Complex Traits Heritability is Highly Clustered in the eQTL Bipartite Network

Katherine Stone[1,2], **John Platig**[3,4,5], **John Quackenbush**[1,2,6], **and Maud Fagny**[1,2,7]

[1]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, United States
[2]Department of Data Science and Center for Cancer Computational Biology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA.
[3]Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia, USA.
[4]Department of Public Health Sciences, University of Virginia, Charlottesville, Virginia, USA.
[5]Department of Biomedical Engineering, University of Virginia, Charlottesville, Virginia, USA.
[6]Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, Massachusetts, United States
[7]Université Paris-Saclay, INRAE, CNRS, AgroParisTech, Genetique Quantitative et Evolution - Le Moulon, Gif-sur-Yvette 91190 France

Single Nucleotide Polymorphisms (SNPs) associated with traits typically explain a small part of the trait genetic heritability—with the remainder thought to be distributed throughout the genome. Such SNPs are likely to alter expression levels of biologically relevant genes. Expression Quantitative Trait Locus (eQTL) networks analysis has helped to functionally characterize such variants. We systematically analyze the distribution of SNP heritability for ten traits across 29 tissue-specific eQTL networks. We find that heritability is clustered in a small number or tissue-specific, functionally relevant SNP-gene modules and that the greatest occurs in local "hubs" that are both the cornerstone of the network's modules and tissue-specific regulatory elements. The network structure could thus both amplify the genotype-phenotype connection and buffer the deleterious effect of the genetic variations on other traits. Together, these results define a conceptual framework for understanding complex trait architecture and identifying key mutations carrying most of the heritability.

GTEx | eQTL | Bipartite Networks | GWAS | Heritability | Complex Traits
Correspondence: ***maud.fagny@inrae.fr***

DRAFT

Introduction

Many recent studies have shown that SNPs associated with complex traits and diseases in genome-wide association studies co-localize with SNPs associated with the expression level of nearby genes (expression quantitative trait loci, eQTLs)and thus likely to play important regulatory role, potentially influencing traits by modifying gene expression as *cis*-acting eQTLs (1, 2). This is consistent with the observation that more than 90% of the SNPs identified by GWAS are localized outside of coding regions (3, 4). Partitioning heritability for these traits across various annotations while correcting for linkage disequilibrium has confirmed that *cis*-eQTLs as a group explain more of complex trait heritability than would be expected by chance (5–7) but still fail to explain a majority of disease heritability (8). Others have shown that including *trans*-eQTLs may capture additional heritability and explain important pathologic mechanisms (9), but this is rarely done because *trans*-eQTL studies are generally underpowered.

Given the number of loci and the number of potential interactions involved, gene regulatory networks provide an efficient tool for understanding the biological interactions behind trait heritability(10). We have proposed a way to integrate both *cis*- and *trans*-eQTL results using a bipartite eQTL network representation (11, 12), that relies on summary statistics from eQTL studies and is relatively insensitive to a high false discovery rate, allowing one to partially compensate for the lack of power of *trans*-eQTL studies; an update to this method allowed us to weigh SNP-gene regulatory relationships by eQTL effect sizes (13). These analyses have shown that highly structured eQTL networks can reliably identify, in a tissue-specific manner, the biological functions disrupted by traits-associated SNPs (11, 14), and further defined network topological features that are useful in explaining in part the link between genotype and phenotype in complex traits.

In this study, we build on eQTL networks to investigate how complex trait heritability is spread within the tissue-specific eQTL networks to better understand the genetic architecture of complex traits. We used GWAS summary statistics from ten complex traits and diseases, and built 29 tissue-specific *cis*- and *trans*-eQTL networks using the GTEx dataset. We then partitioned heritability across various features, including network node topological summary statistics, to identify key determinants of trait heritability. Finally, we investigated whether heritability for each trait is clustered in particular subparts (regulatory modules, also sometimes referred to as communities) of the eQTL networks, and identified the biological functions explaining most of the heritability of each trait. Importantly, we found that heritability is not scattered uniformly across the genome but rather "clustered" in eQTL modules that represent trait-relevant functions.

Results

Local and global hubs carry a large part of trait heritability. We considered a total of ten traits and diseases (see Supplementary Tab. S1) presenting varying levels of polygenic determinism and estimated genetic heritability. For each trait or disease, we obtained summary statistics including SNP chromosome, position, allele 1 and 2, χ^2, and Z-score. GWAS data were obtained from https://alkesgroup.broadinstitute.org/LDSCORE/all_sumstats/. We wanted to understand how heritability is distributed across the network and whether there are particular eQTL

network topological features that correlate with a greater-than-expected ability to explain the heritability of a particular trait. We investigated ten polygenic traits or diseases (see Methods and Supplementary Tab S1) chosen for their medical or evolutionary relevance in populations of European descent. We conditioned on European descent because of the demographics of GTEx. Breast Cancer (BRC), Ovarian Cancer (OVC) and Prostate Cancer (PRC), Alzheimer's Disease (ALZ), Multiple Sclerosis (MS), Schizophrenia (SCZ), high HDL levels (HDL), and Type 2 diabetes (TIID; diabetes) are all increasing in populations of European descent and are known to exhibit partial genetic heritability. Smoking Cessation (SMC) is a measure of smoking dependency that is partly explained by genetic factors and is linked to a host of pulmonary and other diseases. Schizophrenia (SCZ) is a highly heritable but poorly understood polygenic disease. Finally, Height (HGT) is the canonical example of a polygenic trait, and some consider it a natural selection target in populations of European descent. These traits and diseases also represent a wide range of global genetic heritability, as reported in the literature, from 25% for type 2 diabetes, to up to 80-90% for schizophrenia and Alzheimer's. A complete description of these traits and diseases and their GWAS summary statistics is reported in the Supplementary Tab. S1.

We used RNA-seq and genotyping data from GTEx to perform eQTL analyses and built 29 tissue-specific weighted bipartite eQTL networks and identified highly connected modules within each network based on CONDOR's bipartite modularity maximization (12) (see Methods, Supplementary Text and Supplementary Fig. S1 and Tab. S2). We then computed two different summary statistics characterizing the topological properties of SNPs within each network: outdegree and core score. The outdegree measures the centrality of a SNP within the entire network. SNPs within the top 25% of outdegree distribution were considered as global hubs. The core score measures the contribution of SNPs to the modularity of the network module in which it arises. SNPs within the top 25% of core score distribution were considered local hubs (core SNPs).

We investigated whether trait heritability was distributed evenly across all SNPs or instead concentrated in SNPs with local or global centrality. Using a likelihood ratio test accounting for linkage disequilibrium and module size (see Methods), we found for each trait we considered that the SNPs explaining greatest portion of the heritability (top 5% of Z^2, or high heritability SNPs) are more likely to have both high outdegree and core scores (Supplementary Fig. S2 for outdegrees and Supplementary Fig. S3c for core scores). Using the Whole-Blood network as a model, we found that this enrichment in high heritability SNPs increases with the threshold chosen to define high outdegree and core score (from the top 90% to the top 5%, Supplementary Figure S4A.).

Given the demonstrated importance of eQTL network topology, we explored whether the increased heritability was driven simply by being in the network (a proxy for being an eQTL), being a global hub, or being a local hub. These annotations are overlapping and reflect potentially confounding factors such as underlying chromatin annotations (for example, global hubs are enriched for non-genic enhancers, while local hubs are enriched for genic enhancers and promoters (11)). For this reason, we partitioned heritability across various functional annotations while accounting for linked markers using stratified LD Score regression(15, 16) (see Methods), using the 97-levels baseline annotation model, to which we added our three annotations: belonging to eQTL network, being in the top quartile of outdegrees (global hubs), or of core scores (local hubs). The total proportion of heritability explained by SNPs as estimated using the LDScore software is reported in Supplementary Tab. S1. We found that these proportions are not always perfectly correlated with the estimated global genetic heritability reported in published twin and pedigree studies (see Supplementary Tab. S1), but our estimate of the total heritability of traits explained by SNPs as computed by the LDScore regression is coherent with previous reports (see (17) for breast, ovarian and prostate cancer examples).

The LDScore regression confirmed the results we obtained above using likelihood ratio tests: SNPs with high core scores or high outdegrees are enriched for trait heritability, and this enrichment increases with the threshold chosen to define high scores (Supplementary Fig. S4B). The detailed results for enrichment in h^2 among each annotation for each of the ten traits and diseases are reported in Supplementary Tab. S3.

We performed a meta-analysis across the ten uncorrelated traits and diseases for each tissue-specific network. An example of enrichment in h^2 among each annotation for the Whole-Blood network is shown in Fig. 2A. As noted earlier, SNPs that are within the eQTL networks tend to be significantly enriched in h^2 independent of tissue (Fig. 2B). However, the enrichment in h^2 is even greater for both high outdegree SNPs and high core score SNPs in all networks except Artery - Coronary and Liver. In many ways, this trend is exactly what one would expect: SNPs that fall within the eQTL networks have increased heritability because they are potentially capable of affecting the expression of multiple genes, a trend that increases as the SNPs become increasingly connected in the eQTL networks.

SNPs associated at the genome-wide level with traits or diseases are clustered in a few modules. We had previously reported that GWAS SNPs for chronic obstructive pulmonary disease, cancers, and other traits are concentrated in a small number of modules that are enriched for genes that are associated with trait-relevant GO Term biological functions (12). Consequently, we tested whether SNPs that carry the highest proportion of heritability are evenly distributed across the entire network or appear preferentially in specific network modules. We used the summary statistic Z^2 as a proxy for the per-SNP heritability, and SNPs with a GWAS Z^2 in the top 5% were named "high heritability SNPs". We found that the high heritability SNPs cluster in a small number of network modules (Supplementary Tab. S4). As an example, the breast cancer-associated SNPs appear in only 29 (12.6%) of the 230 network modules in the SKN (Skin -

DRAFT

Not sun-exposed–Suprapubic) eQTL network (Fig. 3A) representing a substantial concentration of heritability. We find similar results for other diseases and other tissues, with some combinations exhibiting even more substantial clustering of heritability in a limited number of clusters (Fig. 3B).

Because the highest heritability SNPs often only explain a small proportion of the heritability of associated traits (particularly when the trait is highly polygenic) we examined the distribution of per-SNP heritability across the different eQTL network modules. We found that heritability is not distributed evenly across modules (Kruskal-Wallis test p=0 for all pairs of tissue-specific networks and traits tested). As an example, consider Z^2 calculated for Alzheimer's disease in Brain - Nucleus Accumbens (basal ganglia) networks, which is plotted in Fig. 3C; all the results are in Supplementary Tab. S4). Comparing the distribution of Z^2 for each module with the rest of the network, we found that less than a third of the modules contain high heritability SNPs (54[5-84] modules representing about 23.7%[5.6-33.1] of the total number of modules depending on tissue-specific network/trait pair, with Benjamini-Hochberg-corrected Mann-Whitney U tests $p \leq 0.01$ - see Supplementary Tab. S5).

Heritability is enriched in trait-specific, functionally relevant modules. Finally, we investigated whether the heritability for different, uncorrelated traits was clustered in the same or different modules in the tissue-specific eQTL networks. Depending on the tissue-specific network, about 40% [28-53] of the modules were not enriched for high heritability SNPs associated with any traits, 29% [24-35] were enriched for high heritability SNPs from only one trait and can be considered trait-specific, and 3%[1-8] were enriched for high heritability SNPs from at least 5 of the 10 traits and can be considered as shared across (many) traits (Fig. 4A and Supplementary Fig. S5).

We performed a pairwise comparison of traits in each of the tissue-specific networks between modules enriched for high heritability SNPs. Using 10,000 resamplings, we found that top-heritability SNPs for two independent traits did not cluster in the same modules more than expected by chance in most cases. Indeed, among the 1305 pairwise comparisons possible — $\binom{10}{2}$ comparisons $\times 29$ tissues — only 28 show significant enrichment in overlap compared to what would expected by chance (Benjamini-Hocheberg-corrected $p \leq 0.01$). In half of these cases, the excess of overlap was observed between high HDL, height, and/or type 2 diabetes in various tissue-specific networks (see Supplementary Tab. S6).

We also investigated the biological functions represented by genes with the modules enriched for high heritability SNPs. We performed a GO term enrichment analysis for each module using Bioconductor `R` topGO package (see Methods); the results are presented in Supplementary Tab. S7. This allowed us to identify modules that were tissue-specific (see Methods). We then focused on trait- and tissue-specific modules, defined as those enriched for high heritability SNPs from one or two traits and enriched for tissue-specific GO Terms (Supplementary Tab. S8)—with the goal being to explain the specificity of each trait and/or disease heritability and underlying molecular mechanisms. We found that these modules were enriched in genes involved in biologically and trait-relevant functions.

For example, heritability for Alzheimer's disease and multiple sclerosis were both clustered in module 5 of the brain – nucleus accumbens (basal ganglia) network; module 5 is enriched in genes involved in catecholamine metabolic process (Fig. 4B), a class of molecules whose concentrations are altered in symptomatic Alzheimer's and multiple sclerosis diseases, both of which are amyloid plaque diseases (18, 19). Heritability for schizophrenia was most strongly clustered in module 100 of the same tissue, a molecule that is enriched for dopamine receptor signaling pathway genes (Fig. 4G), and this pathway is known to be functionally disrupted in the brain striatum in schizophrenia patients (20).

Unsurprisingly, heritability for cancers, including breast cancer, ovarian cancer, and prostate cancer, are enriched in epithelial tissues (skin–not sun-exposed exposed, skin–sun-exposed, colon–transverse) in network modules consisting of genes enriched for immune response (prostate cancer and module 76 of skin, not sun-exposed), response to cellular hypoxia (breast cancer, prostate cancer and module 157 of skin–sun-exposed), DNA break repair, cell cycle, apoptosis, and epithelium differentiation and growth (breast cancer and module 149 of skin–sun-exposed, and module 191 of skin (not sun-exposed), ovarian cancer and module 199 and skin, prostate cancer and modules 78, 97 of skin and module 149 of skin–sun-exposed). Particularly interesting, prostate cancer heritability is clustered in module 149 of the colon-transverse network, enriched for TRAIL-activated apoptotic signaling pathway genes, a long-known signaling pathway involved in cancer progression (21) (Fig. 4C). Breast and ovarian cancer heritability are also clustered in modules involved in estrogen metabolism and signaling (breast cancer and module 180 of skin–sun-exposed, ovarian cancer and module 182 of skin–not sun-exposed).

Metabolic traits such as high blood levels of HDL (high-density lipoprotein) and Type 2 Diabetes tend to cluster in modules enriched for lipid and carbohydrate metabolism. High HDL level heritability is enriched in module 142 of adipose–visceral omentum, enriched for genes involved in very-low and high-density lipoprotein particle assembly, remodeling, and clearance (Fig. 4D). Type 2 diabetes heritability clusters in module 17 of whole blood, enriched in genes involved in glycogen metabolism (Fig. 4F). Finally, heritability for height, a trait primarily related to development and, in particular, bone and muscle growth, is clustered in module 90 of muscle–skeletal, enriched for genes involved in muscle tissue morphogenesis, endochondral ossification, bone mineralization, and chondrocyte and osteoblast differentiation (Fig. 4E).

Conclusions

Of the SNPs found through GWAS to be associated with complex traits, most (up to 90%) fall outside of gene coding regions suggesting that they likely play a regulatory role.

DRAFT

Bipartite eQTL networks have allowed us to explore how these SNP loci work together to regulate the many genes involved in the biological processes underlying the trait, in a tissue- or cell-type-specific manner (11, 12, 14). Using eQTL networks, we found that not only do SNPs act in both *cis*- and *trans*- to influence the expression of complex networks of genes, but that these networks have a robust, modular structure consisting of SNP-gene modules with properties that help explain the polygenic influences that underlie most common traits. Specifically, the modules in eQTL networks are enriched for functionally related groups of genes and nucleated around "core SNPs" that are not only local (module) hubs but are also those SNPs with strong GWAS associations with various phenotypes (including disease) (11, 12).

Concurrent with our development of this eQTL network-based model, Boyle, Li, and Pritchard proposed the omnigenic model in which trait association signals are spread across most of the genome in a way that includes many genes lacking an obvious connection to a particular trait and they suggest that most heritability can be explained by effects on genes outside core pathways but that alter the functioning of those pathways (22, 23). This model was important in that it helped bridge the gap between the missing heritability found in many diseases and the growing number of traits in which hundreds, if not thousands, of small effect-size genetic variants contribute to particular traits.

Although it has previously been reported that global hubs in eQTL networks are enriched for tissue-relevant trait heritability (13), as are gene modules appearing in various types of networks(10), there has not been a systematic exploration of these two complementary and important conceptual advances in understanding genetic effects in complex traits. Here we present an analysis of ten polygenic traits representing a wide range of genetic heritability to determine the distribution of trait heritability across 29 tissue-specific eQTL networks.

We found that while heritability is widely distributed as suggested by the omnigenic model, the greatest heritability is concentrated in network modules that contain genes representing trait-specific and biologically relevant functions. This makes logical sense as phenotypic traits arise through alterations of specific biological processes relevant to individual traits. Further, we found that trait heritability was more likely to be explained by SNPs occupying key positions in the eQTL networks, especially among the “core SNPs” that are local hubs in their functional modules. These same SNPs, which we had previously shown to be enriched in tissue-specific activated regulatory elements(11), are thus likely to determine a significant proportion of the heritability of complex traits.

The clustering of heritability in modules is not evenly distributed across traits and differs between tissues. We find that heritability in each phenotype tends to be clustered in a small number of biologically relevant modules within the highly modular eQTL networks that are relevant for understanding the phenotype in question. As previously noted, these modules tend to be trait-specific even when traits are not genetically correlated. This module-dependent concentration of heritability, coupled with the enrichment of heritability in core SNPs of those same modules, could help explain why even highly connected regulatory networks (22, 23) are robust to the genetic perturbations of deleterious genetic variants. Indeed, the highly modular structure of eQTL networks provides a means by which the disruptive effect of regulatory mutations can be buffered in a tissue-specific manner against altering the broader functionality of the wider regulatory networks active in living cells.

Overall, our results both demonstrate the concordance between the eQTL network and omnigenic views of how genetic variants work together to influence traits and provide some important extensions of each that address their respective shortcomings. The value of such a synthesis can be seen in the results derived from the collection of complex traits we chose to analyze, including cancers, metabolic diseases, and auto-immune neurodegenerative diseases. In each of these traits, we can see genetic risk factors identified through GWAS perturbing tissue-specific functional modules (while affecting others), while those modules are simultaneously affected by many other genetic variants of smaller overall effect size. Together, these results define a conceptual framework for understanding disease risk in which one can attempt to prioritize therapeutic targets and consider ways in which treatments can be adjusted to allow for the genetic landscape that lays the groundwork for the phenotypes we manifest.

Methods

GTEx data set. We used genotyping and gene expression level data from the NHGRI GenotypeTissue Expression (GTEx) project version 8.0 (24). For appropriate statistical power in downstream analyses and network stability, we filtered out tissues for which the number of individuals with both genotyping and RNA-Seq data available was less than 200; sex-specific tissues were not included. This left 29 tissues for analyses, as can be found in (Supplementary Tab. S2). Genotyping data were downloaded from the database of Genotypes and Phenotypes (dbGaP): phs000424.v8.p2. Genotyping data were preprocessed on the Bridges system at the Pittsburgh Supercomputing Center (PSC) and the Cannon cluster supported by the Faculty of Arts and Sciences Division of Science, Research Computing Group at Harvard University (see (13)).

The sequencing data were processed in plink 1.90 to retain only SNPs, and we removed variants with genotype missingness greater than 10% or minor allele frequency less than 0.1 (25). SNP imputation was then performed using Eagle2 (26). Fully processed, filtered, and normalized RNA-Seq data were obtained from the GTEx Portal (`www.gtexportal.org`). The GENCODE 26 model was used to collapse transcripts and quantify expression using RNA-SeQC (`https://www.gencodegenes.org/human/release_26.html#`).

Bipartite eQTL networks inference. Expression quantitative trait loci (eQTLs) were obtained from (13). Rapidly, with the `R` MatrixEQTL package (27), the association between

DRAFT

SNP genotypes and gene expression was modeled using linear regression (Eq. 1) that included potential confounding factors as covariates (28): the two first principal components for population structure, sex, age and RIN that measures RNA quality. If $\mathbf{G}$ is an $r \times m$ matrix of gene expression and $\mathbf{S}$ is an $r \times n$ matrix of SNP genotypes, each with r rows representing observations and columns representing n SNPs and m genes, respectively, $\mathbf{X}$ is a covariate matrix, the eQTL of a particular SNP i on a locus's gene expression j is then :

$$\mathbf{G_j} = \mathbf{X^T}\boldsymbol{\alpha} + \beta_{ij}\mathbf{S_i}. \quad \textbf{(1)}$$

Associations were evaluated for SNPs in both *cis* – SNPs within 1MB of a gene's transcription start site – and *trans*. The eQTL associations between all pairs of SNPs and genes were then represented as a sparse, weighted bipartite network. Each SNP and gene was considered a node in the network. Using a fixed cutoff $q = 0.2$ on the false discovery rate (FDR) of the eQTL regression (11, 12), the edge weight $a_{i,j}$ between SNP i and gene j were defined by the function $I_{i,j}\{FDR < q\}|\beta_{ij}|$. Thus, when the estimated FDR of the eQTL regression was below the threshold of 0.2, then $a_{i,j} = |\beta_{ij}|$, indicating there was an edge connecting the nodes, and $a_{i,j} = 0$ otherwise.

Network summary statistics. The module structure of each tissue-specific network was determined using the bipartite modularity maximization approach (12), allowing for a balance between computation time and memory usage. This new implementation (CONDOR:condorSplitMatrixModularity) has been published in the `R` netzooR package fork of `https://github.com/maudf/`.

$$k_i = \sum_{j=1}^{g} I_{i,j}\{FDR < q\}|\beta_{ij}| \quad \textbf{(2)}$$

The SNP core score was defined as the SNP's contribution to the modularity of its module, it measures the centrality of the SNP in the module (12). If $m = s \times g$ is the total number of possible edges in a network made of s SNPs and g genes, $\tilde{a}_{ij}$ is the observed edge value between SNP i and gene j. Here, d_j is the gene i indegree defined as $d_j = \sum_{i=1}^{s} I_{i,j}\{FDR < q\}|\beta_{ij}|$, then, for SNP i in module h, its core score, Q_{ih}, is defined by Eq. 3:

$$Q_{ih} = \frac{1}{m}\sum_{j}\left(\tilde{a}_{ij} - \frac{k_i \times d_j}{m}\right)\delta(C_i, h)\delta(C_j, h) \quad \textbf{(3)}$$

GWAS data. GWAS data were obtained from the Alkes group (see Supplementary Tab. S1). We considered a total of ten traits and diseases presenting varying levels of estimated genetic heritability. For each trait or disease, we obtained summary statistics including SNP chromosome, position, allele 1 and 2, χ^2, and Z-score.
We used Z^2 as a proxy for normalized heritability explained by each SNP. High heritability SNPs were defined as those with a Z^2 in the 95th percentile of the distribution.

GWAS Outdegree and Core Score enrichment among high heritability SNPs. We compared the distribution of SNP outdegrees and core scores between the high heritability SNPs and the rest of the SNPs in the network using a likelihood-ratio test (LRT), correcting for linkage disequilibrium. To control for LD between SNPs, we generated lists of SNPs falling into the same LD block, using the plink1.9 –blocks option, a 5-Mb maximum block size, and an r^2 of 0.8. In each module, for each LD block, we extracted the median of either outdegrees (k_i) or core scores (Q_{ih}) for high and non-high heritability SNPs separately and used these values as input in the linear regressions.
The LRT we used assesses whether a linear model that includes GWAS status (Eq. 5) fits the observed data better than a linear model that does not include this variable (Eq. 4). As the distribution of SNP Q_{ih} is not uniform across modules, we added module identity as a covariate in the linear regression when computing LTR for core scores. In Eqs. 4 and 5, $Score_i$ is the score of SNP i (either outdegree or core score), $I(GWAS = 1)$ is an indicator function equal to 1 if the SNP is a high heritability SNP and equal to 0 otherwise, and $I(C_k = 1)$ is an indicator function equal to 1 if the SNP belongs to module k and equal to 0 otherwise:

$$Score_i \sim \left[\sum_{k=1}^{n-1} I(C_k = 1)\right] + \epsilon \quad \textbf{(4)}$$

$$Score_i \sim I(GWAS = 1) + \left[\sum_{k=1}^{n-1} I(C_k = 1)\right] + \epsilon \quad \textbf{(5)}$$

Genetic correlation between traits and heritability enrichment analyses among local and global hubs. We computed partitioned heritability with the 97 annotation baseline-LD model from (29, 30) for each of the ten traits and diseases selected above. This allowed us to estimate the enrichment and standardized effect size of the baseline annotations and three additional parameters on the heritability: belonging to the network (being an eQTL), having a high outdegree, and having a high core score (29, 30). We considered a binary annotation to ensure sufficiently stable estimates. For outdegrees and core scores, annotations were set equal to 1 if k_i and Q_{ih} were in the top quartile of the distribution and 0 otherwise.
Given that approximately 85% of the GTEx study population consists of individuals of European descent, we used LD scores computed from the 1000 Genomes Project data from individuals with European ancestry using the GRCh38 genome version and regression weights that exclude the HLA region; p-values were computed using a block-jacknife.
Meta-analyses were then performed across uncorrelated traits using the meta.summaries function from the `R` rmeta package v.3.0, with random-effect weights. To identify uncorrelated traits and diseases, pairwise genetic correlations between the ten traits and diseases were first computed using stratified LD score regression (S-LDSC). Traits and diseases with a pairwise genetic correlation of less than 0.3 were considered uncorrelated.

DRAFT

High heritability SNP enrichment among modules. High heritability SNP enrichment among modules was performed using a χ^2-test on data corrected for linkage disequilibrium. Using the same LD-blocks as previously described, we considered that an LD block was a high heritability block if at least one SNP in this LD block had a Z^2 in the top 95th percentile. For each module, we counted the number of high and non-high heritability LD blocks in and outside of the modules and used these data to perform a χ^2-test.

Gene Ontology enrichment analyses and identification of tissue-specific modules. We performed Gene Ontology enrichment analyses using the Bioconductor `R` topGO package v.2.44 (31), using the elim method; this method is more efficient than Fisher's Exact test (32). The GO categories are tested sequentially, following the GO tree structure from bottom to top: if one GO category is found to be significant, the genes involved are removed from the parent nodes before they are tested. The tests are thus not independent, and no multiple testing correction can be applied. Following the guidelines in the topGO users' manual, we filtered uncorrected p-values using a stringent threshold of 0.01. We also filtered out GO categories that did not include at least three genes in the gene set of interest. The gene ontology database used in this analysis was the one from the `R` bioconductor org.Hs.eg.db package v.3.13.0. For each test, the background gene set contains all the genes of the network and the gene set of interest contains all the genes from the module of interest. We identified common and tissue-specific modules in the eQTL networks based on pairwise comparisons of GO Term Assignments. For each module of a first network, the GO ID enriched in the module was compared to the GO ID enriched in each of the modules in a second network using Jaccard Index. The best matching module was determined based on the highest Jaccard Index, and if the best Jaccard Index was ≥ 0.3, the modules were considered as similar and otherwise different. Then, for each module in each tissue-specific network, we counted the number of similar modules in the other 28 networks. If this number of similar functional modules was ≤ 2, we considered it a tissue-specific module.

Data and code availability. All the code used to analyze the data is available at https://github.com/maudf/heritability. The new CONDOR:condorSplitMatrixModularity is available at https://github.com/maudf/netZooR The data used for the analyses described in this manuscript were obtained from: the GTEx Portal on 12/17/19 and dbGaP accession number phs000424.v8 on 12/17/19 for RNA-seq and Genotyping data, and from https://alkesgroup.broadinstitute.org/LDSCORE/all_sumstats/ for GWAS data.

Ethics approval. This work was conducted under dbGaP-approved protocol #9112.

ACKNOWLEDGEMENTS
Thank you to Alkes Price for discussions about LDSC and heritability. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. This work was supported by grants from the US National Institutes of Health, including grants from the National Heart, Lung and Blood Institute (5P01HL105339, 5P01HL114501; J.Q. and J.P.: 5R01HL111759; J.P.: K25HL140186), the National Cancer Institute (J.Q.: R35CA220523, 5P30CA006516; J.Q. and M.F.: 1R35CA197449), the National Institute of Allergy and Infectious Disease (J.Q. and J.P.: 5R01AI099204), the Marie Sklodowska-Curie grant PATTERNS (M.F.: 845083).

Bibliography

Author Contributions

KS, JP, JQ, and MF contributed to the conception of the work. KS, JP, JQ, and MF contributed to the study design and method development. KS, JP, and MF contributed analysis, verified the data, and drafted the manuscript. All authors reviewed the manuscript and approved the submitted work.

Declaration of interests. The authors declare no competing interests.

1. Zhihong Zhu, Futao Zhang, Han Hu, Andrew Bakshi, Matthew R. Robinson, Joseph E. Powell, Grant W. Montgomery, Michael E. Goddard, Naomi R. Wray, Peter M. Visscher, and Jian Yang. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nature Genetics*, 48(5):481, May 2016. ISSN 1546-1718. doi: 10.1038/ng.3538.
2. Farhad Hormozdiari, Martijn van de Bunt, Ayellet V. Segrè, Xiao Li, Jong Wha J. Joo, Michael Bilow, Jae Hoon Sul, Sriram Sankararaman, Bogdan Pasaniuc, and Eleazar Eskin. Colocalization of GWAS and eQTL Signals Detects Target Genes. *American Journal of Human Genetics*, 99(6):1245–1260, December 2016. ISSN 1537-6605. doi: 10.1016/j.ajhg.2016.10.003.
3. Lucas D. Ward and Manolis Kellis. Evidence of abundant purifying selection in humans for recently acquired regulatory functions. *Science (New York, N.Y.)*, 337(6102):1675–1678, September 2012. ISSN 1095-9203. doi: 10.1126/science.1225057.
4. Yu Gyoung Tak and Peggy J. Farnham. Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. *Epigenetics & Chromatin*, 8:57, December 2015. ISSN 1756-8935. doi: 10.1186/s13072-015-0050-4.
5. GTEx Consortium, Eric R. Gamazon, Ayellet V. Segrè, Martijn Van De Bunt, Xiaoquan Wen, Hualin S. Xi, Farhad Hormozdiari, Halit Ongen, Anuar Konkashbaev, Eske M. Derks, François Aguet, Jie Quan, Dan L. Nicolae, Eleazar Eskin, Manolis Kellis, Gad Getz, Mark I. McCarthy, Emmanouil T. Dermitzakis, Nancy J. Cox, and Kristin G. Ardlie. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nature Genetics*, 50(7):956–967, July 2018. ISSN 1061-4036, 1546-1718. doi: 10.1038/s41588-018-0154-4.
6. Jason M. Torres, Eric R. Gamazon, Esteban J. Parra, Jennifer E. Below, Adan Valladares-Salgado, Niels Wacher, Miguel Cruz, Craig L. Hanis, and Nancy J. Cox. Cross-Tissue and Tissue-Specific eQTLs: Partitioning the Heritability of a Complex Trait. *The American Journal of Human Genetics*, 95(5):521–534, November 2014. ISSN 00029297. doi: 10.1016/j.ajhg.2014.10.001.
7. Farhad Hormozdiari, Steven Gazal, Bryce Van De Geijn, Hilary K. Finucane, Chelsea J.-T. Ju, Po-Ru Loh, Armin Schoech, Yakir Reshef, Xuanyao Liu, Luke O'Connor, Alexander Gusev, Eleazar Eskin, and Alkes L. Price. Leveraging molecular quantitative trait loci to understand the genetic architecture of diseases and complex traits. *Nature Genetics*, 50(7):1041–1047, July 2018. ISSN 1061-4036, 1546-1718. doi: 10.1038/s41588-018-0148-2.
8. Douglas W. Yao, Luke J. O'Connor, Alkes L. Price, and Alexander Gusev. Quantifying genetic effects on disease mediated by assayed gene expression levels. *Nature Genetics*, 52(6):626–633, June 2020. ISSN 1546-1718. doi: 10.1038/s41588-020-0625-2.
9. Justin M. Luningham, Junyu Chen, Shizhen Tang, Philip L. De Jager, David A. Bennett, Aron S. Buchman, and Jingjing Yang. Bayesian Genome-wide TWAS Method to Leverage both cis- and trans-eQTL Information through Summary Statistics. *The American Journal of Human Genetics*, 107(4):714–726, October 2020. ISSN 0002-9297. doi: 10.1016/j.ajhg.2020.08.022.
10. Samuel S. Kim, Chengzhen Dai, Farhad Hormozdiari, Bryce van de Geijn, Steven Gazal, Yongjin Park, Luke O'Connor, Tiffany Amariuta, Po-Ru Loh, Hilary Finucane, Soumya Raychaudhuri, and Alkes L. Price. Genes with High Network Connectivity Are Enriched for Disease Heritability. *The American Journal of Human Genetics*, 104(5):896–913, May 2019. ISSN 0002-9297. doi: 10.1016/j.ajhg.2019.03.020.
11. Maud Fagny, Joseph N. Paulson, Marieke L. Kuijjer, Abhijeet R. Sonawane, Cho-Yi Chen, Camila M. Lopes-Ramos, Kimberly Glass, John Quackenbush, and John Platig. Exploring regulation in tissues with eQTL networks. *PNAS*, 114(37):E7841–E7850, September 2017. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1707375114.
12. John Platig, Peter J. Castaldi, Dawn DeMeo, and John Quackenbush. Bipartite Community Structure of eQTLs. *PLOS Computational Biology*, 12(9):e1005033, September 2016. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1005033.
13. Sheila M. Gaynor, Maud Fagny, Xihong Lin, John Platig, and John Quackenbush. Connectivity in eQTL networks dictates reproducibility and genomic properties. *Cell Reports Methods*, 2(5):100218, May 2022. ISSN 26672375. doi: 10.1016/j.crmeth.2022.100218.

DRAFT

14. Maud Fagny, John Platig, Marieke Lydia Kuijjer, Xihong Lin, and John Quackenbush. Non-genic cancer-risk SNPs affect oncogenes, tumour-suppressor genes, and immune function. *British Journal of Cancer*, December 2019. ISSN 1532-1827.
15. 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, and Schizophrenia Working Group of the Psychiatric Genomics Consortium. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, 47(3): 291–295, March 2015. ISSN 1546-1718. doi: 10.1038/ng.3211.
16. H. K. Finucane, B. Bulik-Sullivan, A. Gusev, G. Trynka, Y. Reshef, and P. R. Loh. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet*, 47, 2015. doi: 10.1038/ng.3404.
17. Sara Lindström, Hilary Finucane, Brendan Bulik-Sullivan, Fredrick R. Schumacher, Christopher I. Amos, Rayjean J. Hung, Kristin Rand, Stephen B. Gruber, David Conti, Jennifer B. Permuth, Hui-Yi Lin, Ellen L. Goode, Thomas A. Sellers, Laufey T. Amundadottir, Rachael Stolzenberg-Solomon, Alison Klein, Gloria Petersen, Harvey Risch, Brian Wolpin, Li Hsu, Jeroen R. Huyghe, Jenny Chang-Claude, Andrew Chan, Sonja Berndt, Rosalind Eeles, Douglas Easton, Christopher A. Haiman, David J. Hunter, Benjamin Neale, Alkes L. Price, and Peter Kraft. Quantifying the Genetic Correlation between Multiple Cancer Types. *Cancer Epidemiology, Biomarkers & Prevention*, 26(9):1427–1435, September 2017. ISSN 1055-9965, 1538-7755. doi: 10.1158/1055-9965.EPI-17-0211.
18. Kristi Henjum, Leiv Otto Watne, Kristin Godang, Nathalie Bodd Halaas, Rannveig Saksholm Eldholm, Kaj Blennow, Henrik Zetterberg, Ingvild Saltvedt, Jens Bollerslev, and Anne Brita Knapskog. Cerebrospinal fluid catecholamines in Alzheimer's disease patients with and without biological disease. *Translational Psychiatry*, 12(1):151, April 2022. ISSN 2158-3188. doi: 10.1038/s41398-022-01901-5.
19. Mara Cercignani, Ottavia Dipasquale, Iulia Bogdan, Tiziana Carandini, James Scott, Waqar Rashid, Osama Sabri, Swen Hesse, Michael Rullmann, Leonardo Lopiano, Mattia Veronese, Daniel Martins, and Marco Bozzali. Cognitive fatigue in multiple sclerosis is associated with alterations in the functional connectivity of monoamine circuits. *Brain Communications*, 3(2):fcab023, April 2021. ISSN 2632-1297. doi: 10.1093/braincomms/fcab023.
20. Robert A. McCutcheon, Anissa Abi-Dargham, and Oliver D. Howes. Schizophrenia, Dopamine and the Striatum: From Biology to Symptoms. *Trends in Neurosciences*, 42 (3):205–220, March 2019. ISSN 01662236. doi: 10.1016/j.tins.2018.12.004.
21. Ricky W. Johnstone, Ailsa J. Frew, and Mark J. Smyth. The TRAIL apoptotic pathway in cancer onset, progression and therapy. *Nature Reviews Cancer*, 8(10):782–798, October 2008. ISSN 1474-1768. doi: 10.1038/nrc2465.
22. Evan A. Boyle, Yang I. Li, and Jonathan K. Pritchard. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell*, 169(7):1177–1186, June 2017. ISSN 0092-8674. doi: 10.1016/j.cell.2017.05.038.
23. Xuanyao Liu, Yang I. Li, and Jonathan K. Pritchard. Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. *Cell*, 177(4):1022–1034.e6, May 2019. ISSN 00928674. doi: 10.1016/j.cell.2019.04.014.
24. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science*, 348(6235):648–660, May 2015. doi: 10.1126/science.1262110.
25. Shaun Purcell, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A. R. Ferreira, David Bender, Julian Maller, Pamela Sklar, Paul I. W. de Bakker, Mark J. Daly, and Pak C. Sham. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81(3):559–575, September 2007. ISSN 0002-9297. doi: 10.1086/519795.
26. Po-Ru Loh, Petr Danecek, Pier Francesco Palamara, Christian Fuchsberger, Yakir A Reshef, Hilary K Finucane, Sebastian Schoenherr, Lukas Forer, Shane McCarthy, Goncalo R Abecasis, Richard Durbin, and Alkes L Price. Reference-based phasing using the Haplotype Reference Consortium panel. *Nature Genetics*, 48(11):1443–1448, November 2016. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.3679.
27. Andrey A. Shabalin. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics (Oxford, England)*, 28(10):1353–1358, May 2012. ISSN 1367-4811. doi: 10.1093/bioinformatics/bts163.
28. C. M. Kendziorski, M. Chen, M. Yuan, H. Lan, and A. D. Attie. Statistical Methods for Expression Quantitative Trait Loci (eQTL) Mapping. *Biometrics*, 62(1):19–27, March 2006. ISSN 0006-341X. doi: 10.1111/j.1541-0420.2005.00437.x. Publisher: John Wiley & Sons, Ltd.
29. Steven Gazal, Po-Ru Loh, Hilary K. Finucane, Andrea Ganna, Armin Schoech, Shamil Sunyaev, and Alkes L. Price. Functional architecture of low-frequency variants highlights strength of negative selection across coding and non-coding annotations. *Nature Genetics*, 50(11):1600–1607, November 2018. ISSN 1061-4036, 1546-1718. doi: 10.1038/s41588-018-0231-8.
30. Margaux L.A. Hujoel, Steven Gazal, Farhad Hormozdiari, Bryce Van De Geijn, and Alkes L. Price. Disease Heritability Enrichment of Regulatory Elements Is Concentrated in Elements with Ancient Sequence Age and Conserved Function across Species. *The American Journal of Human Genetics*, 104(4):611–624, April 2019. ISSN 00029297. doi: 10.1016/j.ajhg.2019.02.008.
31. Adrian Alexa and Jorg Rahnenfuhrer. topGO: Enrichment Analysis for Gene Ontology, 2016.
32. Adrian Alexa, Jörg Rahnenführer, and Thomas Lengauer. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*, 22 (13):1600–1607, July 2006. ISSN 1367-4803. doi: 10.1093/bioinformatics/btl140.

DRAFT

Figures

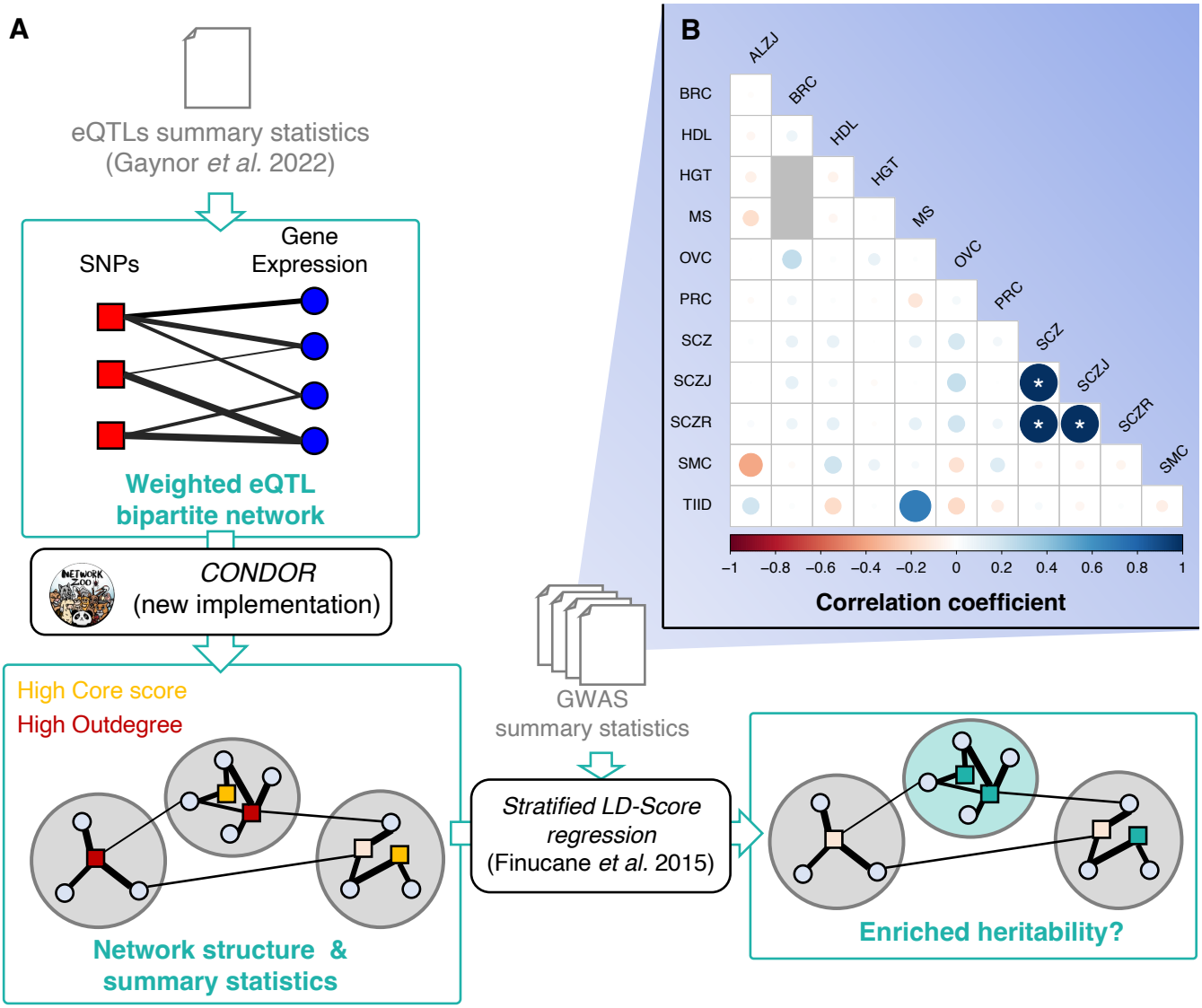


Fig. 1. Pipeline and Traits correlation. A. Pipeline of data analyses. Input data are in grey. GTEx eQTL summary statistics were obtained from a previous study (13) and are described in Methods and Supplementary Tab. S2. GWAS summary statistics were downloaded from `https://alkesgroup.broadinstitute.org/LDSCORE/all_sumstats/` and are described in Supplementary Tab. S1. eQTL network summary statistics are presented in Supplementary Fig. S1. **B.** Pairwise genetic correlations between traits based on GWAS summary statistics.

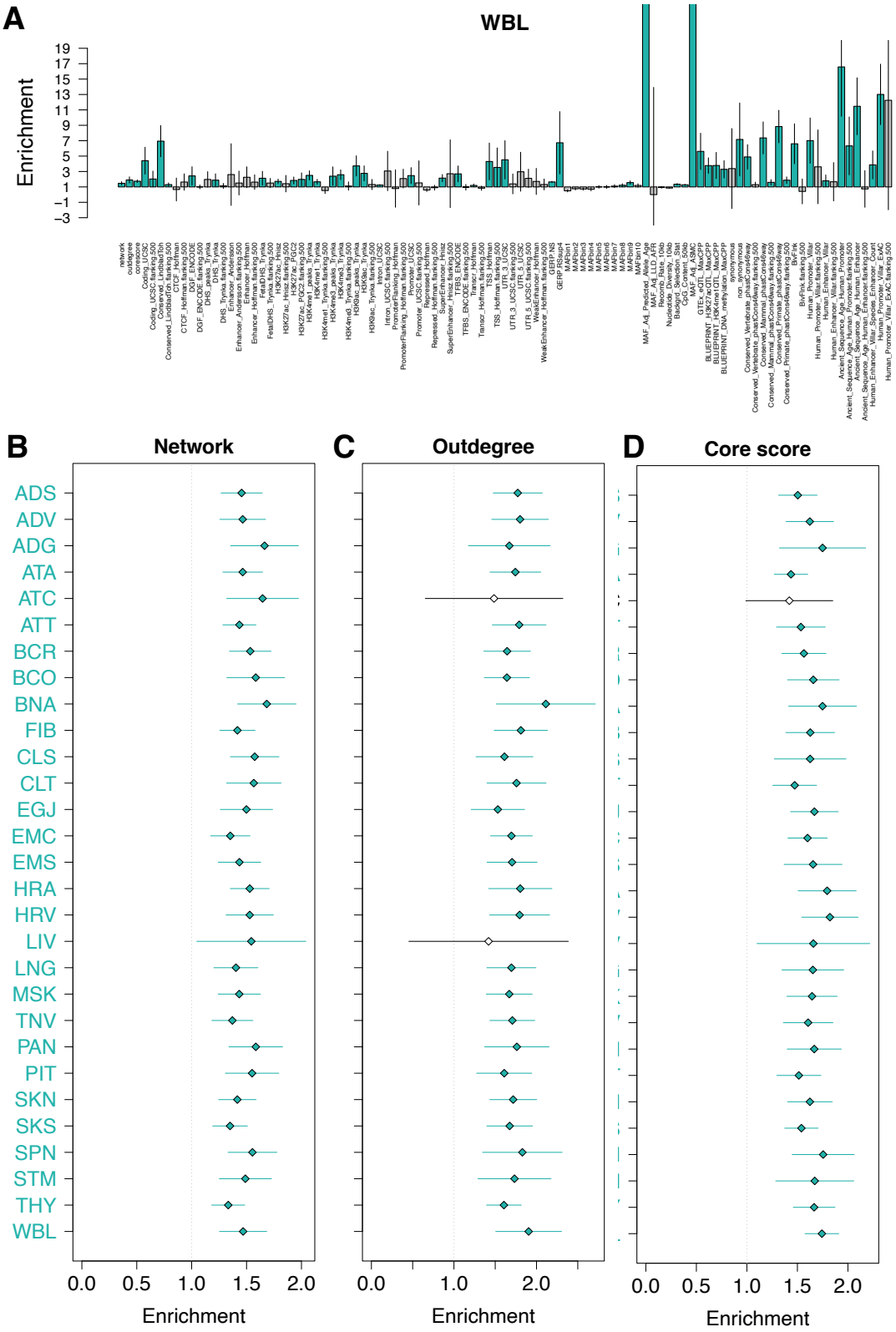


Fig. 2. Global and local hubs explain a larger share of heritability than expected by chance. A. Meta-analysis of h^2 enrichment among various annotations across 10 uncorrelated traits and diseases for the Whole Blood eQTL network. **B-D.** Meta-analysis of h^2 enrichment among SNP subcategories across 10 uncorrelated traits and diseases for each tissue-specific eQTL network. **B.** Among SNPs within the eQTL network. **C.** Among global hubs (high degrees). **D.** Among local hubs (high core scores.)

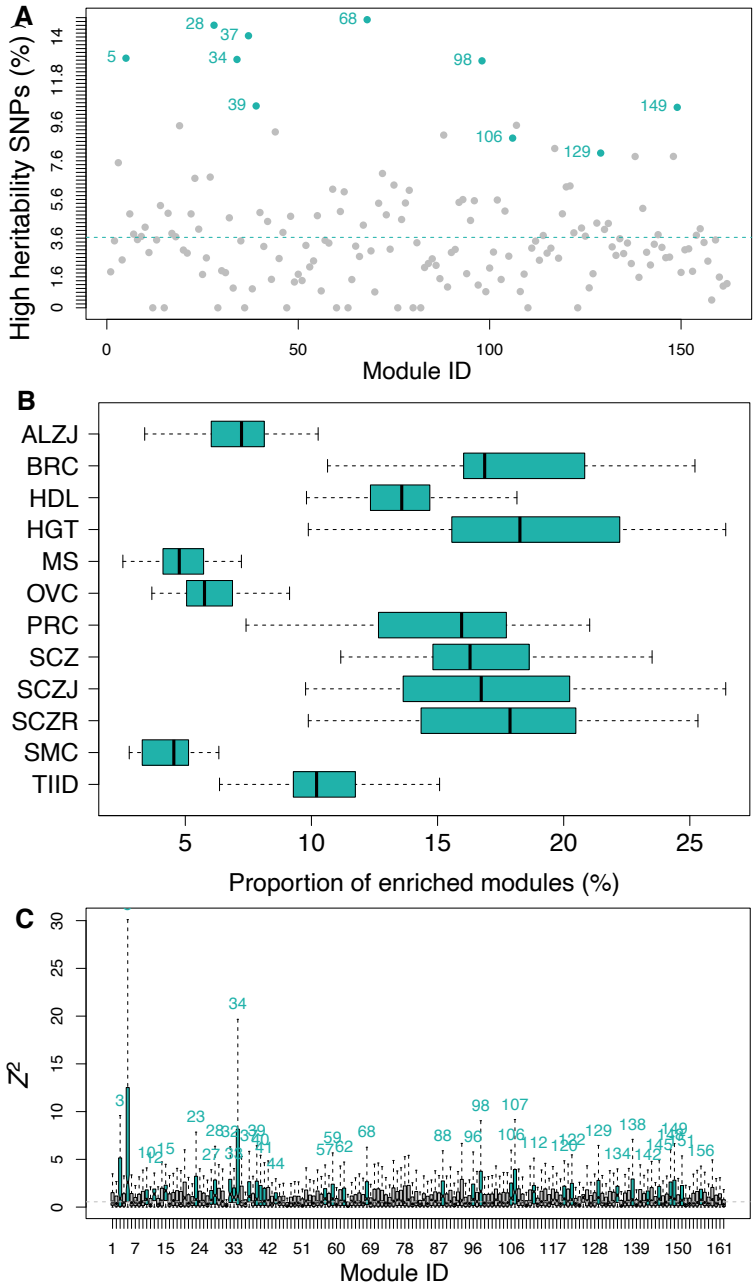


Fig. 3. High heritability SNPs are clustered in a few trait-specific modules. A. Proportion of high heritability SNPs for Breast Cancer in each module of the Skin - Not sun-exposed (Lower Leg) eQTL network. Modules significantly enriched for high heritability SNPs are highlighted in blue (Benjamini-Hochberg corrected $p \leq 0.01$ using a χ^2-test). The dotted blue line represents the expected proportion of high heritability SNPs by chance in each module. **B.** Distribution of the proportion of modules enriched for high heritability SNPs for each trait or disease among all tissue-specific eQTL networks. **C.** Distribution of Z^2 values in each module of Brain - Nucleus Accumbens (basal ganglia) for Alzheimer's Disease (Kruskal-Wallis p=0). Modules with a distribution skewed towards high values are represented in blue (modules with a Benjamini-Hochberg corrected $p \leq 0.01$ using one-sided Mann-Whitney U tests for each module *vs.* the rest of the network were considered as significantly enriched in high Z^2). The dotted blue line represents the expected Z^2 median across the whole network.

Fig. 4. Heritability is clustered in trait-specific, biologically significant modules. A. Average proportion of modules enriched in high heritability SNPs for one to ten traits or diseases across all tissue-specific modules. To remove artifacts due to the high genetic correlation between the three schizophrenia studies, only one schizophrenia study was taken into account in the analysis. The 0 category represents the modules that are not enriched in high heritability SNPs for any trait or disease. For a split by tissue-specific network, see Supplementary Fig. S5. **B-G.** Gene Ontology enrichment analyses results on gene content for a few tissue-specific modules enriched for high heritability SNPs for one or two traits and diseases. **B-C.** Bubble plots for top ten terms or terms with p-value < 0.01 for **B.** Brain (nucleus accumbens - basal ganglia) module 5, enriched for high heritability SNPs for Alzheimer's Disease and multiple sclerosis. **C.** Colon - Transverse module 149, enriched for high heritability SNPs for Prostate Cancer. **D-G.** Word clouds representing word frequencies in gene ontology terms enriched in **D.** Adipose - visceral omentum, module 142, enriched for high heritability SNPs for HDL; **E.** Muscle - skeletal, module 90, enriched for high heritability SNPs for height; **F.** Whole blood, module 17, enriched for high heritability SNPs for type 2 diabetes; **G.** Brain (nucleus accumbens - basal ganglia) module 100, enriched for high heritability SNPs for schizophrenia.

Supplementary Information

Supplementary Text.

eQTL networks are highly structured. Using RNA-seq and genotyping data from 29 tissues with > 200 matching samples from the GTEx data, we performed eQTL studies (see Methods). Including all associations with a Benjamini-Hochberg corrected *p*-value < 0.2, we built 29 tissue-specific weighted bipartite eQTL networks, using $|\beta_{ij}|$, the coefficient of the linear regression for SNP *i* and gene *j* as weights.
The structure of each eQTL network in terms of modules was then inferred using a new implementation of the bipartite maximization algorithm that balances computation time and memory usage (see Methods). All of the networks are highly structured, with modularity values ranging from 0.92 to 0.98 (Supplementary Fig. S1A). To note, as previously shown (11), eQTL network structure does not recapitulate local linkage disequilibrium, but rather summarizes the genetic component of gene expression regulation across the genome, with most modules presenting genes and SNPs from at least 2 different chromosomes (Supplementary Fig. S1B). The description in terms of the number of nodes (SNPs and genes) and edges can be found in Supplementary Fig. S1C-E.

Supplementary Tables.

DRAFT

Table S1. Description of GWAS data. See GWASDescription.txt.

Table S2. Description of GTEx tissues

Tissues	Abbreviations	Sample Counts
Adipose - Subcutaneous	ADS	581
Adipose - Visceral (Omentum)	ADV	469
Adrenal Gland	ADG	233
Artery - Aorta	ATA	387
Artery - Coronary	ATC	213
Artery - Tibial	ATT	584
Brain - Cerebellum	BCR	209
Brain - Cortex	BCO	205
Brain - Nucleus accumbens (basal ganglia)	BNA	202
Cells - Cultured fibroblasts	FIB	483
Colon - Sigmoid	CLS	318
Colon - Transverse	CLT	368
Esophagus - Gastroesophageal Junction	EGJ	330
Esophagus - Mucosa	EMC	497
Esophagus - Muscularis	EMS	465
Heart - Atrial Appendage	HRA	372
Heart - Left Ventricle	HRV	386
Liver	LIV	208
Lung	LNG	515
Muscle - Skeletal	MSK	706
Nerve - Tibial	TNV	532
Pancreas	PAN	305
Pituitary	PIT	237
Skin - Not Sun Exposed (Suprapubic)	SKN	517
Skin - Sun Exposed (Lower leg)	SKS	605
Spleen	SPN	227
Stomach	STM	324
Thyroid	THY	574
Whole Blood	WBL	670

DRAFT

Table S3. Enrichment of h^2 in various annotations obtained with LDSC See results_LDSCORE_all_scores.txt.

Table S4. Enrichment of each module in high mean Z^2 and high heritability SNPs. See all_signif_enrichment_bycluster.txt.

Table S5. Number of modules enriched with per-SNP $|Z^2$ and high heritability SNPs for each trait / tissue-specific network pair. See summary_signif_enrichment_bycluster.txt.

Table S6. Pairwise comparison between traits of modules enriched in high heritability SNPs. See overlapping_enrichment_bycluster_FDR.txt.

Table S7. Gene Ontology enrichment analysis results for each module in each tissue-specific network. See topGO_results.txt.

Table S8. Gene Ontology enrichment analysis results for each module in each tissue-specific network. See tissue-specific_trait-specific_enriched_modules.txt.

Supplementary Figures.

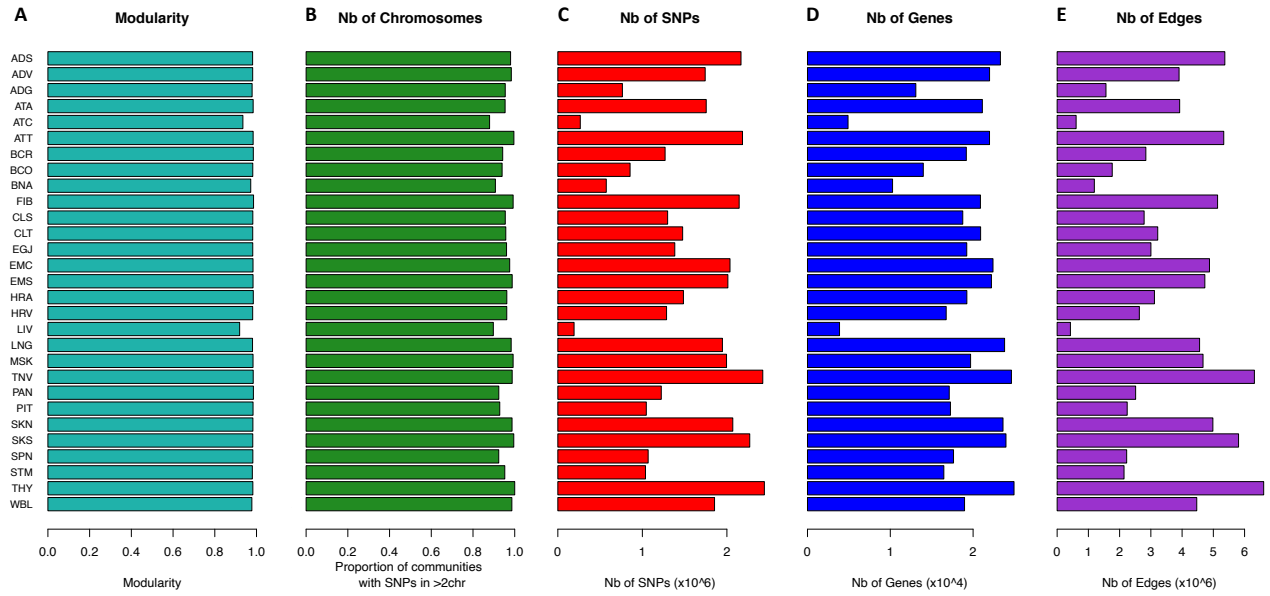
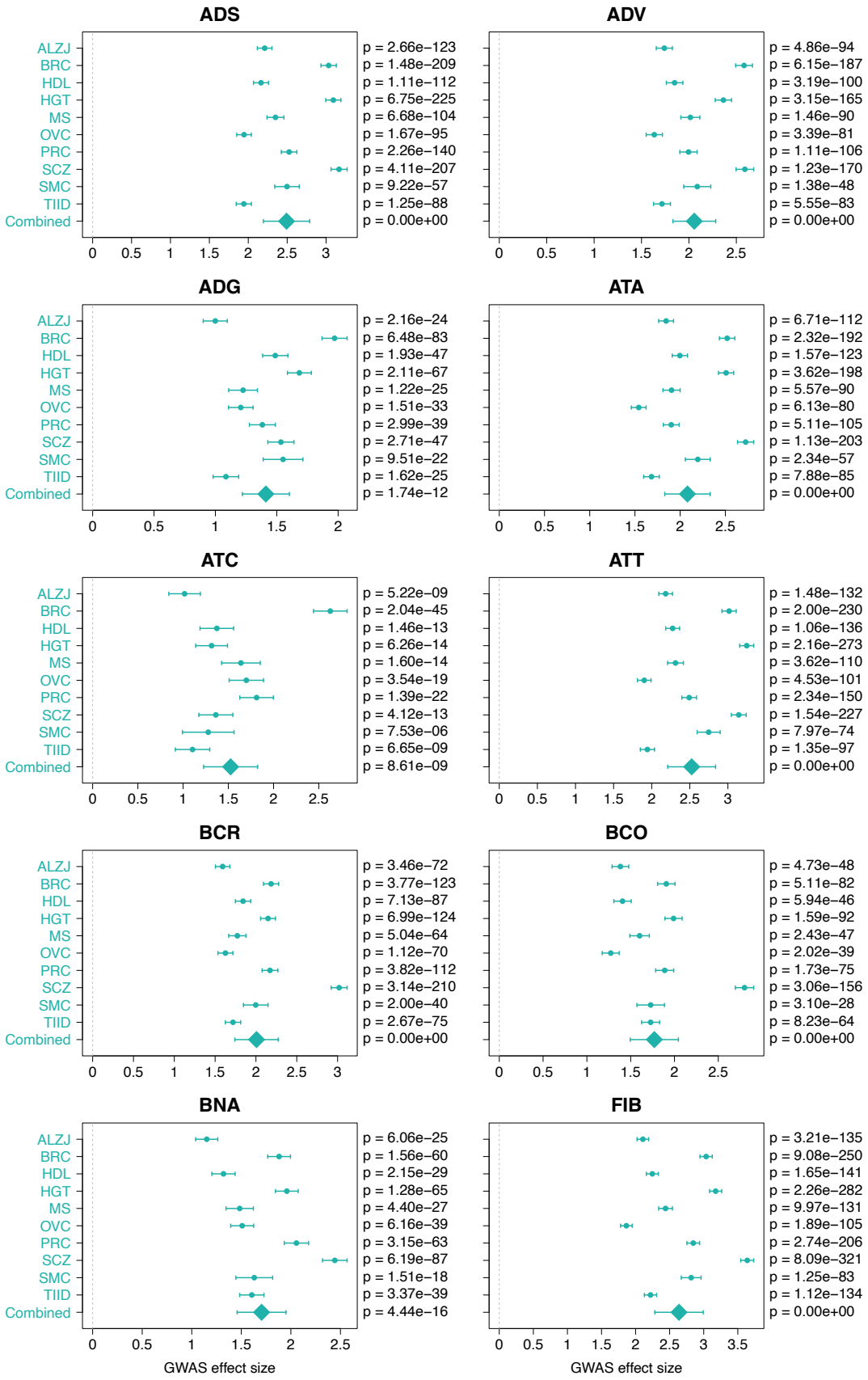
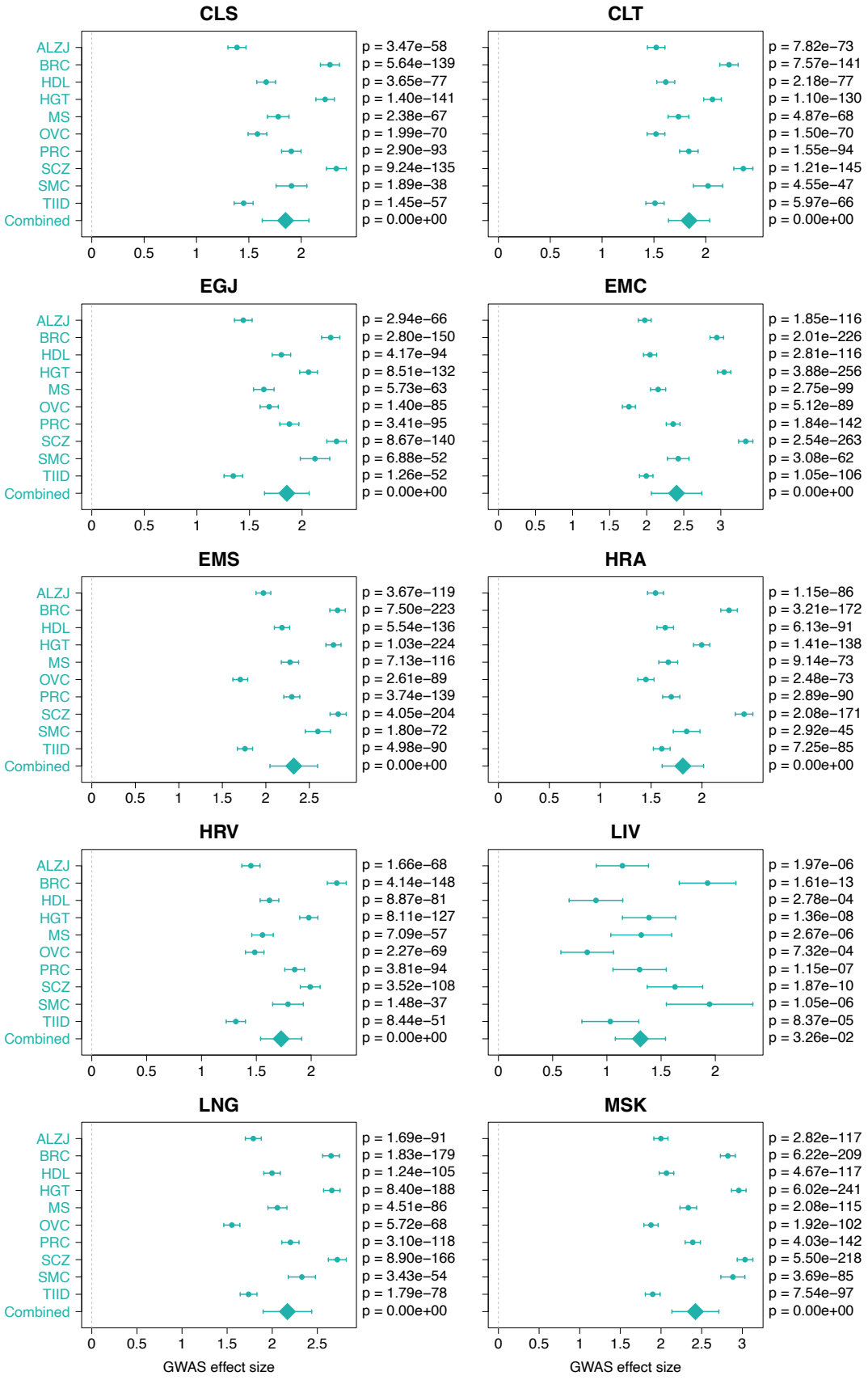


Fig. S1. eQTL networks are highly structured and reflect regulatory relationships A. Modularity of each tissue-specific eQTL network. **B.** Most modules from eQTL networks contain SNPs and genes from several modules and are not driven by linkage disequilibrium. **C-E.** Number of nodes and edges of each tissue-specific eQTL network. **C.** SNPs. **D.** Genes. **E.** Edges at FDR < 0.2.

DRAFT


ADS
ALZJ p = 2.66e−123
BRC p = 1.48e−209
HDL p = 1.11e−112
HGT p = 6.75e−225
MS p = 6.68e−104
OVC p = 1.67e−95
PRC p = 2.26e−140
SCZ p = 4.11e−207
SMC p = 9.22e−57
TIID p = 1.25e−88
Combined p = 0.00e+00
ADV
ALZJ p = 4.86e−94
BRC p = 6.15e−187
HDL p = 3.19e−100
HGT p = 3.15e−165
MS p = 1.46e−90
OVC p = 3.39e−81
PRC p = 1.11e−106
SCZ p = 1.23e−170
SMC p = 1.38e−48
TIID p = 5.55e−83
Combined p = 0.00e+00
ADG
ALZJ p = 2.16e−24
BRC p = 6.48e−83
HDL p = 1.93e−47
HGT p = 2.11e−67
MS p = 1.22e−25
OVC p = 1.51e−33
PRC p = 2.99e−39
SCZ p = 2.71e−47
SMC p = 9.51e−22
TIID p = 1.62e−25
Combined p = 1.74e−12
ATA
ALZJ p = 6.71e−112
BRC p = 2.32e−192
HDL p = 1.57e−123
HGT p = 3.62e−198
MS p = 5.57e−90
OVC p = 6.13e−80
PRC p = 5.11e−105
SCZ p = 1.13e−203
SMC p = 2.34e−57
TIID p = 7.88e−85
Combined p = 0.00e+00
ATC
ALZJ p = 5.22e−09
BRC p = 2.04e−45
HDL p = 1.46e−13
HGT p = 6.26e−14
MS p = 1.60e−14
OVC p = 3.54e−19
PRC p = 1.39e−22
SCZ p = 4.12e−13
SMC p = 7.53e−06
TIID p = 6.65e−09
Combined p = 8.61e−09
ATT
ALZJ p = 1.48e−132
BRC p = 2.00e−230
HDL p = 1.06e−136
HGT p = 2.16e−273
MS p = 3.62e−110
OVC p = 4.53e−101
PRC p = 2.34e−150
SCZ p = 1.54e−227
SMC p = 7.97e−74
TIID p = 1.35e−97
Combined p = 0.00e+00
BCR
ALZJ p = 3.46e−72
BRC p = 3.77e−123
HDL p = 7.13e−87
HGT p = 6.99e−124
MS p = 5.04e−64
OVC p = 1.12e−70
PRC p = 3.82e−112
SCZ p = 3.14e−210
SMC p = 2.00e−40
TIID p = 2.67e−75
Combined p = 0.00e+00
BCO
ALZJ p = 4.73e−48
BRC p = 5.11e−82
HDL p = 5.94e−46
HGT p = 1.59e−92
MS p = 2.43e−47
OVC p = 2.02e−39
PRC p = 1.73e−75
SCZ p = 3.06e−156
SMC p = 3.10e−28
TIID p = 8.23e−64
Combined p = 0.00e+00
BNA
ALZJ p = 6.06e−25
BRC p = 1.56e−60
HDL p = 2.15e−29
HGT p = 1.28e−65
MS p = 4.40e−27
OVC p = 6.16e−39
PRC p = 3.15e−63
SCZ p = 6.19e−87
SMC p = 1.51e−18
TIID p = 3.37e−39
Combined p = 4.44e−16
FIB
ALZJ p = 3.21e−135
BRC p = 9.08e−250
HDL p = 1.65e−141
HGT p = 2.26e−282
MS p = 9.97e−131
OVC p = 1.89e−105
PRC p = 2.74e−206
SCZ p = 8.09e−321
SMC p = 1.25e−83
TIID p = 1.12e−134
Combined p = 0.00e+00
GWAS effect size
GWAS effect size


CLS
ALZJ p = 3.47e−58
BRC p = 5.64e−139
HDL p = 3.65e−77
HGT p = 1.40e−141
MS p = 2.38e−67
OVC p = 1.99e−70
PRC p = 2.90e−93
SCZ p = 9.24e−135
SMC p = 1.89e−38
TIID p = 1.45e−57
Combined p = 0.00e+00
CLT
ALZJ p = 7.82e−73
BRC p = 7.57e−141
HDL p = 2.18e−77
HGT p = 1.10e−130
MS p = 4.87e−68
OVC p = 1.50e−70
PRC p = 1.55e−94
SCZ p = 1.21e−145
SMC p = 4.55e−47
TIID p = 5.97e−66
Combined p = 0.00e+00
EGJ
ALZJ p = 2.94e−66
BRC p = 2.80e−150
HDL p = 4.17e−94
HGT p = 8.51e−132
MS p = 5.73e−63
OVC p = 1.40e−85
PRC p = 3.41e−95
SCZ p = 8.67e−140
SMC p = 6.88e−52
TIID p = 1.26e−52
Combined p = 0.00e+00
EMC
ALZJ p = 1.85e−116
BRC p = 2.01e−226
HDL p = 2.81e−116
HGT p = 3.88e−256
MS p = 2.75e−99
OVC p = 5.12e−89
PRC p = 1.84e−142
SCZ p = 2.54e−263
SMC p = 3.08e−62
TIID p = 1.05e−106
Combined p = 0.00e+00
EMS
ALZJ p = 3.67e−119
BRC p = 7.50e−223
HDL p = 5.54e−136
HGT p = 1.03e−224
MS p = 7.13e−116
OVC p = 2.61e−89
PRC p = 3.74e−139
SCZ p = 4.05e−204
SMC p = 1.80e−72
TIID p = 4.98e−90
Combined p = 0.00e+00
HRA
ALZJ p = 1.15e−86
BRC p = 3.21e−172
HDL p = 6.13e−91
HGT p = 1.41e−138
MS p = 9.14e−73
OVC p = 2.48e−73
PRC p = 2.89e−90
SCZ p = 2.08e−171
SMC p = 2.92e−45
TIID p = 7.25e−85
Combined p = 0.00e+00
HRV
ALZJ p = 1.66e−68
BRC p = 4.14e−148
HDL p = 8.87e−81
HGT p = 8.11e−127
MS p = 7.09e−57
OVC p = 2.27e−69
PRC p = 3.81e−94
SCZ p = 3.52e−108
SMC p = 1.48e−37
TIID p = 8.44e−51
Combined p = 0.00e+00
LIV
ALZJ p = 1.97e−06
BRC p = 1.61e−13
HDL p = 2.78e−04
HGT p = 1.36e−08
MS p = 2.67e−06
OVC p = 7.32e−04
PRC p = 1.15e−07
SCZ p = 1.87e−10
SMC p = 1.05e−06
TIID p = 8.37e−05
Combined p = 3.26e−02
LNG
ALZJ p = 1.69e−91
BRC p = 1.83e−179
HDL p = 1.24e−105
HGT p = 8.40e−188
MS p = 4.51e−86
OVC p = 5.72e−68
PRC p = 3.10e−118
SCZ p = 8.90e−166
SMC p = 3.43e−54
TIID p = 1.79e−78
Combined p = 0.00e+00
MSK
ALZJ p = 2.82e−117
BRC p = 6.22e−209
HDL p = 4.67e−117
HGT p = 6.02e−241
MS p = 2.08e−115
OVC p = 1.92e−102
PRC p = 4.03e−142
SCZ p = 5.50e−218
SMC p = 3.69e−85
TIID p = 7.54e−97
Combined p = 0.00e+00
0 0.5 1 1.5 2 2.5 3
GWAS effect size

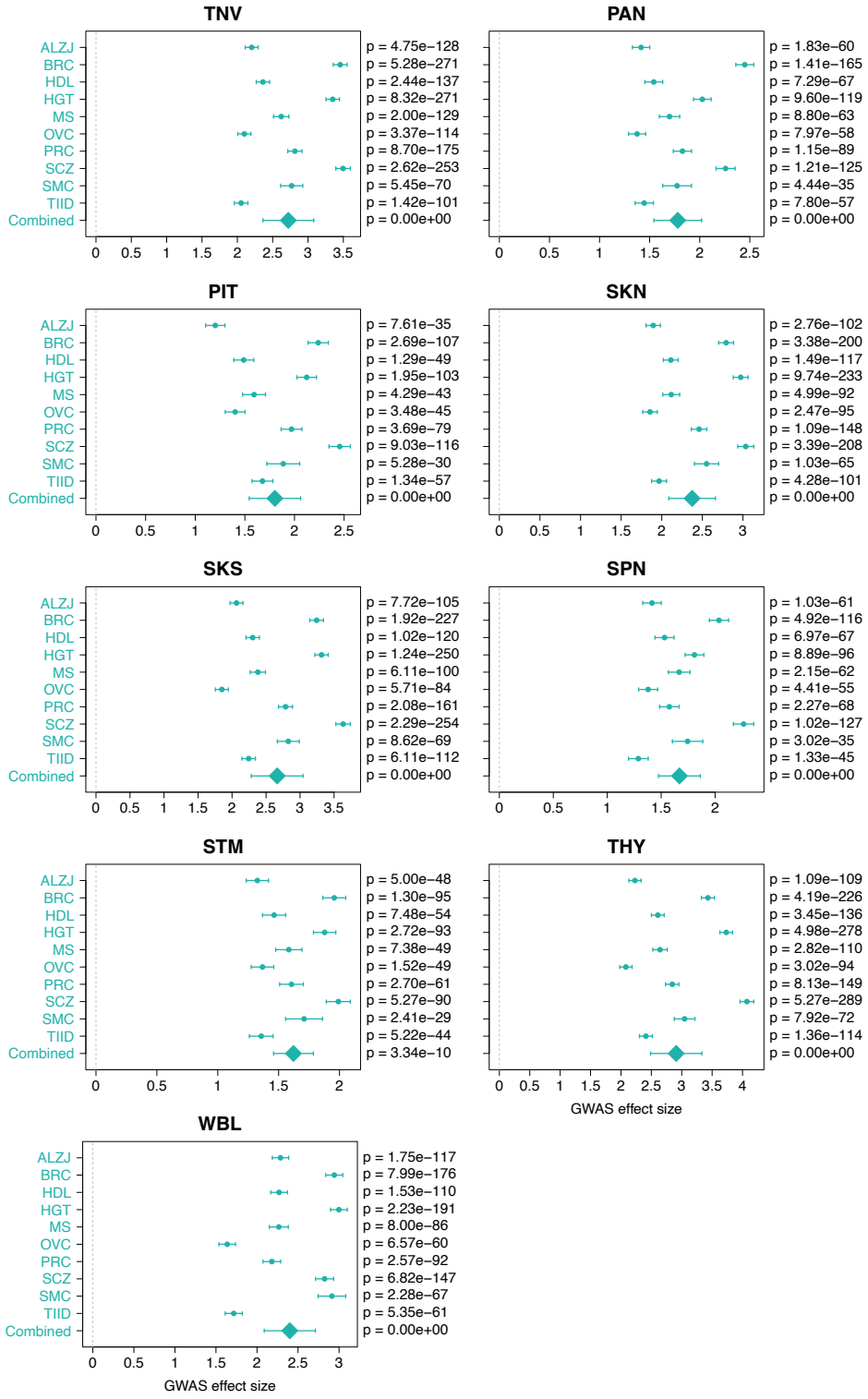
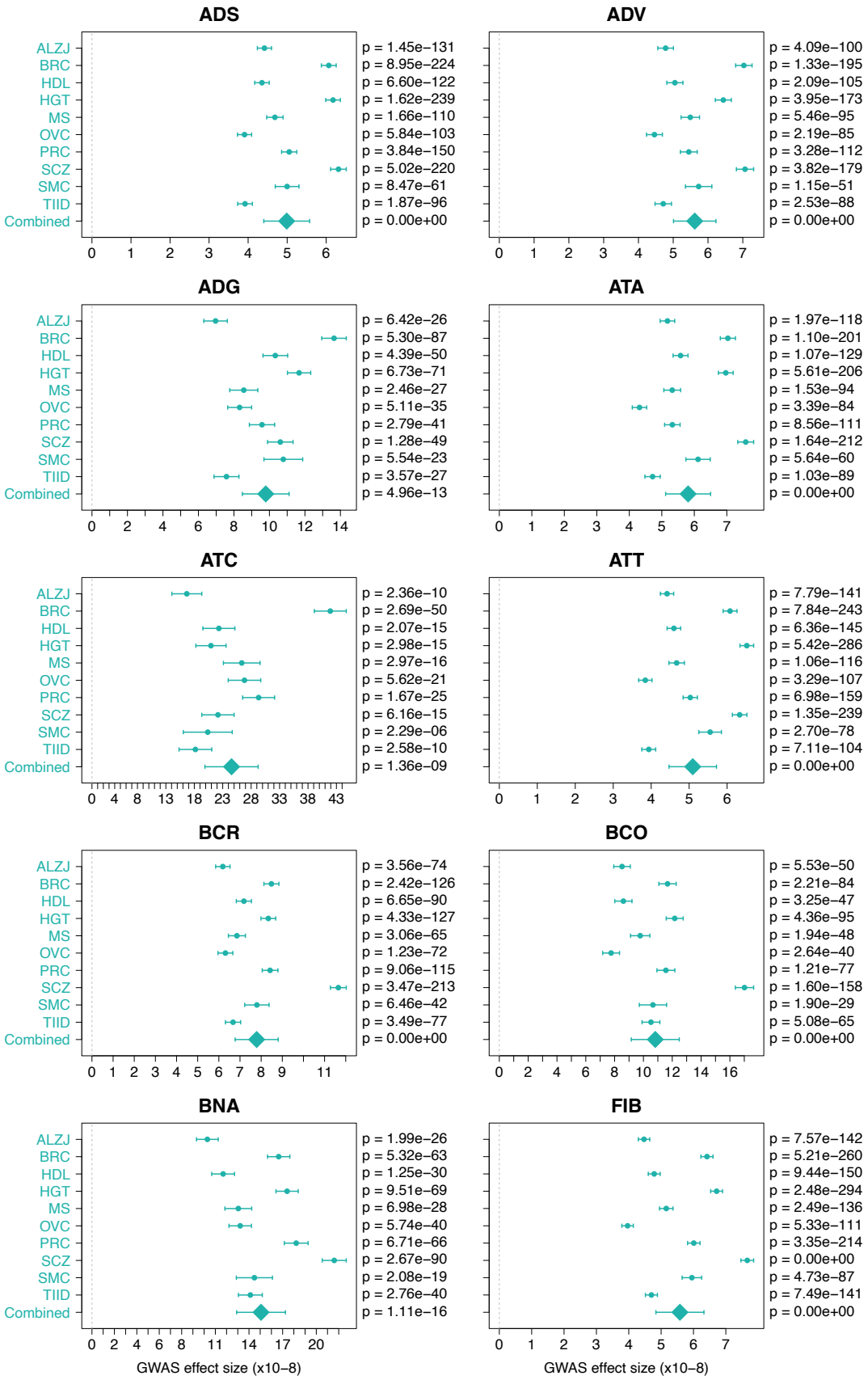
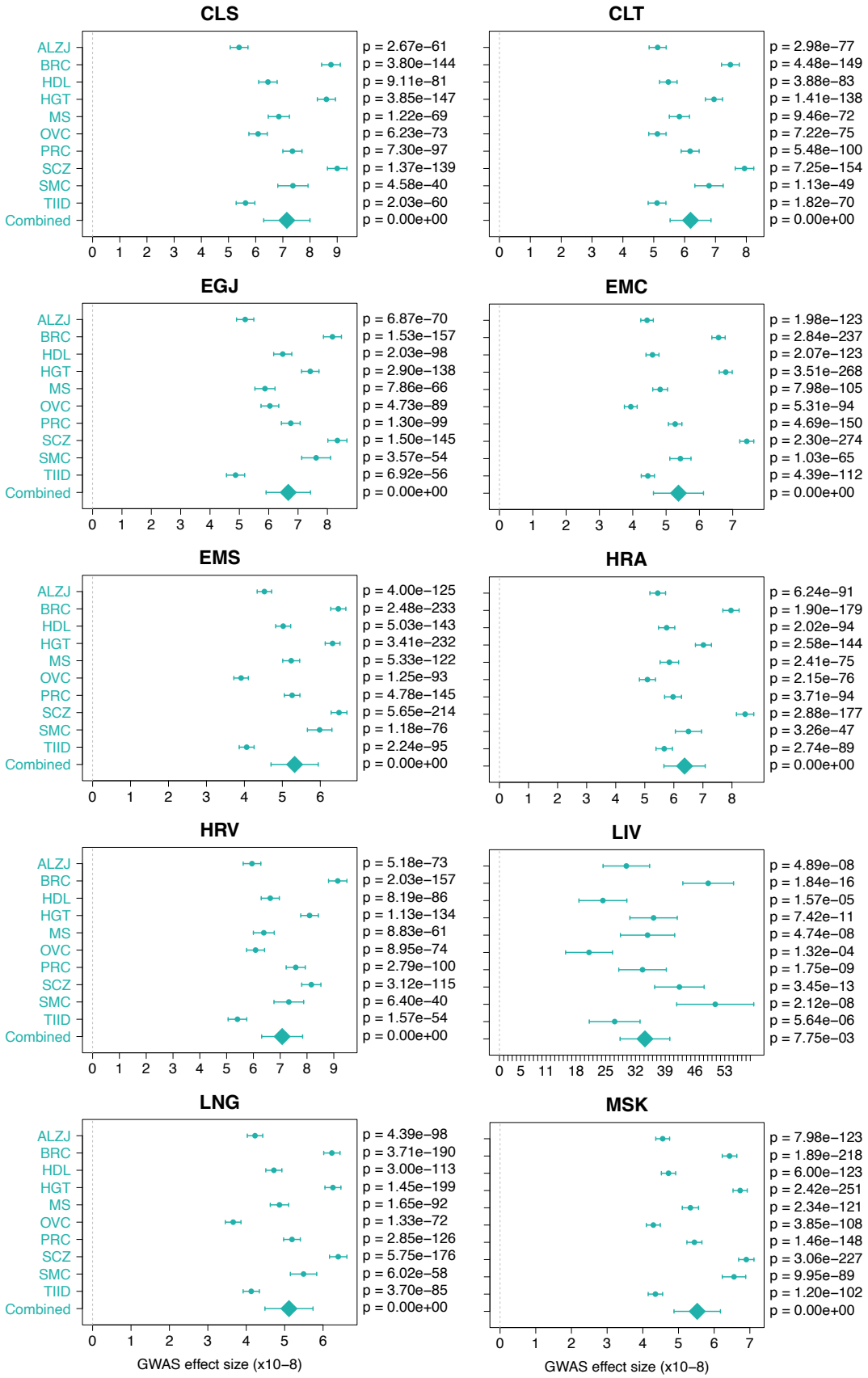


Fig. S2. High heritability SNPs have higher degree. Effect size of being a high heritability SNP on outdegree in each of the tissue-specific networks. *p*-values were obtained using a likelihood ratio test correcting for linkage disequilibrium.


ADS
ALZJ p = 1.45e−131
BRC p = 8.95e−224
HDL p = 6.60e−122
HGT p = 1.62e−239
MS p = 1.66e−110
OVC p = 5.84e−103
PRC p = 3.84e−150
SCZ p = 5.02e−220
SMC p = 8.47e−61
TIID p = 1.87e−96
Combined p = 0.00e+00
0 1 2 3 4 5 6
ADV
p = 4.09e−100
p = 1.33e−195
p = 2.09e−105
p = 3.95e−173
p = 5.46e−95
p = 2.19e−85
p = 3.28e−112
p = 3.82e−179
p = 1.15e−51
p = 2.53e−88
p = 0.00e+00
0 1 2 3 4 5 6 7
ADG
p = 6.42e−26
p = 5.30e−87
p = 4.39e−50
p = 6.73e−71
p = 2.46e−27
p = 5.11e−35
p = 2.79e−41
p = 1.28e−49
p = 5.54e−23
p = 3.57e−27
p = 4.96e−13
0 2 4 6 8 10 12 14
ATA
p = 1.97e−118
p = 1.10e−201
p = 1.07e−129
p = 5.61e−206
p = 1.53e−94
p = 3.39e−84
p = 8.56e−111
p = 1.64e−212
p = 5.64e−60
p = 1.03e−89
p = 0.00e+00
0 1 2 3 4 5 6 7
ATC
p = 2.36e−10
p = 2.69e−50
p = 2.07e−15
p = 2.98e−15
p = 2.97e−16
p = 5.62e−21
p = 1.67e−25
p = 6.16e−15
p = 2.29e−06
p = 2.58e−10
p = 1.36e−09
0 4 8 13 18 23 28 33 38 43
ATT
p = 7.79e−141
p = 7.84e−243
p = 6.36e−145
p = 5.42e−286
p = 1.06e−116
p = 3.29e−107
p = 6.98e−159
p = 1.35e−239
p = 2.70e−78
p = 7.11e−104
p = 0.00e+00
0 1 2 3 4 5 6
BCR
p = 3.56e−74
p = 2.42e−126
p = 6.65e−90
p = 4.33e−127
p = 3.06e−65
p = 1.23e−72
p = 9.06e−115
p = 3.47e−213
p = 6.46e−42
p = 3.49e−77
p = 0.00e+00
0 1 2 3 4 5 6 7 8 9 11
BCO
p = 5.53e−50
p = 2.21e−84
p = 3.25e−47
p = 4.36e−95
p = 1.94e−48
p = 2.64e−40
p = 1.21e−77
p = 1.60e−158
p = 1.90e−29
p = 5.08e−65
p = 0.00e+00
0 2 4 6 8 10 12 14 16
BNA
p = 1.99e−26
p = 5.32e−63
p = 1.25e−30
p = 9.51e−69
p = 6.98e−28
p = 5.74e−40
p = 6.71e−66
p = 2.67e−90
p = 2.08e−19
p = 2.76e−40
p = 1.11e−16
0 2 4 6 8 11 14 17 20
GWAS effect size (x10−8)
FIB
p = 7.57e−142
p = 5.21e−260
p = 9.44e−150
p = 2.48e−294
p = 2.49e−136
p = 5.33e−111
p = 3.35e−214
p = 0.00e+00
p = 4.73e−87
p = 7.49e−141
p = 0.00e+00
0 1 2 3 4 5 6 7
GWAS effect size (x10−8)


CLS
ALZJ p = 2.67e−61
BRC p = 3.80e−144
HDL p = 9.11e−81
HGT p = 3.85e−147
MS p = 1.22e−69
OVC p = 6.23e−73
PRC p = 7.30e−97
SCZ p = 1.37e−139
SMC p = 4.58e−40
TIID p = 2.03e−60
Combined p = 0.00e+00
0 1 2 3 4 5 6 7 8 9
CLT
ALZJ p = 2.98e−77
BRC p = 4.48e−149
HDL p = 3.88e−83
HGT p = 1.41e−138
MS p = 9.46e−72
OVC p = 7.22e−75
PRC p = 5.48e−100
SCZ p = 7.25e−154
SMC p = 1.13e−49
TIID p = 1.82e−70
Combined p = 0.00e+00
0 1 2 3 4 5 6 7 8
EGJ
ALZJ p = 6.87e−70
BRC p = 1.53e−157
HDL p = 2.03e−98
HGT p = 2.90e−138
MS p = 7.86e−66
OVC p = 4.73e−89
PRC p = 1.30e−99
SCZ p = 1.50e−145
SMC p = 3.57e−54
TIID p = 6.92e−56
Combined p = 0.00e+00
0 1 2 3 4 5 6 7 8
EMC
ALZJ p = 1.98e−123
BRC p = 2.84e−237
HDL p = 2.07e−123
HGT p = 3.51e−268
MS p = 7.98e−105
OVC p = 5.31e−94
PRC p = 4.69e−150
SCZ p = 2.30e−274
SMC p = 1.03e−65
TIID p = 4.39e−112
Combined p = 0.00e+00
0 1 2 3 4 5 6 7
EMS
ALZJ p = 4.00e−125
BRC p = 2.48e−233
HDL p = 5.03e−143
HGT p = 3.41e−232
MS p = 5.33e−122
OVC p = 1.25e−93
PRC p = 4.78e−145
SCZ p = 5.65e−214
SMC p = 1.18e−76
TIID p = 2.24e−95
Combined p = 0.00e+00
0 1 2 3 4 5 6
HRA
ALZJ p = 6.24e−91
BRC p = 1.90e−179
HDL p = 2.02e−94
HGT p = 2.58e−144
MS p = 2.41e−75
OVC p = 2.15e−76
PRC p = 3.71e−94
SCZ p = 2.88e−177
SMC p = 3.26e−47
TIID p = 2.74e−89
Combined p = 0.00e+00
0 1 2 3 4 5 6 7 8
HRV
ALZJ p = 5.18e−73
BRC p = 2.03e−157
HDL p = 8.19e−86
HGT p = 1.13e−134
MS p = 8.83e−61
OVC p = 8.95e−74
PRC p = 2.79e−100
SCZ p = 3.12e−115
SMC p = 6.40e−40
TIID p = 1.57e−54
Combined p = 0.00e+00
0 1 2 3 4 5 6 7 8 9
LIV
ALZJ p = 4.89e−08
BRC p = 1.84e−16
HDL p = 1.57e−05
HGT p = 7.42e−11
MS p = 4.74e−08
OVC p = 1.32e−04
PRC p = 1.75e−09
SCZ p = 3.45e−13
SMC p = 2.12e−08
TIID p = 5.64e−06
Combined p = 7.75e−03
0 5 11 18 25 32 39 46 53
LNG
ALZJ p = 4.39e−98
BRC p = 3.71e−190
HDL p = 3.00e−113
HGT p = 1.45e−199
MS p = 1.65e−92
OVC p = 1.33e−72
PRC p = 2.85e−126
SCZ p = 5.75e−176
SMC p = 6.02e−58
TIID p = 3.70e−85
Combined p = 0.00e+00
0 1 2 3 4 5 6
GWAS effect size (x10−8)
MSK
ALZJ p = 7.98e−123
BRC p = 1.89e−218
HDL p = 6.00e−123
HGT p = 2.42e−251
MS p = 2.34e−121
OVC p = 3.85e−108
PRC p = 1.46e−148
SCZ p = 3.06e−227
SMC p = 9.95e−89
TIID p = 1.20e−102
Combined p = 0.00e+00
0 1 2 3 4 5 6 7
GWAS effect size (x10−8)

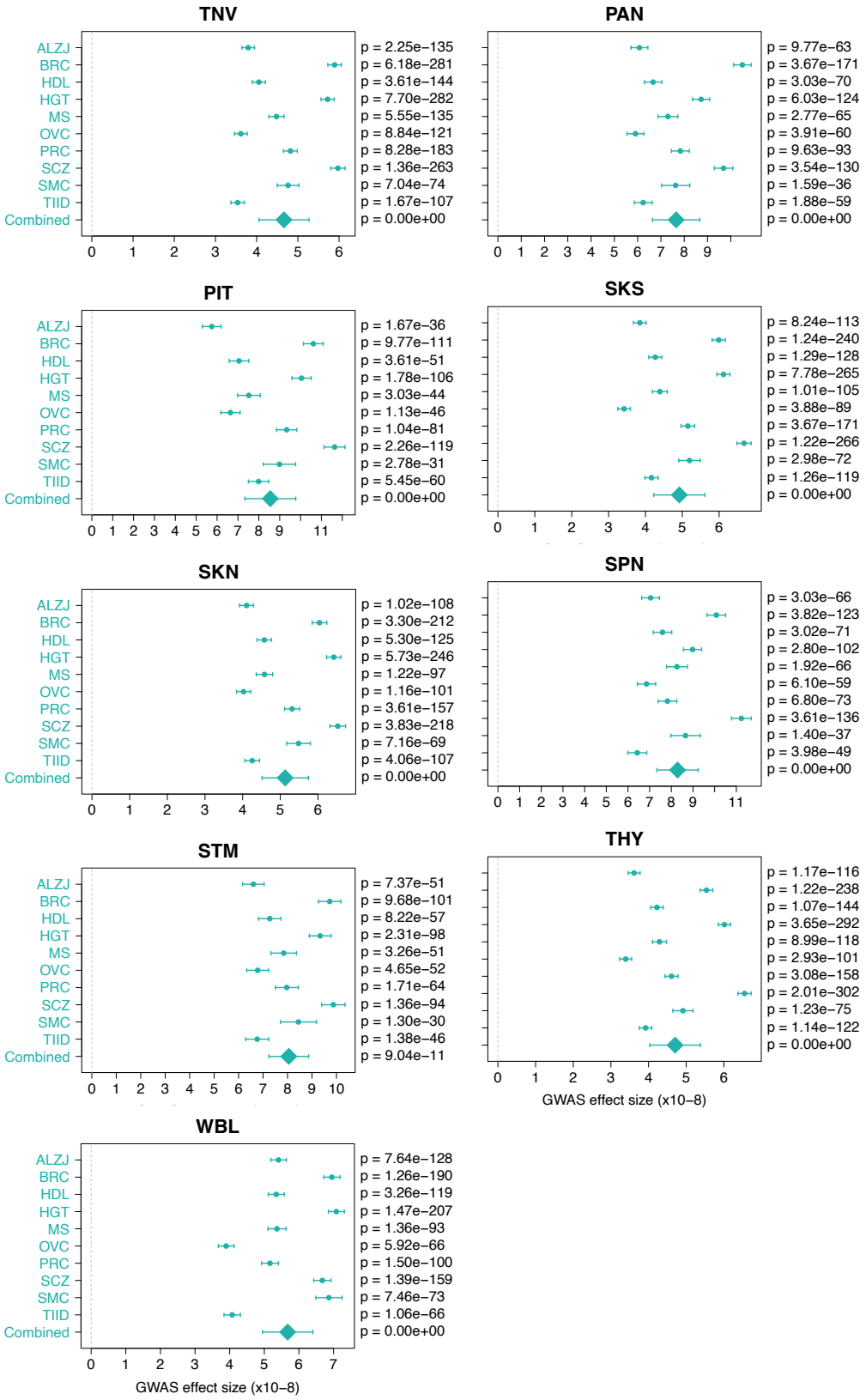


Fig. S3. High heritability SNPs have higher core scores. Effect size of being a high heritability SNP on core score in each of the tissue-specific networks. *p*-values were obtained using a likelihood ratio test correcting for linkage disequilibrium and module size. As core scores depend on the number of nodes in a module, core scores were corrected for module identity.

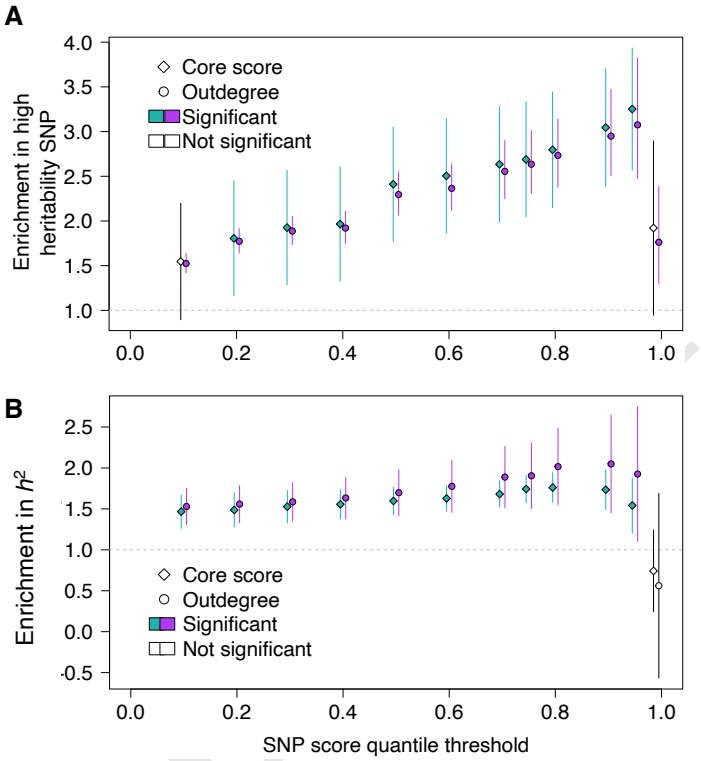


Fig. S4. Local and global hubs explain a larger share of heritability than expected by chance. Enrichments were computed using different thresholds to define high core scores and high outdegree in the whole-blood eQTL network from the first 9 deciles to the top percentile. **A.** Enrichment of high heritability SNPs among local and global hubs obtained using a likelihood ratio test. **B.** Enrichment of h^2 explained by global and local hubs obtained using stratified LD Scores and the basal model. **A-B.** Diamonds = core scores (local hubs). Circles = degrees (global hubs). Colored points = significant enrichment computed using a meta-analysis across 10 uncorrelated traits and diseases.

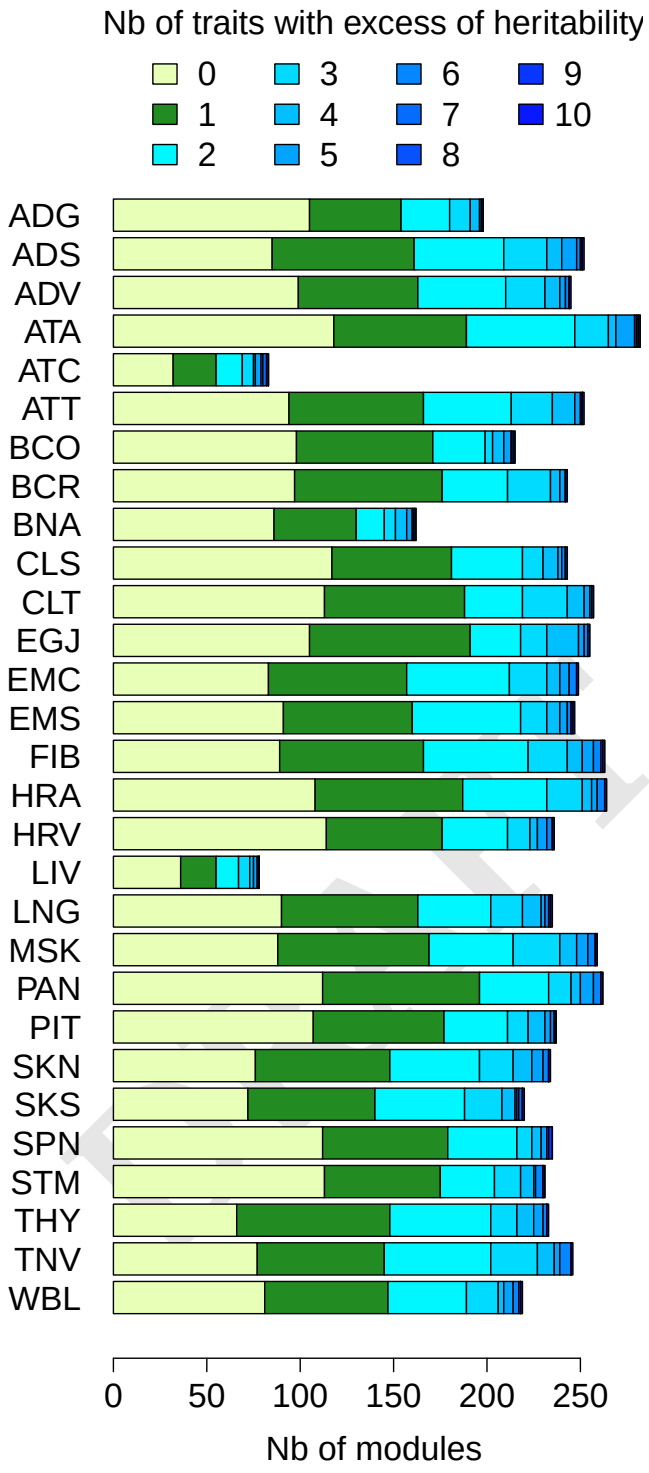


Fig. S5. Proportion of modules enriched for heritability for 0 to 10 traits