol{V}^{-1} \hat{\boldsymbol{\gamma}}_{gwas}] = \boldsymbol{\beta}^\intercal \boldsymbol{V} \boldsymbol{\gamma} + \frac{\rho_e}{n_1 n_2} \mathrm{tr}\left(\boldsymbol{V}\right) = \boldsymbol{\beta}^\intercal \boldsymbol{V} \boldsymbol{\gamma} + \frac{p(\rho - \rho_{g,local}) n_s}{n_1 n_2}, \tag{7}$$

Correcting for bias, we arrive at the unbiased estimator

$$\hat{\rho}_{g,local} = \frac{n_1 n_2 \hat{\boldsymbol{\beta}}_{gwas}^\intercal \boldsymbol{V}^{-1} \hat{\boldsymbol{\gamma}}_{gwas} - n_s p \rho}{n_1 n_2 - n_s p}. \tag{8}$$

For rank-deficient LD matrix $\boldsymbol{V}$, one replaces $\boldsymbol{V}^{-1}$ with the pseudo-inverse $(\boldsymbol{V}^\dagger)$ and p with

$q = \text{rank}(\boldsymbol{V})$, yielding the unbiased estimator

$$\hat{\rho}_{g,local} = \frac{n_1 n_2 \hat{\boldsymbol{\beta}}_{gwas}^{\intercal} \boldsymbol{V}^{-1} \boldsymbol{\gamma}_{gwas} - n_s p \rho}{n_1 n_2 - n_s p}. \tag{9}$$

Thus, in order to obtain an unbiased estimate of genetic covariance between a pair of traits, one needs to know their phenotypic correlation. When phenotypic correlation is not available, one can obtain an estimate from genome-wide summary data using the LD score regression equation[14],

$$\text{E}[z_{\phi,j} z_{\psi,j} | l_j] = \frac{\sqrt{n_1 n_2}\rho_g}{p} l_j + \frac{\rho n_s}{\sqrt{n_1 n_2}}, \tag{10}$$

where $z_{\phi,j}$, $z_{\psi,j}$ are the Z-scores of SNP j in the two traits, and l_j the LD score of SNP j.

In the special case when $\hat{\boldsymbol{\beta}}_{gwas}$ and $\hat{\boldsymbol{\gamma}}_{gwas}$ are obtained for the same trait on the same set of individuals (i.e. $\hat{\boldsymbol{\beta}}_{gwas} = \hat{\boldsymbol{\gamma}}_{gwas}$, $n_1 = n_2 = n_s$, $\rho = 1$) Equation (8) reduces to the local SNP-heritability estimator[18]. When $n_s = 0$ (i.e. no shared individuals between the GWASs), the unbiased estimator is simply $\hat{\rho}_{g,local} = \hat{\boldsymbol{\beta}}_{gwas}^{\intercal} \boldsymbol{V}^{-1} \hat{\boldsymbol{\gamma}}_{gwas}$. An interpretation for this simple formula is that in the absence of sample overlap, the covariance in the noise, ϵ and δ, is 0 and does thus not introduce bias into the estimate of $\rho_{g,local}$.

Following bilinear form theory[41], we obtain the variance of $\hat{\rho}_{g,local}$,

$$\text{Var}[\hat{\rho}_{g,local}] = \left(\frac{n_1 n_2}{n_1 n_2 - n_s p}\right)^2 \left[\left(\frac{p \rho_e n_s}{n_1 n_2}\right)^2 + \frac{\sigma_\epsilon^2 \sigma_\delta^2 p}{n_1 n_2} + \frac{\sigma_\delta^2 h_{g\phi,local}^2}{n_2} + \frac{\sigma_\epsilon^2 h_{g\psi,local}^2}{n_1} + 2\frac{n_s \rho_e \rho_{g,local}}{n_1 n_2}\right] \tag{11}$$

For rank deficient LD matrix with $\text{rank}(\boldsymbol{V}) = q$, one replaces p with q in Equation (11).

Accounting for statistical noise in LD estimates

Limited sample size of external reference panels creates statistical noise in the estimated LD matrix that biases our estimates. Following our previous work[18], we apply truncated-SVD regularization to remove noise in external reference LD. We note that $\hat{\boldsymbol{\beta}}_{gwas}^{\intercal} \boldsymbol{V}^{\dagger} \hat{\boldsymbol{\gamma}}_{gwas} = \sum_{i=1}^{q} s_i = \sum_{i=1}^{q} \frac{1}{w_i} (\hat{\boldsymbol{\beta}}_{gwas}^{\intercal} \boldsymbol{u}_i)(\hat{\boldsymbol{\gamma}}_{gwas}^{\intercal} \boldsymbol{u}_i)$, where w_i, $\boldsymbol{u}_i$ are the eigenvalues and eigenvectors of the LD matrix $\boldsymbol{V}$, and $q = \text{rank}(\boldsymbol{V})$. We use $\hat{s}_i = \frac{1}{\hat{w}_i} (\hat{\boldsymbol{\beta}}_{gwas}^{\intercal} \hat{\boldsymbol{u}}_i)(\hat{\boldsymbol{\gamma}}_{gwas}^{\intercal} \hat{\boldsymbol{u}}_i)$, to denote the counterpart obtained from external reference LD matrix $\hat{\boldsymbol{V}}$. We show through simulations that the bulk of $\hat{\boldsymbol{\beta}}_{gwas}^{\intercal} \boldsymbol{V}^{\dagger} \hat{\boldsymbol{\gamma}}_{gwas}$ comes from s_i where $i \ll q$ and that $s_i \approx \hat{s}_i$ for $i \ll q$,

thus justifying truncated-SVD as an appropriate regularization method when only external reference LD ($\hat{\boldsymbol{V}}$) is available.

Let $g(\hat{\boldsymbol{\beta}}_{gwas}, \hat{\boldsymbol{\gamma}}_{gwas}, k) = \sum_{i=1}^{k} \hat{s}_i = \sum_{i=1}^{k} \frac{1}{\hat{w}_i}(\hat{\boldsymbol{\beta}}_{gwas}^{\intercal}\hat{\boldsymbol{u}}_i)(\hat{\boldsymbol{\gamma}}_{gwas}^{\intercal}\hat{\boldsymbol{u}}_i)$, be the truncated-SVD regularized estimates for $\hat{\boldsymbol{\beta}}_{gwas}^{\intercal}\boldsymbol{V}^{\dagger}\hat{\boldsymbol{\gamma}}_{gwas}$, then it can be shown that

$$\mathrm{E}[g(\hat{\boldsymbol{\beta}}_{gwas}, \hat{\boldsymbol{\gamma}}_{gwas}, k)] = \frac{n_s k(\rho - \rho_g)}{n_1 n_2} + \sum_{i=1}^{k} \hat{w}_i(\boldsymbol{\beta}^{\intercal}\hat{\boldsymbol{u}}_i)(\boldsymbol{\gamma}^{\intercal}\hat{\boldsymbol{u}}_i). \tag{12}$$

Assuming $\hat{w}_i = w_i$ and $\hat{\boldsymbol{u}}_i = \boldsymbol{u}_i$ for $i \ll k$, Equation (12) is a biased approximation of $\rho_{g,local}$, with bias $\frac{n_s k(\rho-\rho_g)}{n_1 n_2}$. Correcting for the bias, we arrive at the estimator

$$\hat{\rho}_{g,local} = \frac{n_1 n_2 g(\hat{\boldsymbol{\beta}}_{gwas}, \hat{\boldsymbol{\gamma}}_{gwas}, k) - n_s \rho k}{n_1 n_2 - n_s k}, \tag{13}$$

which has variance

$$\mathrm{Var}[\hat{\rho}_{g,local}] = \left(\frac{n_1 n_2}{n_1 n_2 - n_s k}\right)^2 \left[\left(\frac{k\rho_e n_s}{n_1 n_2}\right)^2 + \frac{\sigma_\epsilon^2 \sigma_\delta^2 k}{n_1 n_2} + \frac{\sigma_\delta^2 h_{g\phi,local}^2}{n_2} + \frac{\sigma_\epsilon^2 h_{g\psi,local}^2}{n_1} + 2\frac{n_s \rho_e \rho_{g,local}}{n_1 n_2}\right] \tag{14}$$

Extension to multiple independent loci

For genome partitioned into m loci, let

$$\begin{aligned} \phi &= \boldsymbol{x}_1^{\intercal}\boldsymbol{\beta}_1 + \cdots + \boldsymbol{x}_m^{\intercal}\boldsymbol{\beta}_m + \epsilon \\ \psi &= \boldsymbol{x}_1^{\intercal}\boldsymbol{\gamma}_1 + \cdots + \boldsymbol{x}_m^{\intercal}\boldsymbol{\gamma}_m + \delta, \end{aligned} \tag{15}$$

denote the phenotype measurements of two traits at an individuals, where we assume that SNPs in different pairs of loci are independent, i.e. $\mathrm{E}[\boldsymbol{x}_{ik}\boldsymbol{x}_{il}] = 0$ for all $i \neq j$, $k \in \{1, \cdots, p_i\}$, and $l \in \{1, \cdots, p_j\}$, where p_i and p_j are the number of SNPs in locus i and j. Under these assumptions, we decompose the phenotypic covariance, ρ, between ϕ and ψ,

$$\rho = \mathrm{E}[(\boldsymbol{x}_1^{\intercal}\boldsymbol{\beta}_1 + \cdots + \boldsymbol{x}_m^{\intercal}\boldsymbol{\beta}_m + \epsilon)(\boldsymbol{x}_1^{\intercal}\boldsymbol{\gamma}_1 + \cdots + \boldsymbol{x}_m^{\intercal}\boldsymbol{\gamma}_m + \delta)^{\intercal}] = \sum_{i=1}^{m} \boldsymbol{\beta}_i^{\intercal}\boldsymbol{V}_i\boldsymbol{\gamma}_i + \rho_e, \tag{16}$$

where $\rho_{g,local,i} = \boldsymbol{\beta}_i^\intercal \boldsymbol{V}_i \boldsymbol{\gamma}_i$ is the genetic covariance between the pair of traits attributed to genetic variants at locus i. Following strategies outlined in previous sections, we arrive at the estimator for genetic covariance at the i-th locus,

$$\hat{\rho}_{g,local,i} = \frac{n_1 n_2 g(\hat{\boldsymbol{\beta}}_{gwas,i}, \hat{\boldsymbol{\gamma}}_{gwas,i}, k) - n_s(\rho - \sum_{j=1,j\neq i}^{m} \hat{\rho}_{g,local,j}) k_i}{n_1 n_2 - n_s k_i}, \tag{17}$$

which defines a system of linear equation involving m unknown variables and m equations. In the special case where there is no sample overlap, $n_s = 0$, and $\hat{\rho}_{g,local,i}$ reduces to $g(\hat{\boldsymbol{\beta}}_{gwas}, \hat{\boldsymbol{\gamma}}_{gwas}, k)$, and can be estimated independent of all other windows. Following bilinear form theory, we obtain variance estimate for $\hat{\rho}_{g,local,i}$ as,

$$\begin{aligned}\mathrm{Var}[\hat{\rho}_{g,local,i}] = \left(\frac{n_1 n_2}{n_1 n_2 - n_s k_i}\right)^2 \left[\left(\frac{k_i \rho_e n_s}{n_1 n_2}\right)^2 + \frac{\sigma_\epsilon^2 \sigma_\delta^2 k_i}{n_1 n_2} + \frac{\sigma_\delta^2 h_{g\phi,local,i}^2}{n_2} + \frac{\sigma_\epsilon^2 h_{g\psi,local,i}^2}{n_1} + 2\frac{n_s \rho_e \rho_{g,local,i}}{n_1 n_2}\right] \\ + \sum_{j=1,j\neq i}^{m} \left(\frac{n_s k_j}{n_1 n_2 - n_s k_i}\right)^2 \mathrm{Var}[\hat{\rho}_{g,local,j}]\end{aligned} \tag{18}$$

which also defines a system of linear equations with m equations and m variables.

When $k_1 = \cdots = k_m = k$, i.e. all loci use the same number of eigenvectors in the truncated-SVD regularization, summing over i on both sides of Equation (17) yields

$$\begin{aligned}\hat{\rho}_g = \sum_{i=1} \hat{\rho}_{g,local,i} &= \frac{n_1 n_2}{n_1 n_2 - n_s k} \sum_{i=1}^{m} g(\hat{\boldsymbol{\beta}}_{gwas,i}, \hat{\boldsymbol{\gamma}}_{gwas,i}, k) - \frac{k n_s}{n_1 n_2 - n_s k} \sum_{i=1}^{m} \left(r - \sum_{j=1,j\neq i}^{m} \hat{\rho}_{g,local,j}\right) \\ &= \frac{n_1 n_2}{n_1 n_2 - n_s k} \sum_{i=1}^{m} g(\hat{\boldsymbol{\beta}}_{gwas,i}, \hat{\boldsymbol{\gamma}}_{gwas,i}, k) - \frac{k n_s}{n_1 n_2 - n_s k} \sum_{i=1}^{m} (\rho - \hat{\rho}_g + \hat{\rho}_{g,local,i}) \\ &= \frac{n_1 n_2}{n_1 n_2 - n_s k} \sum_{i=1}^{m} g(\hat{\boldsymbol{\beta}}_{gwas,i}, \hat{\boldsymbol{\gamma}}_{gwas,i}, k) + \frac{k n_s m - k n_s}{n_1 n_2 - n_s k} \hat{\rho}_g - \frac{k n_s m \rho}{n_1 n_2 - n_s k}.\end{aligned} \tag{19}$$

Solving for $\hat{\rho}_g$ yields

$$\hat{\rho}_g = \frac{n_1 n_2 \sum_{i=1}^{m} g(\hat{\boldsymbol{\beta}}_{gwas,i}, \hat{\boldsymbol{\gamma}}_{gwas,i}, k) - k n_s m \rho}{n_1 n_2 - k n_s m}, \tag{20}$$

which has variance

$$\mathrm{Var}[\hat{\rho}_g] = \left(\frac{n_1 n_2}{n_1 n_2 - k n_s m}\right)^2 \sum_{i=1}^{m} \mathrm{Var}[g(\hat{\boldsymbol{\beta}}_{gwas,i}, \hat{\boldsymbol{\gamma}}_{gwas,i}, k)]. \tag{21}$$

Thus, if k is chosen such that $(n_1 n_2 - k n_s m)$ is small (i.e. $\frac{n_1 n_2}{n_1 n_2 - k n_s m}$ large), the estimate of total genetic covariance will have large standard error. To reduce standard error in the estimates (at the cost of some bias), we recommend choosing k such that $\frac{n_1 n_2}{n_1 n_2 - k n_s m}$ is less than 2. When testing for statistical significance, we assume that the estimates of local and genome-wide genetic covariance and correlation follow a normal distribution.

Standardizing local genetic covariance

Standardizing local genetic covariance between a pair of traits by the square roots of the local heritability of the two traits yields the local genetic correlation ($r_{g,local}$). Whereas genetic covariance takes into account the magnitude of the effect sizes, genetic correlation provides a measure of similarity between the effects of SNPs on traits comparable across different magnitudes of effect sizes. To estimate local genetic correlation for the i-th locus, we apply the formula

$$\hat{r}_{g,local,i} = \frac{\hat{\rho}_{g,local,i}}{\sqrt{\hat{h}^2_{g\phi,local,i}}\sqrt{\hat{h}^2_{g\psi,local,i}}}, \tag{22}$$

where $\hat{h}^2_{g\phi,local,i}$ and $\hat{h}^2_{g\psi,local,i}$ denote the local SNP-heritability of trait ϕ and ψ at the i-th locus. In simulations, we show that $\hat{r}_{g,local,i}$ is approximately unbiased when both traits are heritable at the i-th locus. In practice, however, the terms, $\hat{h}^2_{g\phi,local,i}$ and $\hat{h}^2_{g\psi,local,i}$, can be close to zero, greatly inflating the standard error of $\hat{r}_{g,local,i}$. Thus, we recommend estimating local genetic correlation only at loci with significant local SNP-heritability. One can also estimate local genetic correlation at a set of loci. For example, to estimate genetic correlation at loci indexed by the index set $\boldsymbol{C}$, one applies the following formula,

$$\hat{r}_{g,\boldsymbol{C}} = \frac{\sum_{i\in\boldsymbol{C}} \hat{\rho}_{g,local,i}}{\sqrt{\sum_{i\in\boldsymbol{C}} \hat{h}^2_{\phi,g,local,i}}\sqrt{\sum_{i\in\boldsymbol{C}} \hat{h}^2_{\psi,g,local,i}}}, \tag{23}$$

We estimate standard error of local genetic correlation at a single locus through a parametric bootstrap approach[42] and local genetic correlation at a set of loci through jackknife.

Simulation framework

Starting from half (202 individuals) of the EUR reference panel from the 1000 Genomes Project[22], we simulated genotype data for 50,000 individuals at HapMap3[43] SNPs with minor allele frequency (MAF) greater than 5% in a 1.5Mb locus on chromosome 1 using HAPGEN2[43]. We used the other half of the EUR reference panel (203 individuals) to obtain external reference LD matrices.

We simulated phenotypes from the genotypes according to the linear model $\boldsymbol{\phi} = \boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\epsilon}$ and $\boldsymbol{\psi} = \boldsymbol{X}\boldsymbol{\gamma}+\boldsymbol{\delta}$, where $\boldsymbol{X}$ is the column-standardized genotype matrix. We drew the effects of causal SNPs ($\boldsymbol{\beta}_C$, $\boldsymbol{\gamma}_C$) from the distribution

$$N\left(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}\frac{h^2_{g\phi}}{|C|}\boldsymbol{I} & \frac{\rho_e}{|C|}\boldsymbol{I}\\ \frac{\rho_e}{|C|}\boldsymbol{I} & \frac{h^2_{g\psi}}{|C|}\boldsymbol{I}\end{bmatrix}\right), \tag{24}$$

where C is the index set of causal SNPs, and set the effects of all other SNPs to be zero. Here, for the convenience of simulation, we assume both traits have the same causal SNPs, although this assumption doesn't need to hold for ρ-HESS to be unbiased. We then drew ($\boldsymbol{\epsilon}$, $\boldsymbol{\delta}$) from the distribution

$$N\left(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}(1-h^2_{g\phi})\boldsymbol{I} & \rho_e\boldsymbol{I}\\ \rho_e\boldsymbol{I} & (1-h^2_{g\psi})\boldsymbol{I}\end{bmatrix}\right). \tag{25}$$

Finally, we simulated GWAS summary statistics using methods outlined in previous sections. For each $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ drawn from the normal distribution, we simulated 500 sets of summary statistics by varying $\boldsymbol{\epsilon}$ and $\boldsymbol{\delta}$, and applied ρ-HESS to estimate genetic covariance and genetic correlation for each set of the simulated summary statistics.

Inferring direction of causality

We systematically search for plausible causal directions (see example 1 and 2 in Figure 9) across all 234 pair of traits that show significant genome-wide correlation. First, we test whether the 95% confidence intervals, defined as 1.96 times standard error on each side of the estimates, of $\hat{r}_{g,local,trait1}$ and $\hat{r}_{g,local,trait2}$ overlap. Then we test whether one of the confidence intervals overlaps with 0 and the other does not. Since we employed jackknife to estimate standard errors, for the robustness of the standard error, we restricted our bi-directional analyses to pairs of traits for which the number of loci ascertained for GWAS risk variants specific to both traits is greater than 10. We mark pairs of trait, for which the confidence intervals of local genetic correlations do not overlap, as pairs of traits having putative causal relationships, and designate the trait, for which the local genetic correlation is significantly non-zero, as the causal trait.

Empirical data sets

We obtained GWAS summary data for 35 complex traits and diseases from 11 GWAS consortia (see Table 1), all of which are based on individuals of European ancestry, and have sample size greater than 20,000. We used approximately independent loci defined in[21] to partition the genome, and restricted our analyses on HapMap3 SNPs with minor allele frequency (MAF) greater than 5% in the European population in the 1000 Genomes data[22]. We also removed stand-ambiguous SNPs prior to our analyses. We follow the method outlined in[18] to estimate and re-inflate λ_{gc}, and to choose the number of eigenvectors to include in estimating local genetic covariance and SNP-heritability. We note that although adjustment factors are needed to correct for ascertainment bias in the estimation of local SNP-heritability and local genetic covariance[18], they are not needed in the estimation of local genetic correlation[14].

Trait Name	**Abbreviation**	**Consortium**	**# gen corr [a] within consortium**	**# gen corr [b] outside consortium**	**Approx. sample size**
Body Mass Index[44]	BMI	GIANT	23 (16)	19 (13)	231K
Height[45]	HEIGHT	GIANT	17 (2)	13 (0)	241K
Hip Circumference[46]	HIP	GIANT	21 (14)	17 (10)	144K
Waist-to-hip Ratio[46]	WHR	GIANT	22 (16)	18 (13)	143K
Waist Circumference[46]	WC	GIANT	23 (17)	19 (13)	153K
Haemoglobin[30]	HB	HaemGen	15 (9)	12 (8)	51K
Mean Cell Haemoglobin[30]	MCH	HaemGen	7 (1)	6 (1)	44K
MCH Concentration[30]	MCHC	HaemGen	6 (4)	1 (1)	47K
Mean Cell Volume[30]	MCV	HaemGen	10 (3)	8 (1)	49K
Packed Cell Volume[30]	PCV	HaemGen	15 (10)	11 (8)	45K
Red Blood Cell Count[30]	RBC	HaemGen	17 (10)	14 (8)	46K
Number of Platelets[47]	PLT	HaemGen	10 (1)	6 (1)	67K
Fasting Glucose[48]	FG	MAGIC	18 (10)	15 (9)	46K
Fasting Insulin[48]	FI	MAGIC	18 (12)	16 (12)	46K
HBA1C[49]	HBA1C	MAGIC	18 (14)	16 (13)	46K
HOMA-B[48]	HOMA-B	MAGIC	16 (9)	13 (9)	46K
HOMA-IR[48]	HOMA-IR	MAGIC	16 (13)	16 (13)	46K
High Density Lipoprotein[26]	HDL	GLGC	18 (11)	16 (10)	96K
Low Density Lipoprotein[26]	LDL	GLGC	15 (6)	13 (4)	91K
Total Cholesterol[26]	TC	GLGC	15 (4)	12 (2)	96K
Triglycerides[26]	TG	GLGC	22 (13)	19 (10)	92K
Education Years[50]	EY	SSGAC	25 (8)	22 (6)	294K
Depressive Symptoms[51]	DS	SSGAC	10 (4)	7 (1)	161K
Neuroticism[51]	NEURO	SSGAC	5 (3)	2 (0)	171K
Subjective Well-being[51]	SWB	SSGAC	7 (2)	4 (0)	298K
Forearm BMD[52]	FA	GEFOS	3 (1)	1 (0)	53K
Femoral Neck BMD[52]	FN	GEFOS	4 (2)	2 (0)	53K
Lumbar Spine BMD[52]	LS	GEFOS	5 (1)	3 (0)	53K
Age at Menarche[53]	AM	ReproGen	17 (3)	17 (3)	133K
Rheumatoid Arthritis[54]	RA	RACI and GARNET	3 (0)	3 (0)	58K
Schizophrnia[55]	SCZ	PGC	13 (4)	13 (4)	75K
Crohn's Disease[56]	CD	IIBD	5 (4)	3 (2)	52K
Inflammatory Bowel Disease[56]	IBD	IIBD	5 (4)	3 (2)	66K
Ulcerative Colitis[56]	UC	IIBD	4 (4)	2 (2)	48K
Type 2 Diabetes[57]	T2D	DIAGRAM	20 (13)	20 (13)	62K

Table 1: A summary of the 35 GWAS summary data sets analyzed. [a]Total number of traits with significant non-zero genome-wide genetic correlation (two-tailed $p < 0.05/595$). [b]Total number of traits outside the consortium with significant non-zero genome-wide genetic correlation. Number of traits for which the magnitude of genetic correlation is both significantly non-zero and greater than 0.2 is shown in parentheses.

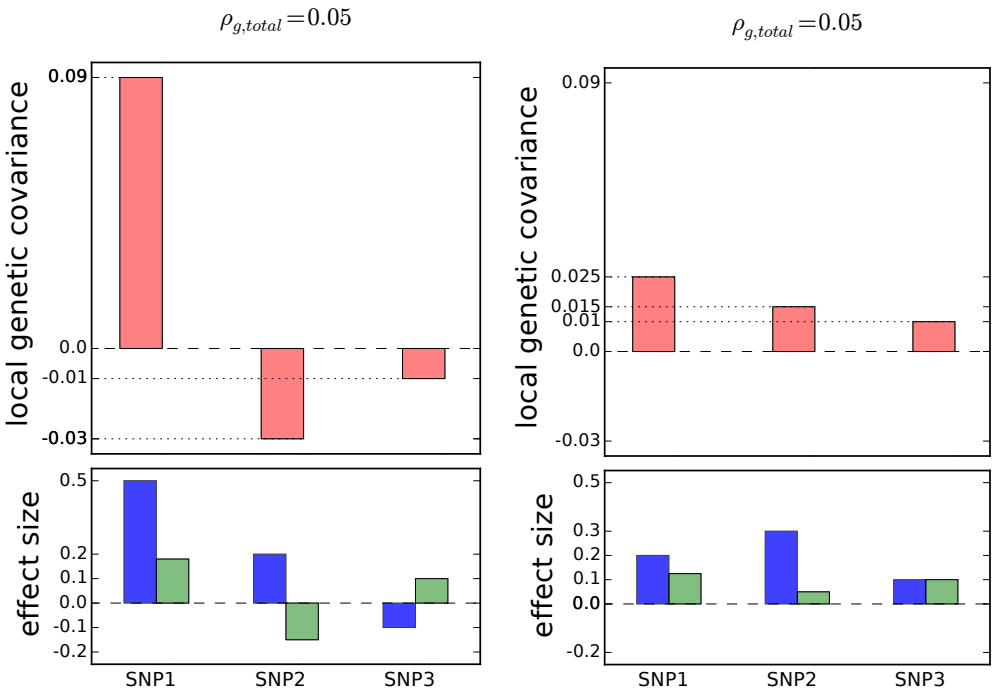

Figure 1: Examples of two different distributions of local genetic covariances (with effect sizes shown at the bottom) that result in the same total genetic covariance ($\rho_{g,total} = 0.05$). In the left example, the total genetic covariance is a summation of a large positive local genetic covariance at SNP1 and two smaller negative local genetic covariances at SNP2 and SNP3 (e.g, SNPs 2 and 3 impact traits through a different pathway than SNP1). In the right side the total genetic covariance is a summation of small positive local genetic covariances (e.g., all three SNPs impact both traits through the same pathway). Positive local genetic covariance can be interpreted as a locus driving a pathway that regulates two traits in the same direction, and negative local genetic covariance the opposite direction.

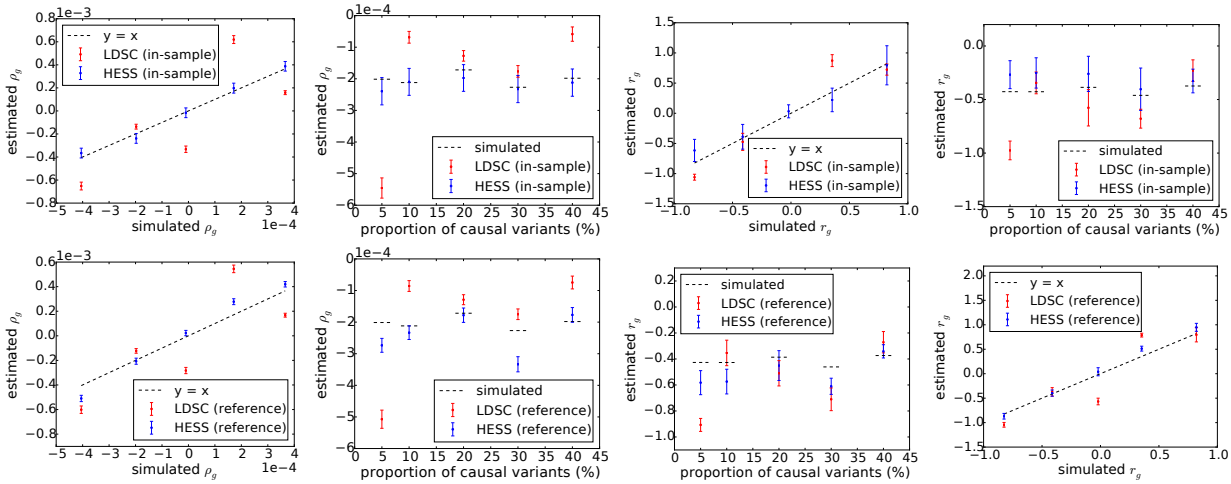

Figure 2: ρ-HESS provides unbiased estimates of covariance when in-sample LD is available (top left) and nearly unbiased estimates of correlation/covariance when LD is estimated from a reference panels (bottom and top right) Mean and standard errors are computed based on 500 simulations. Error bars represent 1.96 times the standard error on each side.

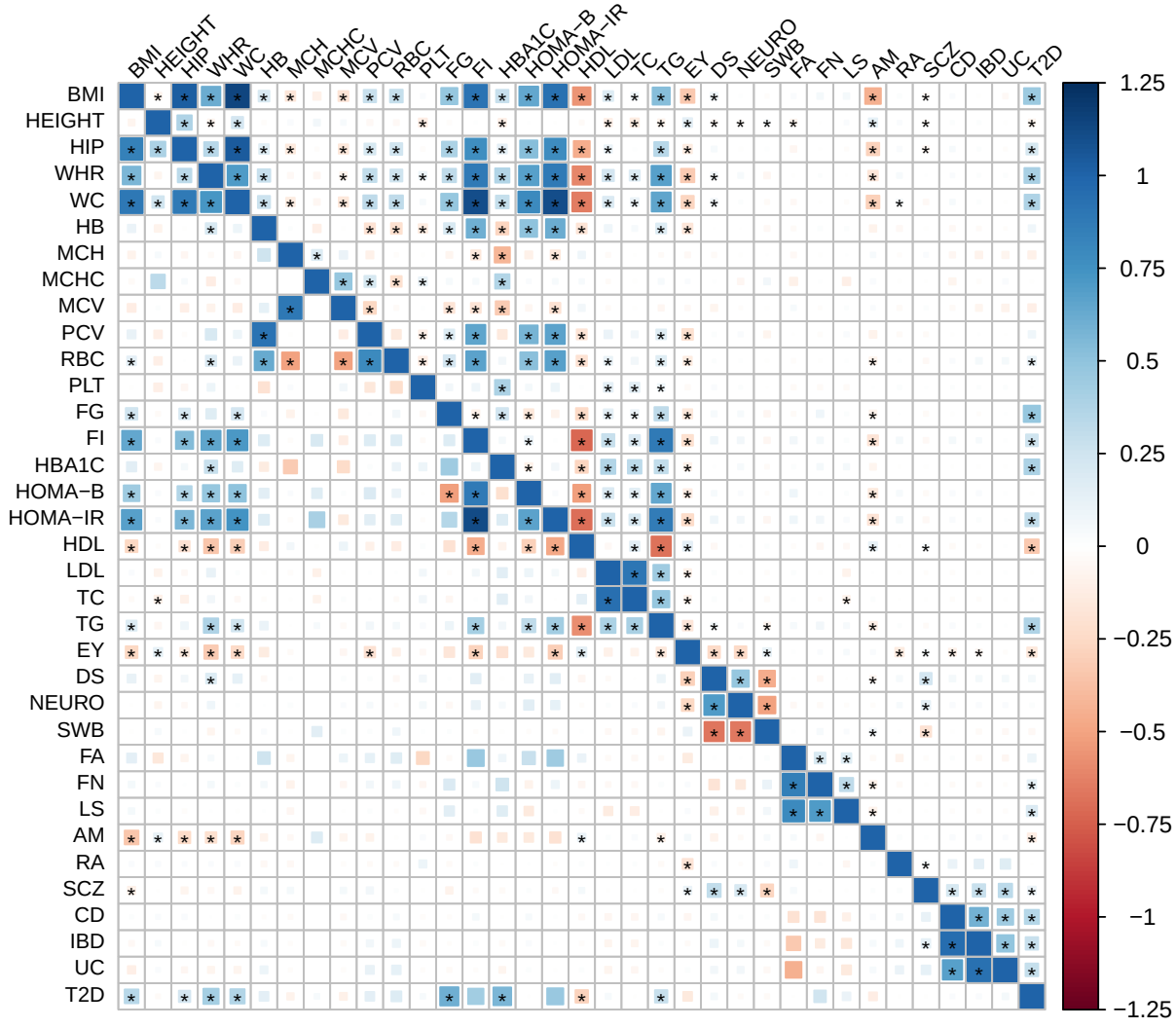

Figure 3: Genetic correlation across the 35 complex traits obtained by ρ-HESS (top half) and cross-trait LDSC[14] (bottom half). The magnitude of the correlation is represented by the color and the size of the square. Among the 595 pairs of traits, ρ-HESS (LDSC) identified 234 (99) pairs showing significant genetic correlation (marked with dots)

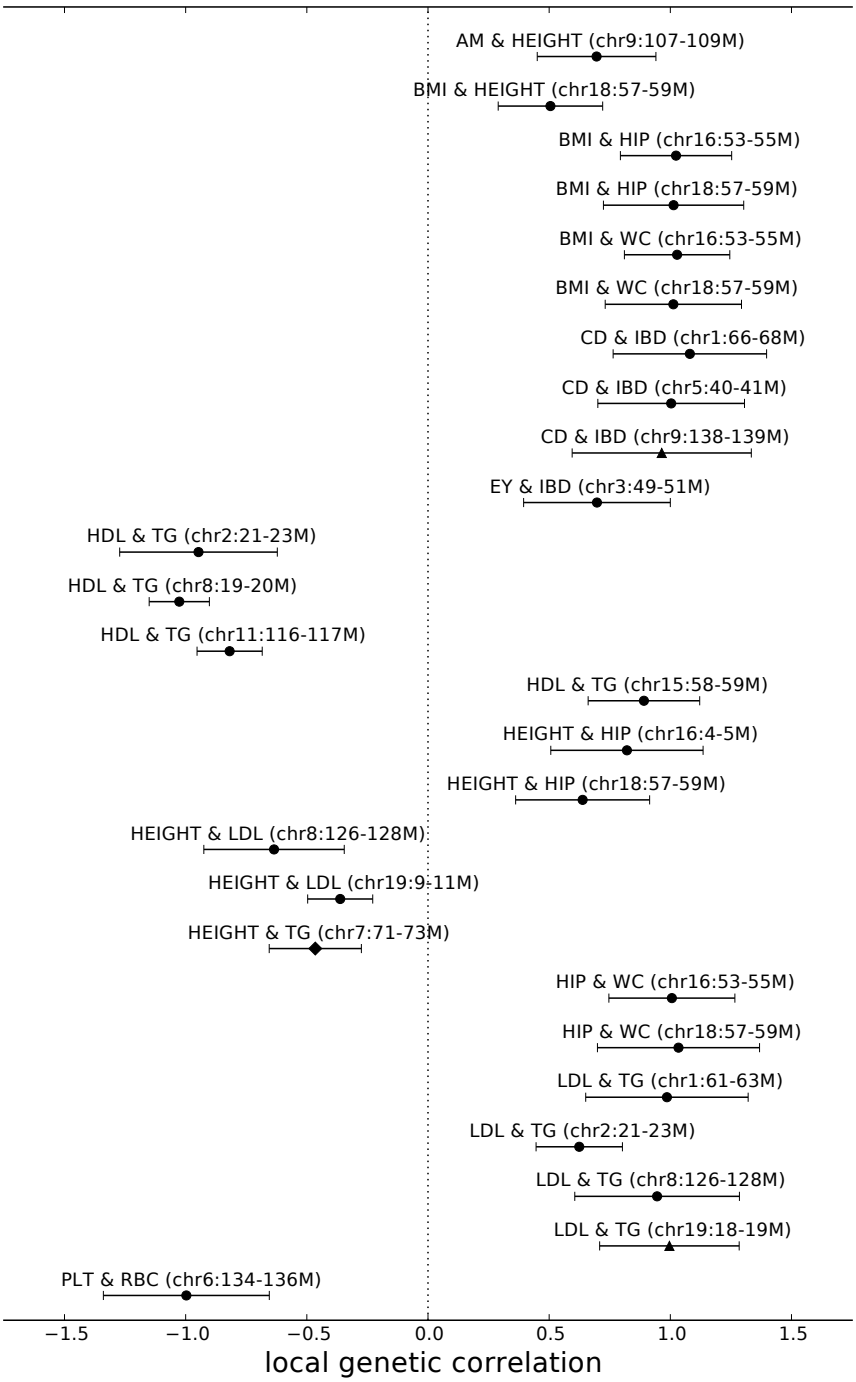

Figure 4: Local genetic correlation at loci displaying significant local genetic covariance and SNP-heritability for pairs of traits with significant genome-wide genetic correlation. To simplify presentation, we excluded all pairs of traits involving TC (see Supplementary Figure 9). We obtain standard error estimates through parametric bootstrapping. Error bars represent 1.96 times the standard error on both sides. Here, triangle represents loci that lack GWAS risk variant for both traits; diamond represents loci that harbor GWAS risk variants for one of the traits; and circle represents loci that contain GWAS risk variants for both traits.

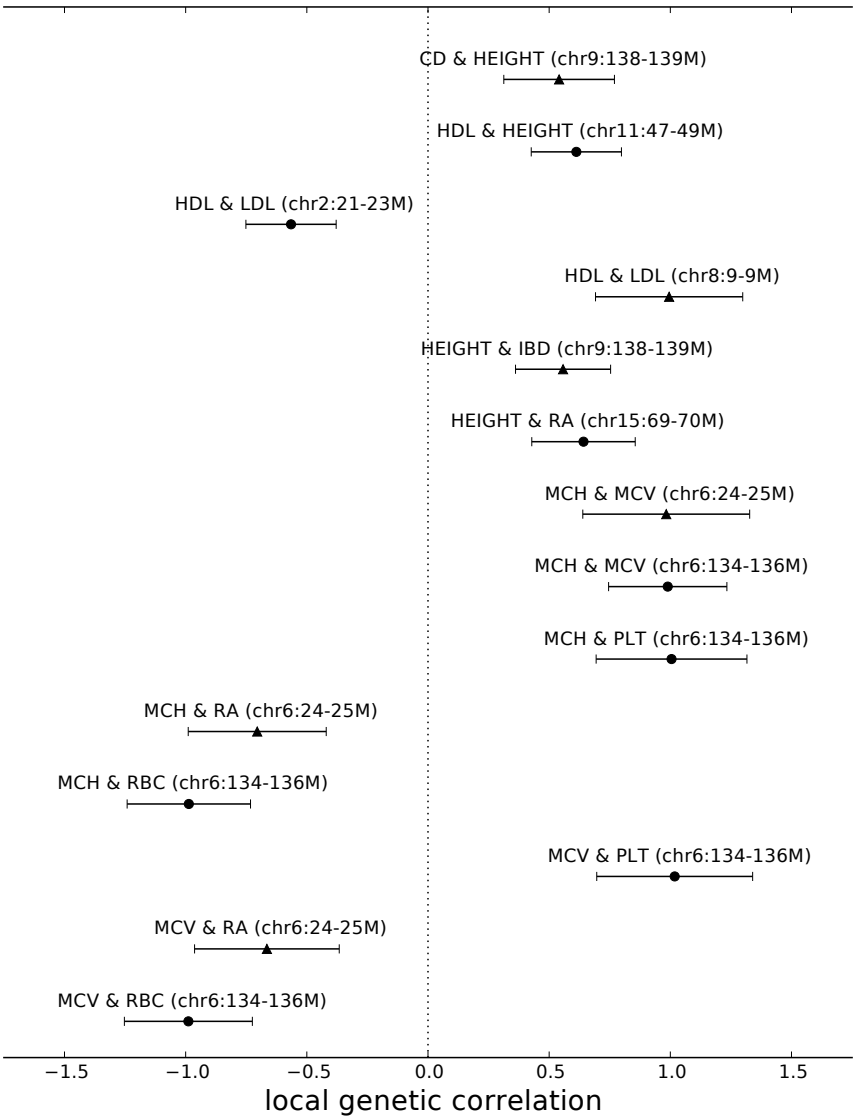

Figure 5: Local genetic correlation at loci displaying significant local genetic covariance and SNP-heritability for pairs of traits without significant genome-wide genetic correlation. We obtain standard error estimates through parametric bootstrapping. Error bars represent 1.96 times the standard error on both sides. Here, triangle represents loci that lack GWAS risk variant for both traits; diamond represents loci that harbor GWAS risk variants for one of the traits; and circle represents loci that contain GWAS risk variants for both traits.

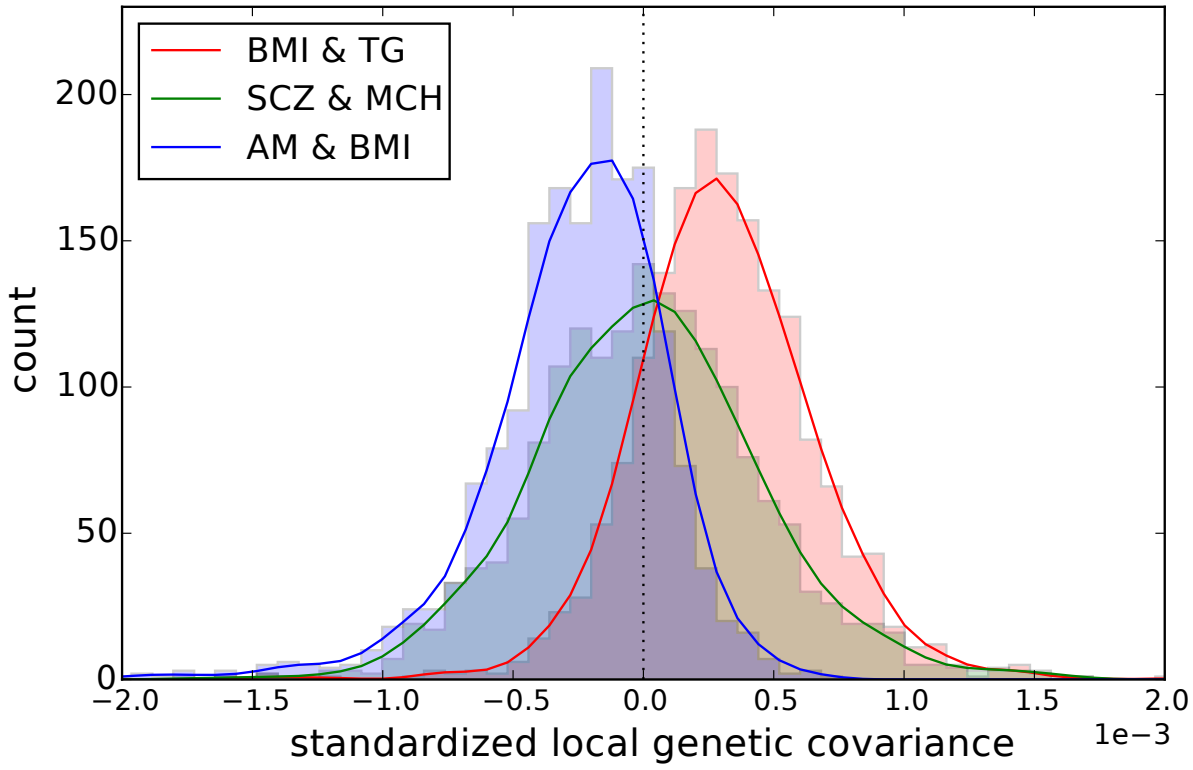

Figure 6: Distribution of standardized local genetic covariance (local genetic covariance standardized by the square roots of heritability of two traits) for the pairs of traits BMI and TG, SCZ and MCH, AM and BMI. Pairs of traits with positive (negative) genome-wide genetic correlation show a shift in the distribution of standardized local genetic covariance away from 0.

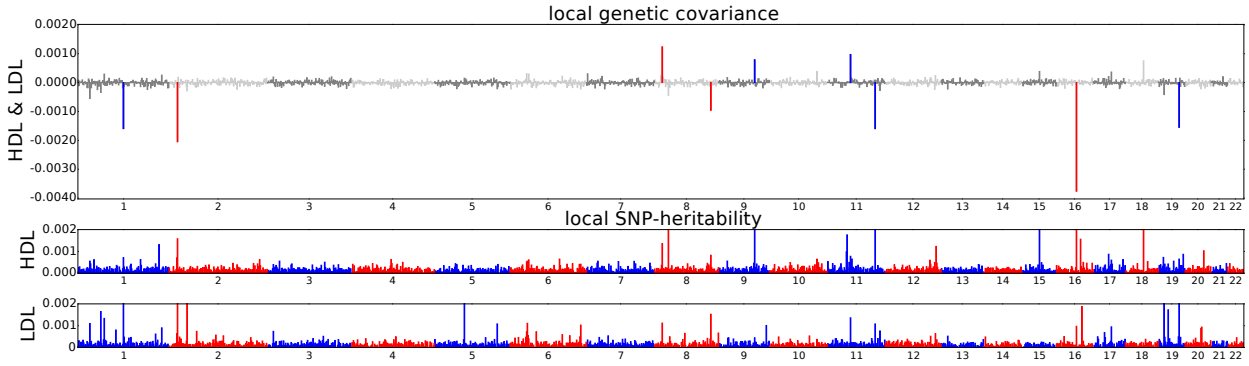

Figure 7: Manhattan-style plots showing the estimates of local genetic covariance for the pairs of traits HDL and LDL. Although the genome-wide genetic correlation between HDL and LDL does not reach the significance level ($p < 0.05/595$), 9 loci exhibit significant local genetic covariance.

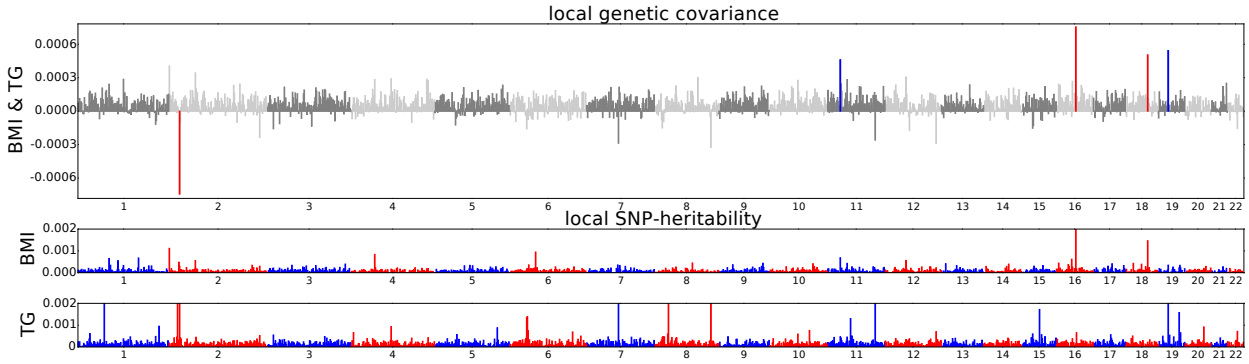

Figure 8: Manhattan-style plots showing the estimates of local genetic covariance for the pairs of traits BMI and TG. That the local genetic covariance between BMI and TG is mostly one-sided implies plausible causal relationship between the two traits

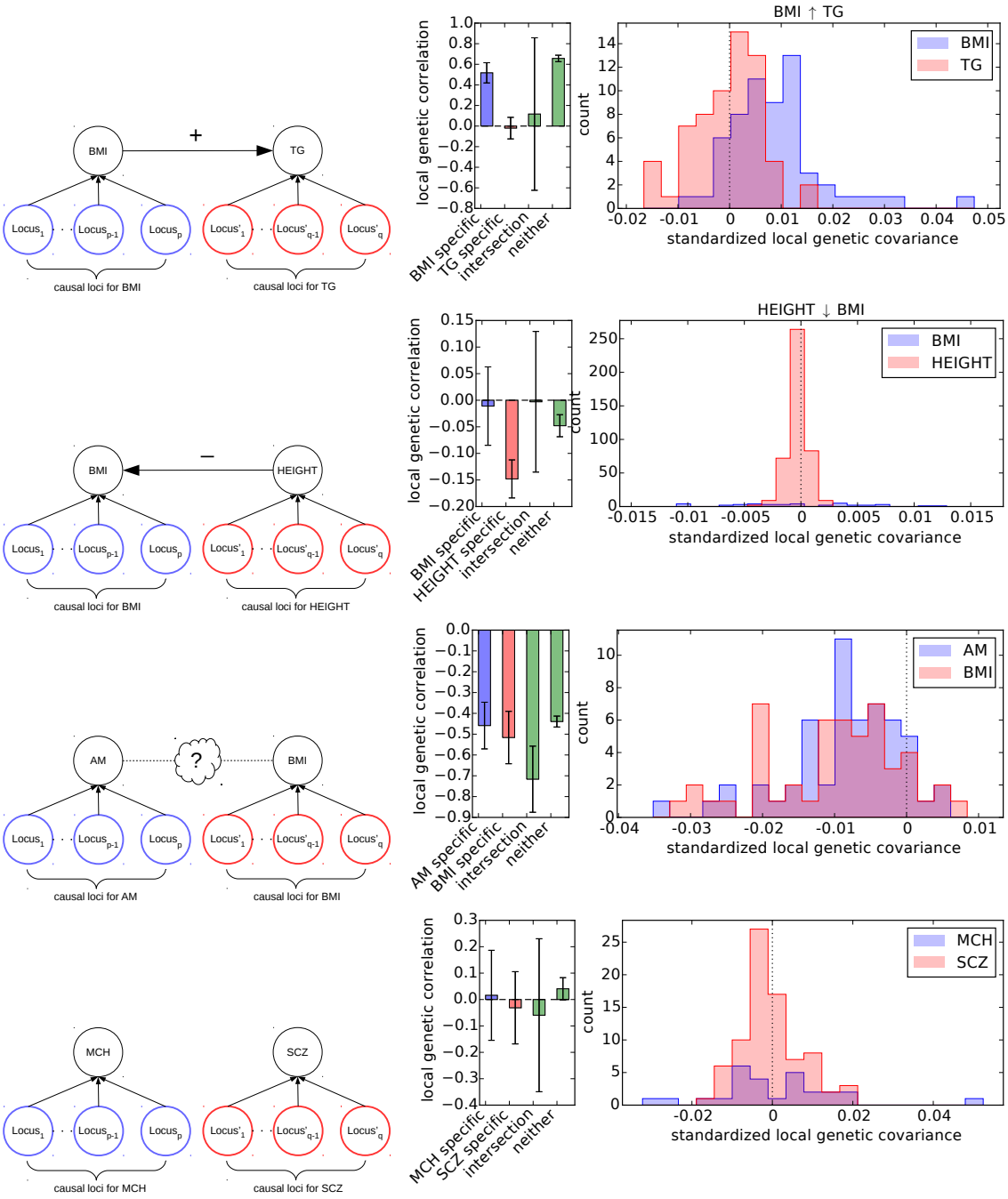

Figure 9: Putative causal relationships between pairs of traits. Example 1: BMI causally increases TG. Local genetic correlation estimate ascertained for loci harboring BMI risk variants is significantly greater than 0. Example 2: Height causally decreases BMI. Local genetic correlation estimate ascertained for loci harboring height risk variants is significantly less than 0. Example 3: More complicated relationships exist between AM and BMI. Local genetic correlation estimates ascertained for loci harboring risk variants for both traits are significantly less than 0. Example 4: Evidence does not support a model in which there is any causal relationship between MCH and schizophrenia (SCZ).

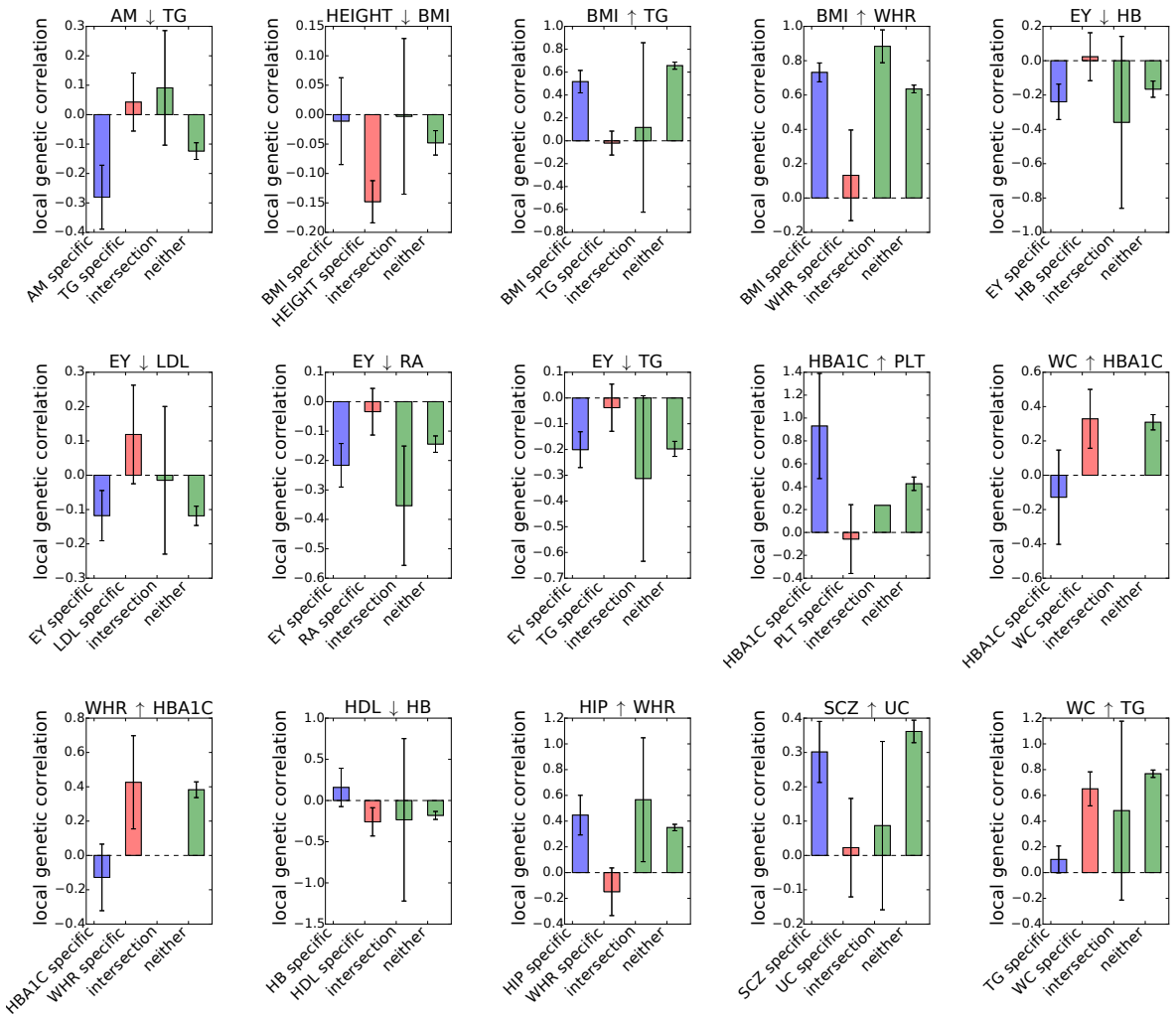

Figure 10: Estimates of local genetic correlation at loci ascertained for GWAS risk variants for each one of 15 pairs of traits that show plausible causal relationship. We obtained standard error using a jackknife approach. Error bars represent 1.96 times the standard error on each side. Here, "↑" means the trait on the left hand side may causally increase the trait on right hand side, whereas "↓" means the trait on the left hand side may causally decrease the trait on the right hand side.