



HAL
open science

The Importance of Regulatory Network Structure for Complex Trait Heritability and Evolution

Katherine L Stone, John Platig, John Quackenbush, Maud Fagny

► To cite this version:

Katherine L Stone, John Platig, John Quackenbush, Maud Fagny. The Importance of Regulatory Network Structure for Complex Trait Heritability and Evolution. *Molecular Biology and Evolution*, 2025, 42 (8), pp.msaf174. <10.1093/molbev/msaf174>. <hal-05314722>

HAL Id: hal-05314722

<https://hal.science/hal-05314722v1>

Submitted on 14 Oct 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC0 1.0 - Universal - International License

The Importance of Regulatory Network Structure for Complex Trait Heritability and Evolution

Katherine L. Stone ^{1,2}, John Platig ^{3,4,5}, John Quackenbush ^{1,2,6}, Maud Fagny ^{1,2,7,*}

¹Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

²Department of Data Science and Center for Cancer Computational Biology, Dana-Farber Cancer Institute, Boston, MA, USA

³Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA

⁴Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA

⁵Department of Biomedical Engineering, University of Virginia, Charlottesville, VA, USA

⁶Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA, USA

⁷INRAE, CNRS, AgroParisTech, Genetique Quantitative et Evolution—Le Moulon, Université Paris-Saclay, Gif-sur-Yvette 91190, France

*Corresponding author: E-mail: maud.fagny@inrae.fr

Associate editor: Anne-Ruxandra Carvunis

Abstract

Complex traits are determined by many loci—mostly regulatory elements—that, through combinatorial interactions, can affect multiple traits. Such high levels of epistasis and pleiotropy have been proposed in the omnigenic model and may explain why such a large part of complex trait heritability is usually missed by genome-wide association studies, while raising questions about the possibility for such traits to evolve in response to environmental constraints. To explore the molecular bases of complex traits and understand how they can adapt, we systematically analyzed the distribution of SNP heritability for 11 traits across 29 tissue-specific expression quantitative trait locus networks. We find that heritability is clustered in a small number of tissue-specific, functionally relevant SNP–gene modules and that the greatest heritability 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. We confirm that this structure has allowed complex traits to evolve in response to environmental constraints, with the local “hubs” being the preferential targets of past and ongoing directional selection. Together, these results provide a conceptual framework for understanding complex trait architecture and evolution.

Keywords: GTEx, eQTL, bipartite networks, GWAS, heritability, complex traits, polygenic selection

Introduction

In many organisms, including yeasts, insects, worms, plants, and mammals, adaptation often leverages polygenic traits to respond to new environmental challenges (Daub et al. 2013; He et al. 2016; Zan and Carlborg 2019). Such “complex” traits have an interconnected genetic architecture involving from a small number to potentially thousands of partly independent loci (Salomé et al. 2011; Pais et al. 2013; Peiffer et al. 2014; Shi et al. 2016). Up to 90% of the SNPs associated with such traits are localized outside of coding regions, according to genome-wide association studies (GWAS) (Ward and Kellis 2012; Tak and Farnham 2015). They are thus likely to affect traits by modifying gene expression—often by altering the sequence of *cis*-regulatory elements. This is consistent with the observation that GWAS-significant regions colocalize with SNPs that are linked to the expression of nearby genes (expression quantitative trait loci, *cis*-eQTLs) (Hormozdiari et al. 2016; Zhu et al. 2016). 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 (Torres et al. 2014; Gamazon et al. 2018; Hormozdiari et al. 2018) but still fail to explain a majority of trait heritability (Yao et al. 2020). Others have shown that including *trans*-eQTLs may capture additional heritability and explain important pathological mechanisms

(Luningham et al. 2020), but this is rarely done because *trans*-eQTL studies are generally underpowered.

Advances in systems biology and functional genomics have highlighted the fact that at the molecular level, complex traits are the results of many regulatory interactions between actors of different biological processes within a complex cellular gene regulatory network (Hallgrímsson et al. 2014; Sonawane et al. 2019). This high level of pleiotropy coupled with significant epistatic interactions may explain the missing heritability observed in quantitative genetics studies and has led to the development of the omnigenic model (Boyle et al. 2017; Liu et al. 2019). In this model, 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. The model posits that most heritability can be explained by effects on genes outside of “core pathways,” often acting in *trans*, but acting in ways that alter the functioning of those pathways and thus account for most of a given trait’s heritability. The amount of pleiotropy implied by this model may limit, at first glance, the possibility for a given complex trait to adapt in response to an environmental change without interfering with other unrelated traits. However, the evidence for polygenic adaptation processes in SNPs associated with complex traits in several species highlights the need to dig deeper into the molecular bases of complex traits to understand their architecture and how they adapt (Barreiro and Quintana-Murci 2010;

Received: September 10, 2024. Revised: May 19, 2025. Accepted: June 27, 2025

Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution 2025.

This work is written by (a) US Government employee(s) and is in the public domain in the US.

Jakobson et al. 2019; Fagny and Austerlitz 2021; Mathieson 2021).

Given the number of loci and the number of potential interactions involved, gene regulatory networks provide an efficient tool for understanding the biological processes behind trait heritability (Kim et al. 2019). Combining GWAS results with gene coexpression networks have proven useful to identify the causal gene within a candidate region in many species (Chan et al. 2011; Lee et al. 2011; Schaefer et al. 2014; Taşan et al. 2015; Calabrese et al. 2017; Schaefer et al. 2018). By integrating SNP–gene or SNP–trait association studies results with coexpression networks, these studies have highlighted the fact that *cis*- and *trans*-regulatory relationships involved in complex trait determination are clustering within coexpression subnetworks enriched for functionally relevant molecular pathways. Due to power issues, integration methods generally do not account directly *trans*-regulatory effects (Siewert-Rocks et al. 2022) and focus on *cis*-regulatory associations that only explain a small fraction of trait heritability (Yao et al. 2020), relying on the coexpression network to provide indirect information on *trans*-regulatory relationships. However, *trans*-regulatory factors that have been shown to carry an important share of the heritability (Westra et al. 2013; Torres et al. 2014; Brynedal et al. 2017), as reflected in the omnigenic model (Boyle et al. 2017; Liu et al. 2019). We previously proposed a way to integrate both *cis*- and *trans*-eQTL results using a bipartite eQTL network representation (Platig et al. 2016; Fagny et al. 2017) 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 (Gaynor et al. 2022). 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 (Fagny et al. 2017, 2019), 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 and among the tissue-specific eQTL networks to better understand both the genetic architecture of complex traits and how they may evolve under directional selection. We used GWAS summary statistics from eleven complex traits and diseases and built 29 tissue-specific *cis*- and *trans*-eQTL networks using the Genotype-Tissue Expression (GTEx) dataset. We then partitioned heritability across various features, including network node topological summary statistics, to identify key determinants of trait heritability. 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 network communities regulating biological functions that explain most of the heritability of each trait. Finally, in order to assess how complex traits may evolve, we looked for signatures of past polygenic adaptive processes in key regulatory loci considered in relationship with their role in the eQTL network. We found that heritability is not scattered uniformly across the genome but rather “clustered” in eQTL modules that represent trait-relevant functions and that loci playing a key role in regulating tissue-specific biological processes were not only more likely to explain trait heritability but also to have been preferential targets of past polygenic adaptation events.

Results

SNPs Associated at the Genome-Wide Level with Traits or Diseases are Clustered in a Few Modules

We wanted to understand how heritability is distributed across the eQTL network. We had previously reported that SNPs significantly associated with chronic obstructive pulmonary disease, cancers, and other traits in GWAS are concentrated in a small number of modules (Platig et al. 2016) and wanted to verify that this result can be extended beyond genome-wide significantly associated SNPs to SNPs explaining a larger part of genetic heritability. We thus used RNA-seq and genotyping data from GTEx to perform eQTL analyses and built 29 tissue-specific weighted bipartite eQTL networks. To compensate for the fact that *trans*-eQTL detection is notoriously underpowered compared with *cis*-eQTL detection, we used a loose false discovery rate (FDR) threshold of 0.2 (see Materials and Methods and [supplementary text, Supplementary Material](#) online). We identified modules containing highly connected SNPs and genes within each network based on CONDOR’s bipartite modularity maximization (Platig et al. 2016) (see Materials and Methods, [supplementary text, Supplementary Material](#) online, [supplementary fig. S1, Supplementary Material](#) online, and [supplementary table S1, Supplementary Material](#) online). We also confirmed that our eQTL networks are summarizing information related to the genetic bases of gene expression coregulation rather than merely gene co-expression profiles (see Materials and Methods, [supplementary text, Supplementary Material](#) online, and [supplementary table S3, Supplementary Material](#) online) and that they are enriched for long-range *cis* 3D chromatin interactions (see Materials and Methods, [supplementary text, Supplementary Material](#) online, and [supplementary table S4, Supplementary Material](#) online).

We considered a total of eleven traits and diseases chosen for their medical or evolutionary relevance in populations of European descent. Breast cancer (BRC), ovarian cancer (OVC), and prostate cancer (PRC), Alzheimer’s disease (ALZ), multiple sclerosis, Schizophrenia (SCZ), high-density lipoprotein 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 is a measure of smoking dependency that is partly explained by genetic factors and is linked to a host of pulmonary and other diseases. 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 TIID, to up to 80% to 90% for SCZ and Alzheimer’s. These traits and diseases also represent a wide range of global genetic heritability, as reported in the literature, from 25% for TIID, to up to 80% to 90% for SCZ and Alzheimer’s. We conditioned on European descent because of the number of GWAS data available and the demographics of GTEx. A complete description of these traits and diseases and their GWAS summary statistics is reported in the [supplementary table S2, Supplementary Material](#) online. For each trait or disease, we obtained summary statistics (see Materials and Methods).

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

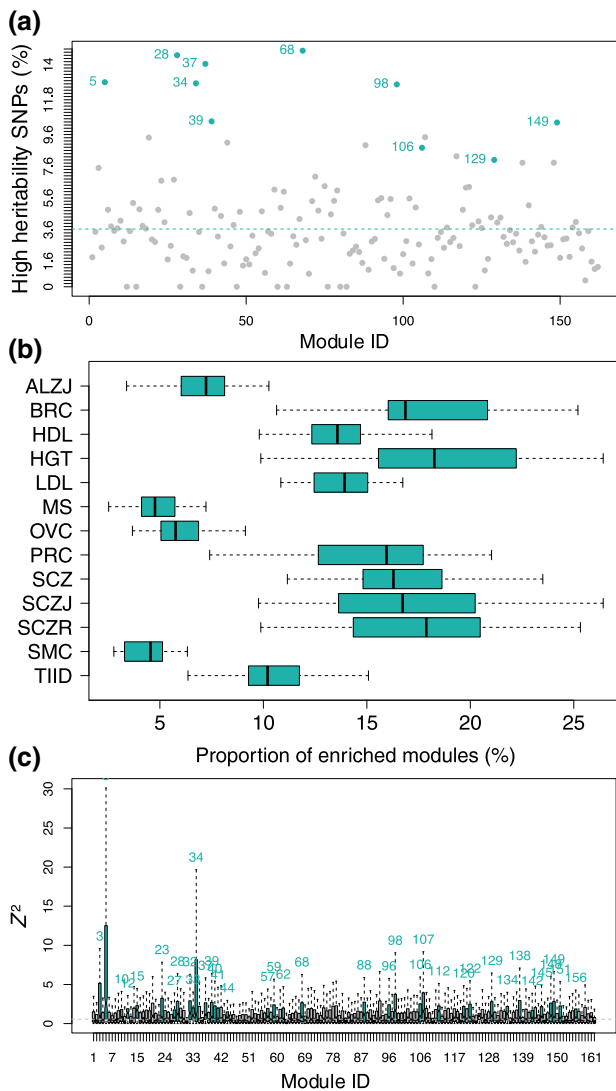


Fig. 1. High heritability SNPs are clustered in a few trait-specific modules. (a) Proportion of high heritability SNPs for BRC in each module of the skin—not sun-exposed (lower leg) eQTL network. Modules significantly enriched for high heritability SNPs are highlighted in color, with module number printed on the left (Benjamini–Hochberg corrected $P \leq 0.01$ using a χ^2 -test). The dotted 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 $P < 2 \times 10^{-16}$). Modules with a distribution skewed towards high values are represented in color, with module number printed on top (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.

(supplementary table S5, Supplementary Material online). As an example, the breast cancer-associated SNPs appear in only 29 (12.6%) of the 230 network modules in the SKN (skin—not sun-exposed—suprapubic) eQTL network (Fig. 1a) representing a substantial concentration of heritability. We found similar results for other diseases and other tissues, with some combinations exhibiting even greater clustering of heritability in a limited number of network modules (Fig. 1b). We validated that these results were not dependent on the method we use to infer the clusters in our networks, because we obtained the same results using a different starting point for the same louvain-clustering based algorithm, and two

other clustering algorithm altogether (see Materials and Methods, supplementary text, Supplementary Material online, and supplementary fig. S2, Supplementary Material online).

Because the top 5% SNPs only explain a small proportion of the heritability of the associated traits (particularly for highly polygenic traits), we examined the distribution of per-SNP heritability across the different eQTL network modules. We found that heritability was 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 ALZ in brain—nucleus accumbens (basal ganglia) networks, which is plotted in Fig. 1c; all the results are in supplementary table S5, Supplementary Material online). 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 (55 [5 to 85] modules representing about 23.8% [5.6 to 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 table S6, Supplementary Material online).

Heritability Is Enriched in Trait-Specific, Functionally Relevant Modules

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 38% [26 to 52] of the modules were not enriched for high heritability SNPs associated with any traits, 28% [22 to 33] were enriched for high heritability SNPs from only one trait and can be considered trait-specific, and 2% [1 to 6] were enriched for high heritability SNPs from at least six of the eleven traits and can be considered as shared across many traits (Fig. 2a and supplementary fig. S3, Supplementary Material online).

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 1,595 pairwise comparisons possible— $\binom{11}{2}$ comparisons \times 29 tissues—only 54 show significant enrichment in overlap compared to what would be expected by chance (Benjamini–Hochberg-corrected $P \leq 0.01$). In more than half of these cases (33), the excess of overlap was observed between HDL and low-density lipoprotein (LDL), HGT, and/or TIID in various tissue-specific networks (see supplementary table S7, Supplementary Material online). This co-occurrence of HDL and LDL is also found in all four ADV network clustering (LCS1, LCS2, FG, and LEC), with $FDR \leq 0.05$ (supplementary table S8, Supplementary Material online).

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 Materials and Methods); the results are presented in supplementary table S9, Supplementary Material online. This allowed us to identify modules that were tissue-specific (see Materials and Methods). The list of tissue-specific modules can change marginally depending on the Jaccard Index cutoff chosen. Based on the distribution of Jaccard indexes for pairwise comparison of tissue-specific networks, we chose two different thresholds: 0.3 and 0.4 (see supplementary fig. S4, Supplementary Material online for all

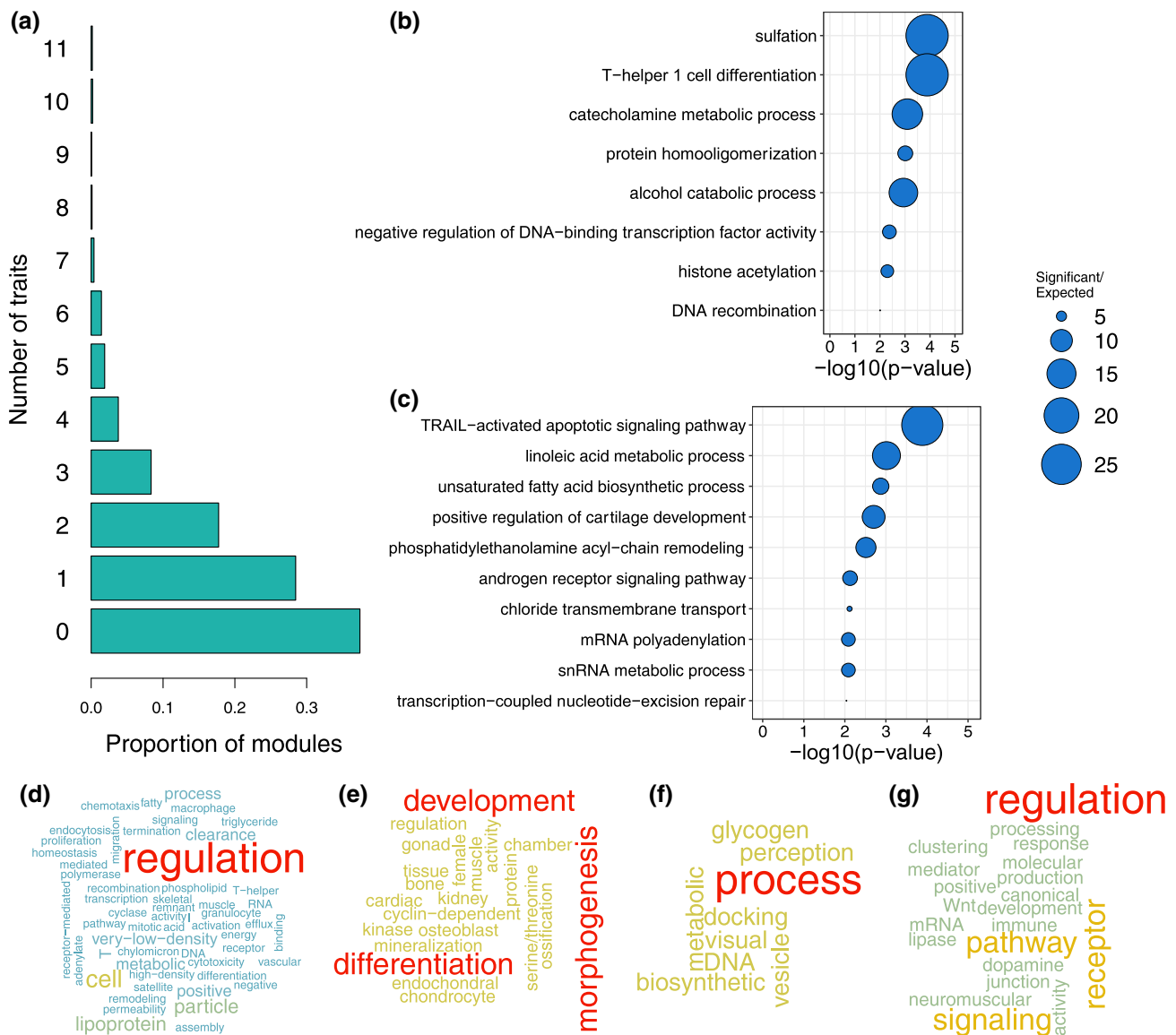


Fig. 2. Heritability is clustered in trait-specific, biologically significant modules. a) Average proportion of modules enriched in high heritability SNPs for one to eleven traits or diseases across all tissue-specific modules. To remove artifacts due to the high genetic correlation between the three 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. S3, Supplementary Material](#) online. b to 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. 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 and c) colon—transverse module 149, enriched for high heritability SNPs for prostate cancer. 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.

comparisons and [supplementary fig. S5, Supplementary Material](#) online for tissue-by-tissue comparisons with the adipose visceral (omentum) network. If choosing a more stringent 0.3 threshold mechanically decreases the number of tissue-specific modules from 2,748 to 1,615, it does not change our main conclusions. We then focused on trait- and tissue-specific modules, defined as those enriched for high heritability SNPs from one to three traits and enriched for tissue-specific GO Terms ([supplementary table S10, Supplementary Material](#) online)—to investigate the specificity of each trait and/or disease heritability and explicit the underlying molecular mechanisms. We found that these modules were enriched in genes involved in biologically and trait-relevant functions.

For example, heritability for ALZ 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. 2b](#)), a class of molecules whose concentrations are altered in symptomatic Alzheimer’s and multiple sclerosis diseases, both of which are amyloid plaque diseases ([Cercignani et al. 2021; Henjum et al. 2022](#)). Heritability for SCZ was most strongly clustered in module 100 of the same tissue, a molecule that is enriched for dopamine receptor signaling pathway genes ([Fig. 2g](#)), and this pathway is known to be functionally disrupted in the brain striatum in SCZ patients ([McCutcheon et al. 2019](#)).

Unsurprisingly, heritability for cancers, including BRC, OVC, and PRC, was enriched in epithelial tissues (skin—not

sun-exposed exposed, skin—sun-exposed, colon—transverse) in network modules consisting of genes enriched for immune response (PRC and module 76 of skin, not sun-exposed), response to cellular hypoxia (BRC, 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), OVC and module 199 and skin, prostate cancer and modules 78, 97 of skin and module 149 of skin—sun-exposed). Particularly interesting, PRC 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 (Johnstone et al. 2008) (Fig. 2c). Breast and OVC heritability are also clustered in modules involved in estrogen metabolism and signaling (BRC and module 180 of skin—sun-exposed, OVC and module 182 of skin—not sun-exposed).

Metabolic traits such as high blood levels of HDL or LDL and TIID tend to cluster in modules enriched for lipid and carbohydrate metabolism. More generally, heritability for HDL and LDL tend to cocluster in several tissue-specific networks, including in adipose–visceral omentum, where almost half of the modules enriched for one trait being also enriched for the other one, the gastro-intestinal networks (CLS-, EGJ, EMC), the heart-related networks (HRA, HRV), the liver network, and the esophagus muscularis and fibroblasts networks. Heritability for high HDL and LDL levels are for example enriched in module 142 of adipose–visceral omentum, enriched for genes involved in very-low-density lipoprotein and HDL particle assembly, remodeling, and clearance (Fig. 2d). They also cocluster in modules enriched for genes involved in lipid metabolism (modules 92 and 197), and cholesterol efflux (module 206). High HDL and LDL heritability can also cluster in different modules enriched for genes belonging to similar pathways: inflammation (modules 4, 133, and 189 for HDL—modules 5, 71, and 128 for LDL) and phosphorylation of STAT protein (module 179 for HDL, 49 for LDL). High HDL heritability is also found in modules related to cholesterol homeostasis (module 189) and ketone metabolism (module 133), while high LDL heritability is found in modules related to lipoprotein lipase activity (module 144), response to stresses (modules 156 and 239) and redox metabolism (modules 57 and 109). These results are robust to the clustering algorithm, in particular the enrichment of both high HDL and high LDL heritability in ADV modules with genes belonging to the lipoprotein metabolic and clearance pathway (module 71 in LCS2; 44 in FG; and modules 1 and 52, with a less significant $P \leq 0.05$ for enrichment of heritability, in LEC, [supplementary tables S11 and S12, Supplementary Material](#) online).

Finally, TIID heritability clusters in module 17 of whole blood, enriched in genes involved in glycogen metabolism (Fig. 2f). Finally, heritability for HGT, 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. 2e).

Local and Global Hubs Carry a Large Part of Trait Heritability

We then explored how heritability is distributed across the SNPs nodes in the eQTL network and whether there are

particular network topological features that correlate with a greater-than-expected ability to explain the heritability of a particular trait. We thus computed two different summary statistics characterizing the topological properties of SNPs within each network: outdegree and core score (Platig et al. 2016). The outdegree measures the centrality of an 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 Materials and Methods), we found for almost every trait we considered that the SNPs explaining the 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. S6, Supplementary Material](#) online for outdegrees and [supplementary fig. S7c, Supplementary Material](#) online for core scores). The enrichment in high heritability among local “hubs” appears stronger and more systematic than for global “hubs”; in artery coronary and liver, the global hubs do not show any significant enrichment at all. Using the adipose visceral (omentum) and the whole blood networks as a model, we found that enrichment of high heritability SNPs in local hubs increases with the threshold chosen to define high outdegree and core score [from the top 90% to the top 5%, [supplementary fig. S8a, Supplementary Material](#) online for whole blood and [supplementary fig. S8c, Supplementary Material](#) online for adipose visceral (omentum)]. This result is also robust to clustering method (see Materials and Methods and [supplementary figs. S8c to S8f, Supplementary Material](#) online and detailed results for threshold = 0.75 in [supplementary fig. S9a, Supplementary Material](#) online).

Given the demonstrated importance of eQTL network topology, we explored whether the increased heritability we found 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 (e.g. global hubs are enriched for non-genic enhancers, while local hubs are enriched for genic enhancers and promoters (Fagny et al. 2017)). For this reason, we partitioned heritability across various functional annotations while accounting for linked markers using stratified LD score regression (Bulik-Sullivan et al. 2015; Finucane et al. 2015) (see Materials and Methods) using the 97-level 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 LDSC software is reported in [supplementary table S2, Supplementary Material](#) online. 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 table S2, Supplementary Material](#) online), but our estimate of the total heritability of traits explained by SNPs as computed by the LD-score regression is coherent with previous reports (see Lindström et al. 2017 for breast, ovarian, and prostate cancer examples).

The LD-score 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. S8b, Supplementary Material online). The detailed results that show enrichment in h^2 among SNPs within each annotation for each of the 11 traits and diseases are reported in supplementary table S13, Supplementary Material online. These results are once again independent of the chosen clustering algorithm (supplementary fig. S9b, Supplementary Material online).

We performed a meta-analysis across the 11 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. 3a. As noted earlier, SNPs that are within the eQTL networks tend to be significantly enriched in h^2 independent of tissue (Fig. 3b). 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. These results are robust to clustering method (see Materials and Methods and supplementary fig. S9b, Supplementary Material online).

Local Hubs Evolve Under Polygenic Adaptation

The clustering of complex traits and diseases heritability in a few modules and in hubs in eQTL networks, together with previous results suggesting that global hubs are likely to evolve under negative selection (Wollenberg Valero 2024) suggest that directional selection on complex traits may preferentially be mediated by local hubs. The importance of the local hubs makes logical sense. Indeed, selection acts at the level of the traits. As each specific trait is likely to be under the control of one or several network modules consisting of genes involved in the same biological function, the local hubs play an important role in their determination. Targeted selection acting on local hubs would moreover mitigate the potential perturbative effects on the network as a whole or other functional modules. To test for this hypothesis, we searched for enrichment among global or local hubs of high scores for different statistics measuring three types of selection signatures. Genomic evolutionary rate profiling (GERP) score detects regions that lack substitutions, which can indicate negative selection (Cooper et al. 2005). Integrated haplotype score (iHS) detects longer-than-expected haplotypes around one of the alleles at one locus, a signature of recent, strong directional selection event or selective sweeps (Voight et al. 2006). F_{ST} measures population differentiation that may reflect older and sometimes milder directional selection events and has proven to be powerful in detecting the shifts in beneficial allele frequencies observed in polygenic selection (Wright 1965; Barghi et al. 2020).

We found that both global and local hubs, while enriched for complex traits and disease heritability, were more likely to be located in constrained genomic regions with high GERP scores in all the tissue-specific networks (see Fig. 4a and supplementary fig. S10a, Supplementary Material online). Global hubs were even more likely than local hubs to

belong to constrained regions (supplementary fig. S11, Supplementary Material online). Both types of hubs are globally evolving under negative selection, in particular global hubs.

Neither local nor global hubs presented a clear pattern of enrichment for high liHSI in all tissue-specific networks (see Fig. 4b and supplementary fig. S10b, Supplementary Material online). However, local hubs were significantly enriched for high liHSI in more tissue-specific networks than global hubs (18 of the 29 networks for local hubs vs. 13 of the 29 networks for global hubs, see Fig. 4b and supplementary fig. S10b, Supplementary Material online). Local hubs were also enriched for high F_{ST} between European and both African and East Asian populations in almost all networks. Conversely, global hubs were depleted for high F_{ST} in almost all networks. Local hubs thus showed more signatures corresponding to polygenic selection than the other SNPs in the network, while global hubs seemed to be evolving under strict negative selection. These results were robust to clustering method (see Materials and Methods and supplementary figs. S12 and S13, Supplementary Material online). Using different thresholds to define high outdegree and core score (from the top 90% to the top 5%), we show that the results obtained reinforce our conclusion on the selection dynamic in the global and local hubs (see Materials and Methods, supplementary text, Supplementary Material online, and supplementary fig. S14, Supplementary Material online).

Discussion

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. 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 traits, in a tissue- or cell-type-specific manner (Platig et al. 2016; Fagny et al. 2017, 2019). 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 determinism that underlie most common traits and provide clues on how they can evolve despite the high degree of pleiotropy. Specifically, the modules in eQTL networks are enriched for functionally related groups of genes and nucleated around “core SNPs” that are both local (module) hubs and the SNPs most likely to exhibit strong GWAS associations with various phenotypes (including diseases) due to their association with module-driven, trait-related processes (Platig et al. 2016; Fagny et al. 2017).

The omnigenic model was important in that it helped bridge the gap between the missing heritability found in many diseases and the growing number of traits for which hundreds, if not thousands, of small effect-size genetic variants contribute to a given trait. Our eQTL network-based model is complementary to the omnigenic model but provides a more nuanced view of the importance of SNP–gene “regulatory” associations in defining phenotypes, identifying disease associations and functions, and explaining both missing heritability and selection. Although it has been reported that global hubs in eQTL networks are enriched for tissue-relevant trait heritability (Gaynor et al. 2022), as are gene modules appearing in various types of networks (Kim et al. 2019), there has not been a

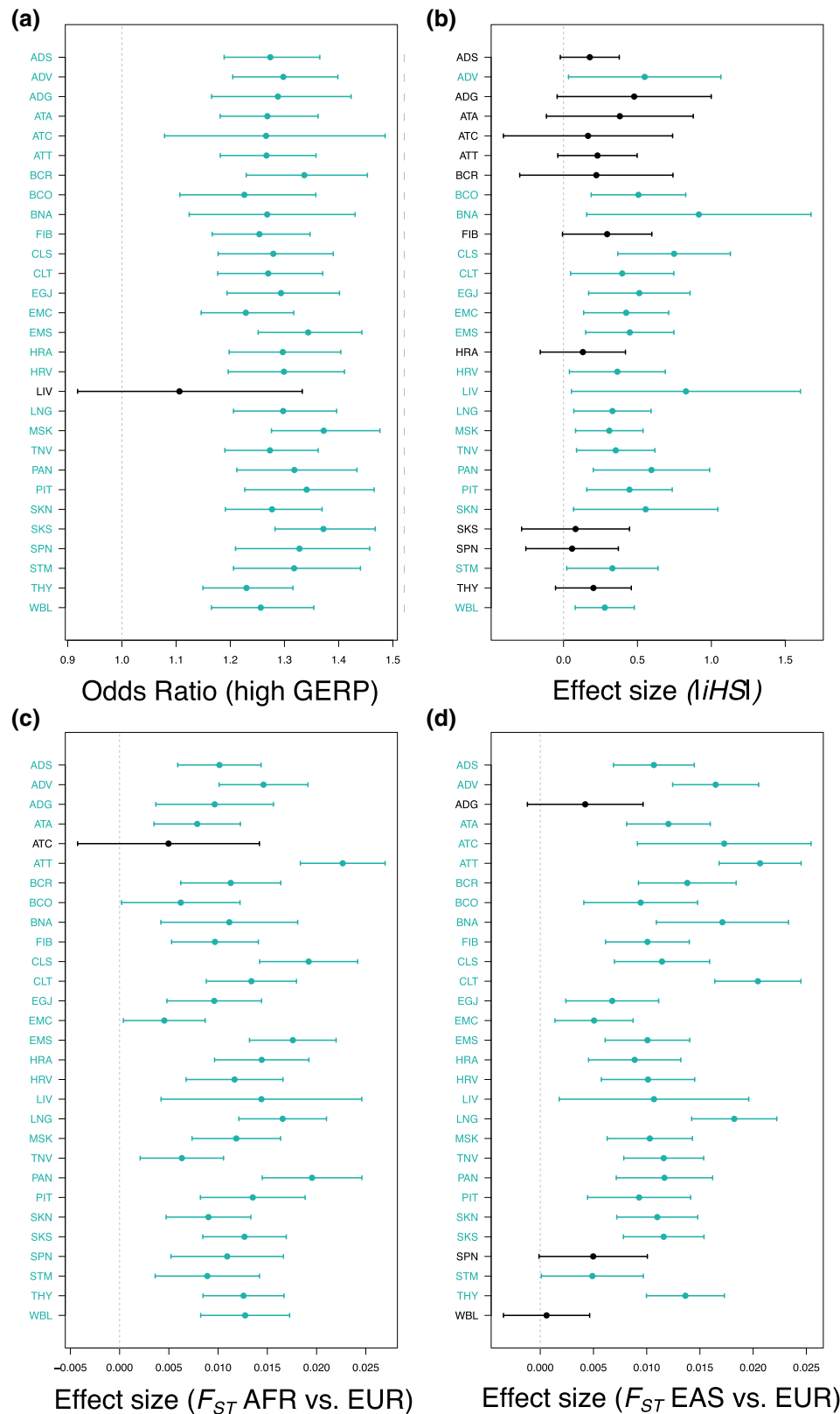


Fig. 4. While overall under negative selection, local hubs are enriched for positive selection signals in almost all tissue-specific eQTL networks. a) Odds ratio of being conserved when the SNP is a local hub vs. a leaf. Bars indicate confidence intervals. Confidence intervals were computed using a logistic regression. b) *iHS* that detect recent selective sweeps signals. c) F_{ST} (AFR, EUR) measuring population differentiation between African (AFR) and European (EUR) samples from 1000 genomes. d) F_{ST} (EAS, EUR) measuring population differentiation between East-Asian (EAS) and European (EUR) samples from 1000 genomes. b to d) Effect size of being a local hub vs a leaf on the value of several statistics sensitive to positive selection. Bars indicate $2 \times$ SE. Standard error (SE) were computed using a linear regression. a to d) For each tissue-specific network, significant enrichment (an odds ratio significantly different from 1 or effect size significantly different from 0) are highlighted in blue. Significance was called when $P \leq 0.05$.

heritability across 29 tissue-specific eQTL networks. We found that although heritability is widely distributed across loci, as suggested by the omnigenic model, the distribution of heritability is far from homogeneous. Instead, there is an uneven distribution with the greatest heritability concentrated in network modules containing genes that represent trait-specific and biologically relevant functions. This makes sense as variations of a given polygenic trait arise through alterations of the expression of specific biological processes relevant to this trait. 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. The only exception are the artery coronary and liver networks, for which this enrichment is not observed. They also behave differently from the other tissue-specific networks in other aspects (tissue-specificity of modules, enrichment of local hubs in negative and polygenic selection signals). This is likely due the relatively small sample size for RNA-seq data in these tissues, leading to smaller eQTL networks, despite their high modularity. These same SNPs, which we had previously shown to be enriched in tissue-specific activated regulatory elements (Fagny et al. 2017), are thus likely to determine a significant proportion of the heritability of complex traits.

The clustering of heritability in modules is also not evenly distributed across traits and differs among tissues. We found that heritability in each phenotype tended 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 (Boyle et al. 2017; Liu et al. 2019) 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.

We also found that global hubs are evolving under strong constraints, as observed in previous studies (Wollenberg Valero 2024). However, our results revealed that another category of nodes, the local hubs that regulate many genes involved in the same biological processes and articulate modules, have been preferential targets of past and ongoing polygenic selection. Because of their strategic position in the eQTL regulatory network, these loci greatly influence the regulation of biological processes in a tissue-specific manner, and, because their effect is limited beyond their “home” module, they are ideal candidates for adaptation. Finally, our results confirm the hypothesis, first presented in an opinion paper by Fagny and Austerlitz (Fagny and Austerlitz 2021), that local hubs are preferential targets for polygenic selection, and provide a means to understand how complex traits may evolve despite the strong pleiotropy that exists in biological networks.

All our results were obtained on bipartite eQTL networks inferred using statistical correlations between individual *cis* and *trans* SNPs to each gene independently. However, other relationships are hidden in our network, including spurious gene co-expression patterns, SNPs-SNPs interactions

caused by linkage disequilibrium, and other SNP-SNP interactions such as epistasis. For the first part, we have demonstrated using comparison with co-expression networks that our approach manage to root out a lot of the spurious co-expression patterns observed in co-expression networks, focusing instead on the portion of gene expression correlated with genetic factors. We also controlled for linkage disequilibrium in our analyses, using two different strategies: (i) treating SNPs within a linkage block as one marker in our downstream analyses and (ii) using the LD-score software, a power tool to account for both SNP linkage and annotations overlaps. Finally, regularized regression models have been recently developed that take these interactions into account. However, the genome-wide application of such methods to a large genome like the human genome one is still limited by computation time, in particular when the eQTL detection is repeated across 29 tissue-specific datasets. Indeed, regularized regression models are generally applied on filtered datasets to limit the number of tests to perform and reduce the computational burden. They are either applied exclusively to *cis*-eQTLs (Marchetti-Bowick et al. 2019; Yan et al. 2020), or applied to heavily filtered and pruned data, such as functionally annotated regions with prior pruning of SNPs for linkage disequilibrium (Banerjee et al. 2021). This may lead to miss some important regulatory SNPs. We consequently arbitrated our pipeline choices in favor of increasing the number of candidates and to the detriment of the potential interactions, and choose a voluntarily loose FDR threshold to balance for the multiple testing issues of *trans*-eQTL detection.

Conclusions

Overall, our results demonstrate a synergy between the eQTL network and omnigenic models in explaining how genetic variants work together to influence traits while addressing 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 disorders. 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. These complex relationships define a conceptual framework for understanding disease risk, helps to define phenotypes for health and disease, and provide a means by which one could potentially prioritize therapeutic targets and design treatment protocols that account for the network architecture. Our findings also allow a better understanding of how complex traits can change as populations adapt to local changes through selection acting on traits mediated at the genetic and molecular level by key regulatory elements. This conceptual framework also provides an explanation, at the molecular level, of how polygenic selection of regulatory mutations beneficial in one environment could lead to the phenomenon of maladaptation that many suspect is the origin of many complex diseases.

Finally, the transferability of these results obtained on human data to other organisms is an open question. One main issue is the impact of the genome structure on the results. In plants genomes, there are many more structural variants than in the human genome (Marroni et al. 2014; Saxena et al. 2014). These variants are more likely to have a greater

impact on the value of the phenotype and are also often associated to adaptation to the environment (Gui et al. 2022; Kang et al. 2023; Li et al. 2023). While they could be integrated in our eQTL network, it is not clear how their presence would affect the conclusions. Keeping in mind these caveats, studying the transferability of these results to other organisms could be of tremendous value to breeders by pointing out which loci are the more likely to improve the tolerance of crops or cattle to environmental stresses without jeopardizing their agricultural values.

Materials and Methods

GTEx Dataset

We used genotyping and gene expression level data from the NHGRI GTEx project version 8.0 (GTEx Consortium 2015). 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 table S1, Supplementary Material](#) online. 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 Gaynor et al. 2022).

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 (Purcell et al. 2007). SNP imputation was then performed using Eagle2 (Loh et al. 2016).

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 Network Inference

eQTLs were obtained from Gaynor et al. (2022). Rapidly, with the R MatrixEQTL package (Shabalin 2012), the association between SNP genotypes and gene expression was modeled using linear regression (Eq. 1) that included potential confounding factors as covariates (Kendzioriski et al. 2006): 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 :

$$\mathbf{G}_j = \mathbf{X}^T \boldsymbol{\alpha} + t_{ij} \mathbf{S}_i. \quad (1)$$

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 (Fig. 5). Each SNP and gene was considered a node in the network. Using a fixed cutoff $q = 0.2$ on the FDR of the eQTL regression (Platig et al. 2016; Fagny et al. 2017), the edge weight a_{ij} between SNP i and gene j was defined by the function

$I_{i,j}\{\text{FDR} < q\}|t_{ij}|$, where t_{ij} is the value of the t statistic computed by the MatrixEQTL package. Thus, when the estimated FDR of the eQTL regression was below the threshold of 0.2, then $a_{ij} = |t_{ij}|$, indicating that there was an edge connecting the nodes, and $a_{ij} = 0$ otherwise.

Overlap with Hi-C Data

3D chromatin loop data stemming from Hi-C experiments for 9 different tissues corresponding to nine of the 29 explored tissues were downloaded from the 3D genome website (see Data Availability). Loops had a resolution of 10 kb, with intrachromosomal loops length spanning between 70 and 2,880 kb. To match our *cis*-eQTL data, we retained only loops spanning a maximum of 1 Mb (on average 99.5% [97.1 to 99.9] of all loops, corresponding to 13,775 [7,812 to 18,749] loops), and for which one side of the loop's 10 kb block was containing a gene TSS. This filtered dataset corresponded to our “null” dataset.

For this analysis, we merged eQTLs by blocks of linkage disequilibrium (with an FDR cutoff of 0.2 for SNP–gene association significance, we retained an average of 749,672 [403,759 to 1,104,908] significant block–gene associations. We retrieved overlaps of Hi-C loops and eQTL data, keeping only cases where the 10-kb block containing the eQTL's LD block was on one side of one loop and the 10-kb block containing the TSS of the associated gene was on the other side of the same loop (on average 646 [304 to 1,068]). We then used a χ^2 test to compare proportions of Hi-C loops in LD blocks that carried significant eQTL and in LD blocks that did not.

Network Summary Statistics

The module structure of each tissue-specific network was determined using the bipartite modularity maximization approach (Platig et al. 2016) implemented in the netZooR Bioconductor package. This is a two-step approach : (i) project the eQTL network on the gene space and infer a first cluster structure using one of the three available algorithms : louvain clustering (LCS), leading eigenvector community detection (LEC), and the fastgreedy community detection (FG); and (ii) improve the clustering by adding SNPs back in the network and searching for improved modularity by reattributing SNPs to modules to obtain the best modularity possible and then genes. The second step is iterated as many times as necessary to reach convergence. We tested all three algorithms on step (i) to ensure that our results are robust to clustering method and implemented a new version of the second step of the algorithm (CONDOR:condorSplitMatrixModularity) that allows for a balance between computation time and memory usage (Fig. 5) and has been published in the R netzooR package fork of <https://github.com/maudf/>.

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

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 (Platig et al. 2016). 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 $d_j = \sum_{i=1}^s I_{i,j}\{\text{FDR} < q\}|t_{ij}|$, then, for SNP i in module h , its core score, Q_{ih} , is defined by Eq. 3:

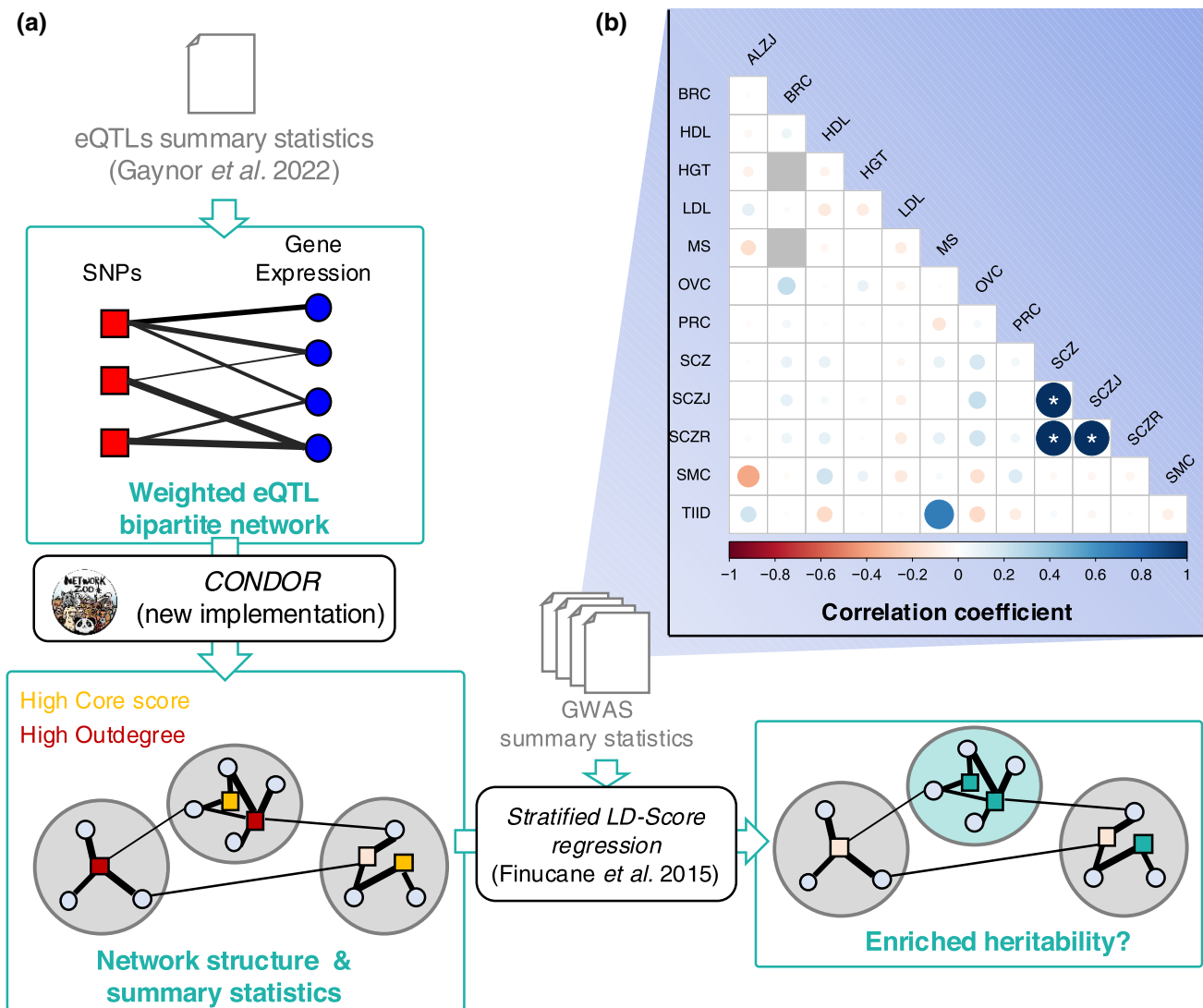


Fig. 5. Pipeline and traits correlation. a) Pipeline of data analyses. Input data are in grey. GTEx eQTL summary statistics were obtained from a previous study (Gaynor et al. 2022) and are described in Materials and Methods and [supplementary table S1, Supplementary Material](#) online. GWAS summary statistics were downloaded from <https://console.cloud.google.com/storage/browser/broad-alkesgroup-public-requester-pays> and are described in [supplementary table S2, Supplementary Material](#) online. eQTL network summary statistics are presented in [supplementary fig. S1, Supplementary Material](#) online. b) Pairwise genetic correlations between traits based on GWAS summary statistics.

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

GWAS Data

GWAS data were obtained from the Alkes group (see [supplementary table S2, Supplementary Material](#) online). We considered a total of eleven traits and diseases presenting varying levels of estimated genetic heritability. For each trait or disease, we obtained summary statistics including SNP chromosome, position, alleles 1 and 2, χ^2 , and Z-score. Here, the relationship between an individual's i genotype at SNP s and its phenotype Y , accounting for covariates M , is usually tested using the following regression: $Y_{i,s} = \beta_{1,s} X_{i,s} + \beta_2 M_i + \epsilon_{i,s}$, where $\epsilon_{i,s}$ is the error term. Then, $Z = \beta_{1,s} / \text{se}(\beta_{1,s})$ and $\chi^2 = q \text{norm}((P/2)^2)$ where $P = 2 \times p \text{norm}(-|Z|)$.

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_{ib}) for high and nonhigh 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_{ib} 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:

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

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

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 [Gazal et al. \(2018\)](#) and [Hujuel et al. \(2019\)](#) for each of the eleven traits and diseases selected above ([Fig. 5](#)). 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 ([Gazal et al. 2018](#); [Hujuel et al. 2019](#)). 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_{ib} 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-jackknife.

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 eleven 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.

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 nonhigh 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 ([Alexa and Rahnenfuhrer 2016](#)), using the `elim` method; this method is more efficient than Fisher's exact test ([Alexa et al. 2006](#)). 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` `bioc` `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 ≤ 3 , we considered it a tissue-specific module. To evaluate the impact of the jaccard index threshold, we also investigated a threshold of 0.4.

Selection Scores Enrichment Analyses

Conserved SNPs were determined by intersecting the annotations of the SNPs present in our eQTL networks with conserved regions determined using GERP scores downloaded from Ensembl 111 release (https://ftp.ensembl.org/pub/release-111/bed/ensembl-compara/91_mammals.gerp_constrained_element/gerp_constrained_elements.homo_sapiens.bb [Martin et al. 2023](#)). Enrichment in conserved SNPs among local and/or global hubs were computed using the following logistic regression model, where $I(GERP = 1)$ is an indicator function equal to 1 if the SNP falls in a conserved region and equal to 0 otherwise, $I(Cat = 1)$ is an indicator function equal to 1 if the SNP belongs to a given category (local or global hubs) 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:

$$\text{Logit}(I(GERP_i = 1)) \sim I(Cat_i = 1) + \left[\sum_{k=1}^{n-1} I(C_{ik} = 1) \right] + \epsilon$$

iHS, which measures the differences in haplotype lengths between haplotypes carrying the ancestral and the derived SNPs, detect signatures of recent selective sweeps ([Voight et al. 2005](#)). We downloaded iHS standardized scores ([Johnson and Voight 2018](#)) computed for 26 populations for 1000 genomes project Phase 3 dataset from <https://zenodo.org/records/7842512>. We extracted scores computed on the CEU population (Utah residents (CEPH) with Northern and Western European ancestry) samples as being the closest population from the European descent samples that represent 85% of the GTEx dataset v8 release.

F_{ST} between African, Eurasian, and European samples from the 1000 genomes data were computed using `plink2.0` ([Purcell et al. 2007](#)) on genotypic data for samples from the EUR, AFR, and EAS populations from the 1000 genomes Phase 3 dataset. Genotyping data were downloaded from ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000G_2504_high_coverage/working/20201028_3202_raw_GT_with_annot/.

Effect sizes of being a local or global hub on iHS or F_{ST} were computed using the following linear regression model, where $Score_i$ is the value of the statistics for SNP i , $I(Cat = 1)$ is an indicator function equal to 1 if the SNP belongs to a given category (local or global hubs) and equal to 0 otherwise, and $I(C_k = 1)$ is an indicator function equal to 1 if the SNP i belongs to module k and equal to 0 otherwise:

$$Score_i \sim I(Cat_i = 1) + \left[\sum_{k=1}^{n-1} I(C_{ik} = 1) \right] + \epsilon_i$$

The community ID terms were only added when local hubs were one of the two categories compared, as the q -score determining local hubs depends on community ID.

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online.

Acknowledgments

Thank you to Alkes Price for discussions about LDSC and heritability and for making the GWAS summary statistics data available. 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 National Human Genome Research Institute (J.Q.: R01HG011393); and the Marie Skłodowska-Curie grant PATTERNS (M.F.: 845083).

Author Contributions

K.L.S., J.P., J.Q., and M.F. contributed to the conception of the work. K.L.S., J.P., J.Q., and M.F. contributed to the study design and method development. K.L.S., J.P., and M.F. contributed analysis, verified the data, and drafted the manuscript. All authors reviewed the manuscript and approved the submitted work.

Ethics Approval

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

Conflict of Interest

The authors declare no competing interests.

Data 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 17 December 2019 and dbGaP accession number phs000424.v8 on 12/17 19 for RNA-seq and Genotyping data, from <https://console.cloud.google.com/storage/browser/>

broad-alkesgroup-public-requester-pays for GWAS data, and from <https://3dgenome.fsm.northwestern.edu/downloads/loops-hg38.zip> on 01/02/2025 for the Hi-C data.

References

- Alexa A, Rahnenführer J. *topGO: enrichment analysis for gene ontology*. R package version v.2.44; 2016. <https://doi.org/10.18129/B9.bioc.topGO>.
- Alexa A, Rahnenführer J, Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*. 2006;22(13):1600–1607. <https://doi.org/10.1093/bioinformatics/btl140>.
- Banerjee S, Simonetti FL, Detrois KE, Kaphle A, Mitra R, Nagial R, Söding J. Tejaas: reverse regression increases power for detecting trans-eQTLs. *Genome Biol*. 2021;22(1):142. <https://doi.org/10.1186/s13059-021-02361-8>.
- Barghi N, Hermisson J, Schlötterer C. Polygenic adaptation: a unifying framework to understand positive selection. *Nat Rev Genet*. 2020;21:769–781. <https://doi.org/10.1038/s41576-020-0250-z>.
- Barreiro LB, Quintana-Murci L. From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nat Rev Genet*. 2010;11(1):17–30. <https://doi.org/10.1038/nrg2698>.
- Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. *Cell*. 2017;169(7):1177–1186. <https://doi.org/10.1016/j.cell.2017.05.038>.
- Brynedal B, Choi J, Raj T, Bjornson R, Stranger BE, Neale BM, Voight BF, Cotsapas C. Large-scale trans-eQTLs affect hundreds of transcripts and mediate patterns of transcriptional co-regulation. *Am J Hum Genet*. 2017;100(4):581–591. <https://doi.org/10.1016/j.ajhg.2017.02.004>.
- Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Patterson N, Daly MJ, Price AL, Neale BM. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47(3):291–295. <https://doi.org/10.1038/ng.3211>.
- Calabrese GM, Mesner LD, Stains JP, Tommasini SM, Horowitz MC, Rosen CJ, Farber CR. Integrating GWAS and co-expression network data identifies bone mineral density genes SPTBN1 and MARK3 and an osteoblast functional module. *Cell Syst*. 2017;4(1):46–59.e4. <https://doi.org/10.1016/j.cels.2016.10.014>.
- Cercignani M, Dipasquale O, Bogdan I, Carandini T, Scott J, Rashid W, Sabri O, Hesse S, Rullmann M, Lopiano L, et al. Cognitive fatigue in multiple sclerosis is associated with alterations in the functional connectivity of monoamine circuits. *Brain Commun*. 2021;3(2):fcab023. <https://doi.org/10.1093/braincomms/fcab023>.
- Chan EKF, Rowe HC, Corwin JA, Joseph B, Kliebenstein DJ. Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in *Arabidopsis thaliana*. *PLoS Biol*. 2011;9(8):e1001125. <https://doi.org/10.1371/journal.pbio.1001125>.
- Cooper GM, Stone EA, Asimenos G, Green ED, Batzoglou S, Sidow A. Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res*. 2005;15(7):901–913. <https://doi.org/10.1101/gr.3577405>.
- Daub JT, Hofer T, Cutivet E, Dupanloup I, Quintana-Murci L, Robinson-Rechavi M, Excoffier L. Evidence for polygenic adaptation to pathogens in the human genome. *Mol Biol Evol*. 2013;30(7):1544–1558. <https://doi.org/10.1093/molbev/mst080>.
- Fagny M, Austerlitz F. Polygenic adaptation: integrating population genetics and gene regulatory networks. *Trends Genet*. 2021;37(7):631–638. <https://doi.org/10.1016/j.tig.2021.03.005>.
- Fagny M, Paulson JN, Kuijjer ML, Sonawane AR, Chen C-Y, Lopes-Ramos CM, Glass K, Quackenbush J, Platig J. Exploring regulation in tissues with eQTL networks. *PNAS*. 2017;114(37):E7841–E7850. <https://doi.org/10.1073/pnas.1707375114>.
- Fagny M, Platig J, Kuijjer ML, Lin X, Quackenbush J. Nongenic cancer-risk SNPs affect oncogenes, tumour-suppressor genes, and immune function. *Br J Cancer*. 2019;122:569–577. <https://doi.org/10.1038/s41416-019-0614-3>.

- Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R, Anttila V, Xu H, Zang C, Farh K, *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet.* 2015;47(11):1228–1235. <https://doi.org/10.1038/ng.3404>.
- Gamazon ER, Segrè AV, van de Bunt M, Wen X, Xi HS, Hormozdiari F, Ongen H, Konkashbaev A, Derks EM, Aguet F, *et al.* Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nat Genet.* 2018;50(7):956–967. <https://doi.org/10.1038/s41588-018-0154-4>.
- Gaynor SM, Fagny M, Lin X, Platig J, Quackenbush J. Connectivity in eQTL networks dictates reproducibility and genomic properties. *Cell Rep Methods.* 2022;2(5):100218. <https://doi.org/10.1016/j.crmeth.2022.100218>.
- Gazal S, Loh P-R, Finucane HK, Ganna A, Schoech A, Sunyaev S, Price AL. Functional architecture of low-frequency variants highlights strength of negative selection across coding and non-coding annotations. *Nat Genet.* 2018;50(11):1600–1607. <https://doi.org/10.1038/s41588-018-0231-8>.
- GTEX Consortium. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science.* 2015;348(6235):648–660. <https://doi.org/10.1126/science.1262110>.
- Gui S, Wei W, Jiang C, Luo J, Chen L, Wu S, Li W, Wang Y, Li S, Yang N, *et al.* A pan-zea genome map for enhancing maize improvement. *Genome Biol.* 2022;23(1):178. <https://doi.org/10.1186/s13059-022-02742-7>.
- Hallgrímsson B, Miao W, Marcucio RS, Spritz R. Let's face it—complex traits are just not that simple. *PLoS Genet.* 2014;10(11):e1004724. <https://doi.org/10.1371/journal.pgen.1004724>.
- He F, Arce AL, Schmitz G, Koornneef M, Novikova P, Beyer A, de Meaux J. The footprint of polygenic adaptation on stress-responsive *cis*-regulatory divergence in the *arabidopsis* genus. *Mol Biol Evol.* 2016;33(8):2088–2101. <https://doi.org/10.1093/molbev/msw096>.
- Henjum K, Watne LO, Godang K, Halaas NB, Eldholm RS, Blennow K, Zetterberg H, Saltvedt I, Bollerslev J, Knapskog AB. Cerebrospinal fluid catecholamines in Alzheimer's disease patients with and without biological disease. *Transl Psychiatry.* 2022;12(1):151. <https://doi.org/10.1038/s41398-022-01901-5>.
- Hormozdiari F, Gazal S, van de Geijn B, Finucane HK, Ju CJ-T, Loh P-R, Schoech A, Reshef Y, Liu X, O'Connor L, *et al.* Leveraging molecular quantitative trait loci to understand the genetic architecture of diseases and complex traits. *Nat Genet.* 2018;50(7):1041–1047. <https://doi.org/10.1038/s41588-018-0148-2>.
- Hormozdiari F, van De Bunt M, Segre A, Li X, Joo JWJ, Bilow M, Sul JH, Sankararaman S, Pasaniuc B, Eskin E. Colocalization of GWAS and eQTL signals detects target genes. *Am J Hum Genet.* 2016;99(6):1245–1260. <https://doi.org/10.1016/j.ajhg.2016.10.003>.
- Hujoel ML, Gazal S, Hormozdiari F, Van De Geijn B, Price AL. Disease heritability enrichment of regulatory elements is concentrated in elements with ancient sequence age and conserved function across species. *Am J Hum Genet.* 2019;104(4):611–624. <https://doi.org/10.1016/j.ajhg.2019.02.008>.
- Jakobson CM, She R, Jarosz DF. Pervasive function and evidence for selection across standing genetic variation in *S. cerevisiae*. *Nat Commun.* 2019;10(1):1222. <https://doi.org/10.1038/s41467-019-09166-1>.
- Johnson KE, Voight BF. Patterns of shared signatures of recent positive selection across human populations. *Nat Ecol Evol.* 2018;2(4):713–720. <https://doi.org/10.1038/s41559-018-0478-6>.
- Johnstone RW, Frew AJ, Smyth MJ. The TRAIL apoptotic pathway in cancer onset, progression and therapy. *Nat Rev Cancer.* 2008;8(10):782–798. <https://doi.org/10.1038/nrc2465>.
- Kang M, Wu H, Liu H, Liu W, Zhu M, Han Y, Liu W, Chen C, Song Y, Tan L, *et al.* The pan-genome and local adaptation of *Arabidopsis thaliana*. *Nat Commun.* 2023;14(1):6259. <https://doi.org/10.1038/s41467-023-42029-4>.
- Kendziorski CM, Chen M, Yuan M, Lan H, Attie AD. Statistical methods for expression quantitative trait loci (eQTL) mapping. *Biometrics.* 2006;62(1):19–27. <https://doi.org/10.1111/biom.2006.62.issue-1>.
- Kim SS, Dai C, Hormozdiari F, van de Geijn B, Gazal S, Park Y, O'Connor L, Amariuta T, Loh P-R, Finucane H, *et al.* Genes with high network connectivity are enriched for disease heritability. *Am J Hum Genet.* 2019;104(5):896–913. <https://doi.org/10.1016/j.ajhg.2019.03.020>.
- Lee I, Blom UM, Wang PI, Shim JE, Marcotte EM. Prioritizing candidate disease genes by network-based boosting of genome-wide association data. *Genome Res.* 2011;21(7):1109–1121. <https://doi.org/10.1101/gr.118992.110>.
- Li N, He Q, Wang J, Wang B, Zhao J, Huang S, Yang T, Tang Y, Yang S, Aisimutuola P, *et al.* Super-pangenome analyses highlight genomic diversity and structural variation across wild and cultivated tomato species. *Nat Genet.* 2023;55(5):852–860. <https://doi.org/10.1038/s41588-023-01340-y>.
- Lindström S, Finucane H, Bulik-Sullivan B, Schumacher FR, Amos CI, Hung RJ, Rand K, Gruber SB, Conti D, Permut JB, *et al.* Quantifying the genetic correlation between multiple cancer types. *Cancer Epidemiol Biomarkers Prev.* 2017;26(9):1427–1435. <https://doi.org/10.1158/1055-9965.EPI-17-0211>.
- Liu X, Li YI, Pritchard JK. Trans effects on gene expression can drive omnigenic inheritance. *Cell.* 2019;177(4):1022–1034.e6. <https://doi.org/10.1016/j.cell.2019.04.014>.
- Loh P-R, Danecek P, Palamara PF, Fuchsberger C, Reshef YA, Finucane HK, Schoenherr S, Forer L, McCarthy S, Abecasis GR, *et al.* Reference-based phasing using the haplotype reference consortium panel. *Nat Genet.* 2016;48(11):1443–1448. <https://doi.org/10.1038/ng.3679>.
- Luningham JM, Chen J, Tang S, De Jager PL, Bennett DA, Buchman AS, Yang J. Bayesian genome-wide TWAS method to leverage both *cis*- and *trans*-eQTL information through summary statistics. *Am J Hum Genet.* 2020;107(4):714–726. <https://doi.org/10.1016/j.ajhg.2020.08.022>.
- Marchetti-Bowick M, Yu Y, Wu W, Xing EP. A penalized regression model for the joint estimation of eQTL associations and gene network structure. *Ann Appl Stat.* 2019;13(1):248–270. <https://doi.org/10.1214/18-AOAS1186>.
- Marroni F, Pinosio S, Morgante M. Structural variation and genome complexity: is dispensable really dispensable? *Genome Stud Mol Genet.* 2014;18:31–36. <https://doi.org/10.1016/j.pbi.2014.01.003>.
- Martin FJ, Amode MR, Aneja A, Austin-Orimoloye O, Azov AG, Barnes I, Becker A, Bennett R, Berry A, Bhai J, *et al.* Ensembl 2023. *Nucleic Acids Res.* 2023;51(D1):D933–D941. <https://doi.org/10.1093/nar/gkac958>.
- Mathieson I. The omnigenic model and polygenic prediction of complex traits. *Am J Hum Genet.* 2021;108(9):1558–1563. <https://doi.org/10.1016/j.ajhg.2021.07.003>.
- McCutcheon RA, Abi-Dargham A, Howes OD. Schizophrenia, dopamine and the striatum: from biology to symptoms. *Trends Neurosci.* 2019;42(3):205–220. <https://doi.org/10.1016/j.tins.2018.12.004>.
- Pais TM, Foulquié-Moreno MR, Hubmann G, Duitama J, Swinnen S, Goovaerts A, Yang Y, Dumortier F, Thevelein JM, Jinks-Robertson S. Comparative polygenic analysis of maximal ethanol accumulation capacity and tolerance to high ethanol levels of cell proliferation in yeast. *PLoS Genet.* 2013;9(6):e1003548. <https://doi.org/10.1371/journal.pgen.1003548>.
- Peiffer JA, Romay MC, Gore MA, Flint-Garcia SA, Zhang Z, Millard MJ, Gardner CAC, McMullen MD, Holland JB, Bradbury PJ, *et al.* The genetic architecture of maize height. *Genetics.* 2014;196(4):1337–1356. <https://doi.org/10.1534/genetics.113.159152>.
- Platig J, Castaldi PJ, DeMeo D, Quackenbush J. Bipartite community structure of eQTLs. *PLoS Comput Biol.* 2016;12(9):e1005033. <https://doi.org/10.1371/journal.pcbi.1005033>.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, *et al.* PLINK: a tool set for whole-genome association and population-based linkage

- analyses. *Am J Hum Genet.* 2007;81(3):559–575. <https://doi.org/10.1086/519795>.
- Salomé PA, Bomblies K, Laitinen RAE, Yant L, Mott R, Weigel D. Genetic architecture of flowering-time variation in *Arabidopsis thaliana*. *Genetics.* 2011;188(2):421. <https://doi.org/10.1534/genetics.111.126607>.
- Saxena RK, Edwards D, Varshney RK. Structural variations in plant genomes. *Brief Funct Genomics.* 2014;13(4):296–307. <https://doi.org/10.1093/bfpg/elu016>.
- Schaefer RJ, Briskine R, Springer NM, Myers CL. Discovering functional modules across diverse maize transcriptomes using COB, the co-expression browser. *PLoS One.* 2014;9(6):e99193. <https://doi.org/10.1371/journal.pone.0099193>.
- Schaefer RJ, Michno J-M, Jeffers J, Hoekenga O, Dilkes B, Baxter I, Myers CL. Integrating coexpression networks with GWAS to prioritize causal genes in maize. *Plant Cell.* 2018;30(12):2922–2942. <https://doi.org/10.1105/tpc.18.00299>.
- Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics.* 2012;28:1353–1358. <https://doi.org/10.1093/bioinformatics/bts163>.
- Shi H, Kichaev G, Pasaniuc B. Contrasting the genetic architecture of 30 complex traits from summary association data. *Am J Hum Genet.* 2016;99(1):139–153. <https://doi.org/10.1016/j.ajhg.2016.05.013>.
- Siewert-Rocks KM, Kim SS, Yao DW, Shi H, Price AL. Leveraging gene co-regulation to identify gene sets enriched for disease heritability. *Am J Hum Genet.* 2022;109(3):393–404. <https://doi.org/10.1016/j.ajhg.2022.01.005>.
- Sonawane AR, Weiss ST, Glass K, Sharma A. Network medicine in the age of biomedical big data. *Front Genet.* 2019;10. <https://doi.org/10.3389/fgene.2019.00294>.
- Tak YG, Farnham PJ. 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.* 2015;8(1):57. <https://doi.org/10.1186/s13072-015-0050-4>.
- Taşan M, Musso G, Hao T, Vidal M, MacRae CA, Roth FP. Selecting causal genes from genome-wide association studies via functionally coherent subnetworks. *Nat Methods.* 2015;12(2):154–159. <https://doi.org/10.1038/nmeth.3215>.
- Torres JM, Gamazon ER, Parra EJ, Below JE, Valladares-Salgado A, Wacher N, Cruz M, Hanis CL, Cox NJ. Cross-tissue and tissue-specific eQTLs: partitioning the heritability of a complex trait. *Am J Hum Genet.* 2014;95(5):521–534. <https://doi.org/10.1016/j.ajhg.2014.10.001>.
- Voight BF, Adams AM, Frishe LA, Qian Y, Hudson RR, Di Rienzo A. Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. *Proc Natl Acad Sci U S A.* 2005;102(51):18508–18513. <https://doi.org/10.1073/pnas.0507325102>.
- Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS Biol.* 2006;4(3):e72. <https://doi.org/10.1371/journal.pbio.0040072>.
- Ward LD, Kellis M. Evidence of abundant purifying selection in humans for recently acquired regulatory functions. *Science.* 2012;337(6102):1675–1678. <https://doi.org/10.1126/science.1225057>.
- Westra H-J, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, Christiansen MW, Fairfax BP, Schramm K, Powell JE, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet.* 2013;45(10):1238–1243. <https://doi.org/10.1038/ng.2756>.
- Wollenberg Valero KC. Brief communication: the predictable network topology of evolutionary genomic constraint. *Mol Biol Evol.* 2024;41(3):msae033. <https://doi.org/10.1093/molbev/msae033>.
- Wright S. The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution.* 1965;19(3):395. <https://doi.org/10.2307/2406450>.
- Yan KK, Zhao H, Wu JT, Pang H. An enhanced machine learning tool for cis-eQTL mapping with regularization and confounder adjustments. *Genet Epidemiol.* 2020;44(8):798–810. <https://doi.org/10.1002/gepi.v44.8>.
- Yao DW, O'Connor LJ, Price AL, Gusev A. Quantifying genetic effects on disease mediated by assayed gene expression levels. *Nat Genet.* 2020;52(6):626–633. <https://doi.org/10.1038/s41588-020-0625-2>.
- Zan Y, Carlborg Ö. A polygenic genetic architecture of flowering time in the worldwide *Arabidopsis thaliana* population. *Mol Biol Evol.* 2019;36(1):141–154. <https://doi.org/10.1093/molbev/msy203>.
- Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, Montgomery GW, Goddard ME, Wray NR, Visscher PM, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet.* 2016;48(5):481. <https://doi.org/10.1038/ng.3538>.