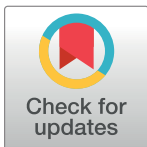


## RESEARCH ARTICLE

## Integrating comprehensive functional annotations to boost power and accuracy in gene-based association analysis

Corbin Quick<sup>1,2†\*</sup>, Xiaoquan Wen<sup>2</sup>, Gonçalo Abecasis<sup>1,3</sup>, Michael Boehnke<sup>1</sup>, Hyun Min Kang<sup>1\*</sup>**1** Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA, **2** Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA, USA, **3** Regeneron Genetics Center, Regeneron Pharmaceuticals, Tarrytown, NY, USA

† This work was primarily completed while at the University of Michigan, Ann Arbor.

\* [qcorbin@hsph.harvard.edu](mailto:qcorbin@hsph.harvard.edu) (CQ); [hmkang@umich.edu](mailto:hmkang@umich.edu) (HMK)

## OPEN ACCESS

**Citation:** Quick C, Wen X, Abecasis G, Boehnke M, Kang HM (2020) Integrating comprehensive functional annotations to boost power and accuracy in gene-based association analysis. *PLoS Genet* 16(12): e1009060. <https://doi.org/10.1371/journal.pgen.1009060>**Editor:** Vincent Plagnol, University College London, UNITED KINGDOM**Received:** October 29, 2019**Accepted:** August 18, 2020**Published:** December 15, 2020**Copyright:** © 2020 Quick et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.**Data Availability Statement:** All GWAS and annotation data analysed in this manuscript are publicly available online. GWAS and genotype data sets are available from the following URLs: UK Biobank SAIGE Summary Statistics: ([ftp://share.sph.umich.edu/UKBB\\_SAIGE\\_HRC/](ftp://share.sph.umich.edu/UKBB_SAIGE_HRC/)) 1000 Genomes Project Data: (<https://www.internationalgenome.org/category/data-access/>) Regulatory element annotations and eQTL predictive weights are available from the following URLs: PredictDB: (<http://predictdb.org/>) TWAS/FUSION: (<http://gusevlab.org/projects/fusion/>)

## Abstract

Gene-based association tests aggregate genotypes across multiple variants for each gene, providing an interpretable gene-level analysis framework for genome-wide association studies (GWAS). Early gene-based test applications often focused on rare coding variants; a more recent wave of gene-based methods, e.g. TWAS, use eQTLs to interrogate regulatory associations. Regulatory variants are expected to be particularly valuable for gene-based analysis, since most GWAS associations to date are non-coding. However, identifying causal genes from regulatory associations remains challenging and contentious. Here, we present a statistical framework and computational tool to integrate heterogeneous annotations with GWAS summary statistics for gene-based analysis, applied with comprehensive coding and tissue-specific regulatory annotations. We compare power and accuracy identifying causal genes across single-annotation, omnibus, and annotation-agnostic gene-based tests in simulation studies and an analysis of 128 traits from the UK Biobank, and find that incorporating heterogeneous annotations in gene-based association analysis increases power and performance identifying causal genes.

## Author summary

Gene-based association tests are statistical methods used in genome-wide association studies (GWAS) to identify genes that affect heritable traits. Gene-based tests are formed by aggregating genotypes across multiple genetic variants for each gene, often including only variants that are likely to affect gene function or regulation. In this work, we present a unified framework to integrate heterogeneous classes of functional variants in gene-based association analysis. This approach enables us to simultaneously assess multiple distinct biological mechanisms underlying GWAS association signals, and to construct powerful omnibus tests by aggregating across functional classes for each gene. We evaluated the performance of gene-based association test methods and strategies to identify causal genes by conducting extensive simulation studies, and by analyzing 128 human traits

[#reference-functional-data](#)) RoadmapLinks: ([www.biolchem.ucla.edu/labs/ernst/roadmaplinking](http://www.biolchem.ucla.edu/labs/ernst/roadmaplinking)) JEME: (<http://yiplab.cse.cuhk.edu.hk/jeme/>) GeneHancer: (<https://www.genecards.org/>) The GAMBIT software is open source and available at (<https://github.com/corbinq/GAMBIT>).

**Funding:** This work was supported by National Institutes of Health (NIH) grants U01HL137182 (PI: HMK) and HG009976 (PI: MB). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** Goncalo Abecasis is an employee of Regeneron Pharmaceuticals. We declare no other conflicts of interest.

from the UK Biobank and comparing our results against lists of high-confidence putative causal genes. Our analysis suggests that incorporating heterogeneous functional variants in gene-based association tests increases power to detect gene-based association and helps identify causal genes.

## Introduction

Genome-wide association studies (GWAS) have identified thousands of genetic loci associated with complex traits [1]; however, the biological mechanisms underlying these associations are often poorly understood. Gene-based association tests can provide a more interpretable analysis framework compared to single-variant analysis, interrogating association at the gene level by aggregating genotypes across multiple variants for each gene. This strategy can also increase power to detect association by aggregating small effects across variants, reducing the burden of multiple testing, and weighting or filtering to prioritize functional variants [2, 3].

In gene-based analysis, variants are often grouped or weighted by putative functional effect, for example, a common strategy for exome analysis is to include only rare non-synonymous or loss-of-function (LoF) variants in gene-based tests such as SKAT and the CMC burden test [4, 5]. A more recent wave of gene-based methods, e.g. PrediXcan [6, 7] and TWAS [8], use eQTL variants to construct gene-based tests of association between the predicted genetic component of gene expression and GWAS trait. Incorporating regulatory variants is expected to be particularly valuable for gene-based analysis of complex traits, since most genetic associations discovered to date are in non-coding regions [9]. However, while coding variants generally implicate a single known gene, the gene(s) affected by regulatory variants are often less clear [10, 11].

Incorporating multiple types of annotation in gene-based analysis provides several advantages over analysis methods using annotations of a single type. First, including variants from multiple annotation categories is expected to increase accuracy (e.g., odds that the most significant gene at a locus is causal), since signals that overlap a single annotation type (e.g., eQTL variants) may be driven by linkage disequilibrium (LD) or pleiotropic regulatory effects [12, 13]. Second, it can increase power by increasing the signal-to-noise ratio, and capturing a wider range of possible mechanisms driving genetic associations with complex traits (e.g., [14–16]). For example, tests that incorporate both coding and eQTL variants are expected to have high power to detect both protein-altering associations as well as associations driven by effects on gene expression levels. One-dimensional annotation scores derived from multiple annotation data sets can be used to weight variants in gene-based tests (e.g., [17–19]), which can increase power by assigning higher weight to functional variants. However, aggregating variants separately for multiple annotation types and combining the result allows us to explicitly model multiple distinct genes and biological mechanisms underlying associations.

Here, we present a statistical framework and computational tool to integrate heterogeneous functional annotations with GWAS association summary statistics for gene-based analysis. We analyze a diverse set of functional annotation data including multiple tissue-specific eQTL annotation data sets, multiple epigenetic annotation sets mapping regulatory elements to putative target genes, coding variant annotations, and TSS annotations. We compare the performance of single-annotation, omnibus, and annotation-agnostic (not stratified or weighted by functional annotation) gene-based analysis methods through simulation studies, and by analyzing GWAS summary statistics from the UK Biobank [20]. Our contributions are to 1) expound a general statistical framework for gene-based analysis with heterogeneous functional

annotations, which includes several existing single-annotation gene-based association methods as components or special cases; 2) provide a computationally efficient open-source tool for gene-based analysis from summary statistics; and 3) conduct a comprehensive analysis of statistical power and accuracy identifying causal genes across gene-based association methods through extensive simulation studies and analysis of GWAS data for 128 human traits.

## Results

We first outline a statistical framework and open-source tool for gene-based analysis with heterogeneous functional annotations. Next, we describe simulations to evaluate 1) the Type I error rates of gene-based test statistics, 2) statistical power, and 3) specificity to identify causal genes. Finally, we discuss applications to empirical data using GWAS summary statistics from the UK Biobank. We assess 1) the empirical power of gene-based tests by comparing the numbers of significant independent gene-based associations discovered for each UK Biobank trait, and 2) concordance with benchmark gene lists compiled from the ClinVar database [21] and the Human Phenotype Ontology (HPO) [22].

### GAMBIT framework

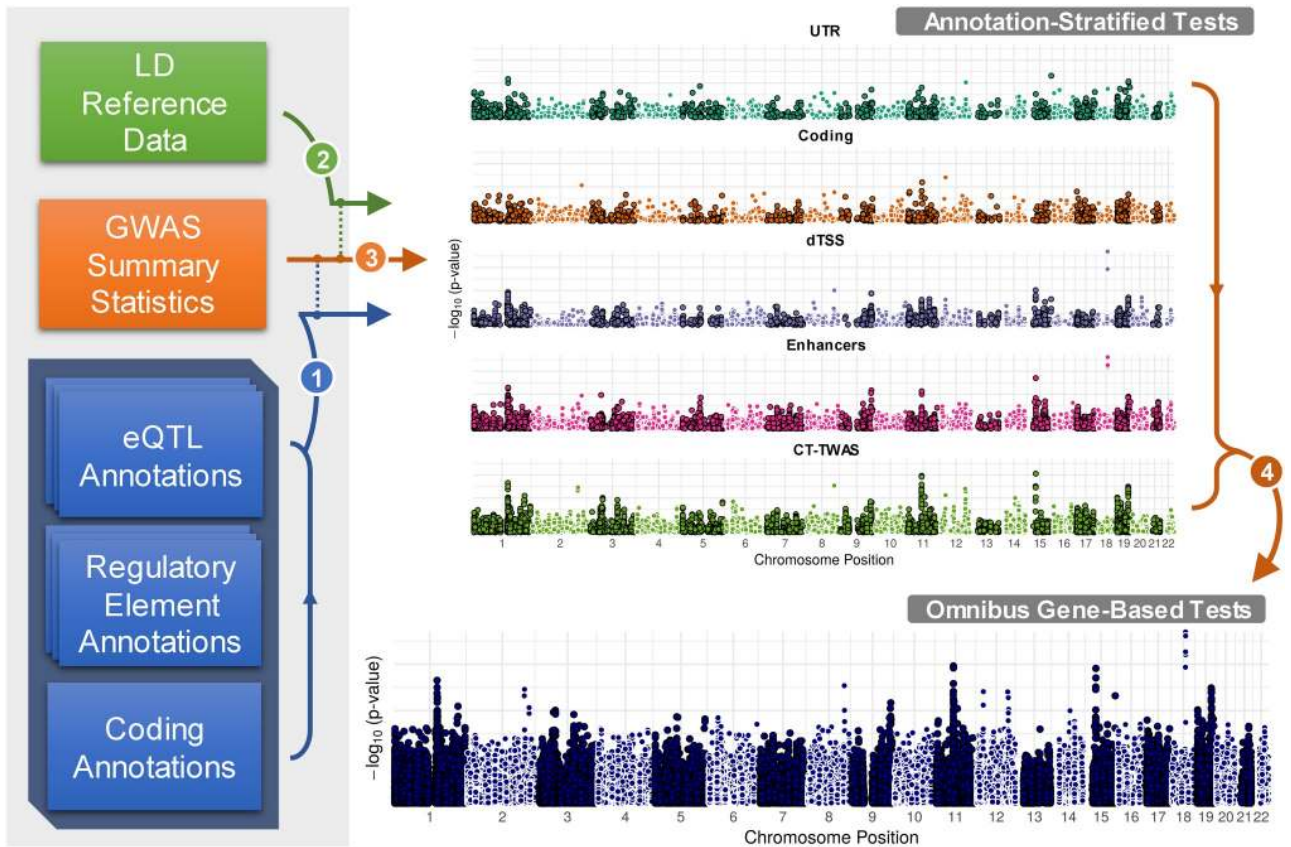
GAMBIT (Gene-based Analysis with oMniBus, Integrative Tests) is an open-source tool for calculating and combining annotation-stratified gene-based tests using GWAS summary statistics (single-variant association z-scores). Broadly, GAMBIT's strategy is to first separately calculate single-annotation gene-based association tests stratified by functional annotation class, and aggregate across classes for each gene to construct omnibus gene-based tests (illustrated in Fig 1). Here and elsewhere, we refer to this omnibus test statistic as the GAMBIT gene-based test. GAMBIT calculates four general forms of gene based test statistics, described briefly in Table 1 and detailed in Materials and Methods. To account for LD between neighboring variants and genes, GAMBIT relies on an LD reference panel from an appropriately matched population (e.g., [23, 24]). GAMBIT is implemented in C++, open source, and freely available.

### Functional annotation data

We considered 5 broad annotation classes in our analysis: 1) proximity-based annotations, 2) coding annotations, 3) UTR regions, 4) enhancer and promoter regions, and 5) eQTL predictive weights. Each of these annotation classes comprises multiple subclasses; for example, annotations include non-synonymous, splice-site, and other variant categories; and eQTL variants are stratified by tissue. Briefly, we annotated coding and UTR variants using TabAnno [32] and EPACTS [33]; obtained enhancer element and enhancer-target gene weight annotations from RoadmapLinks [10, 34], GeneHancer [35], and JEME [11]; and pre-computed tissue-specific eQTL predictive weights from PredictDB [6, 7] and FUSION/TWAS [8]. Enhancer annotations were largely derived from NIH Roadmap Epigenomics and ENCODE project data [36, 37], as well as from the FANTOM Consortium [11, 38, 39]. All eQTL variant annotations were estimated using the GTEx project v7 data [40]. Fig 2 illustrates a subset of these annotations at the *CELSR2* locus on chromosome 1; detailed descriptions of annotation data and statistical methods used to aggregate test statistics within and across classes are provided in Materials and Methods.

### GWAS simulations

We simulated GWAS summary statistics at 2,000 loci using haplotype data from the European subset of the 1000 Genomes Project (1KGP) Phase 3 reference panel [24]. Briefly, each locus



**Fig 1. GAMBIT analysis framework & workflow.** Broad overview of GAMBIT software pipeline. (1) GWAS association summary statistics (single-variant z-scores, or effect size estimates and standard errors) are cross-referenced and linked with multiple sets of functional annotations. (2) Annotated GWAS variants are cross-referenced with LD reference data (a haplotype reference panel to estimate LD as needed). (3) GWAS summary statistics, annotations, and LD estimates are used to calculate stratified gene-based test statistics. (4) Stratified gene-based tests are combined for each gene to construct omnibus test statistics. GAMBIT supports multiple single-annotation test methods and multiple omnibus test methods to combine single-annotation tests. Statistical tests are listed in Table 1; basic annotation types are illustrated in Fig 2 and listed in Table 2. A complete description of statistical methods and annotation types can be found in Materials and Methods.

<https://doi.org/10.1371/journal.pgen.1009060.g001>

**Table 1. Forms of gene-based test statistics.**

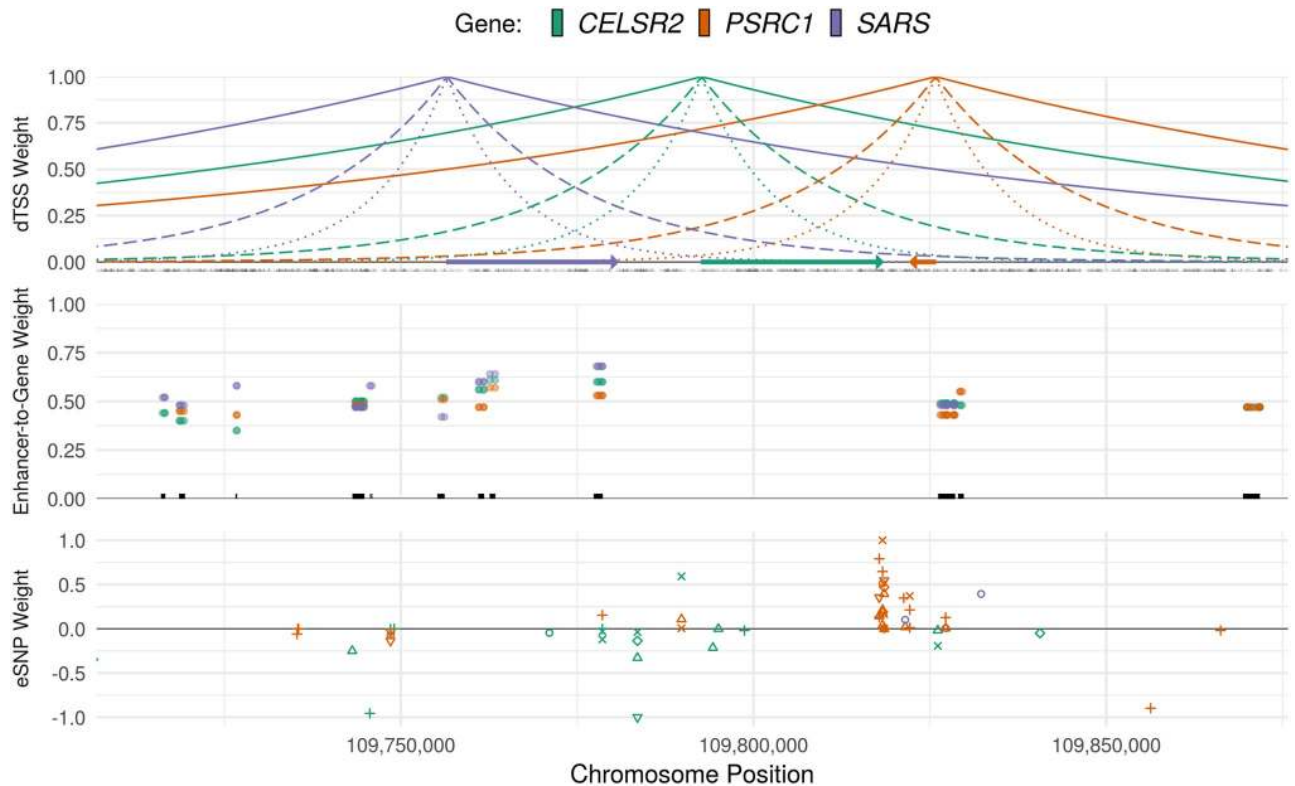
	Statistic	Null Distribution	References & Examples
L-type	$\sum_k w_k Z_k$	$\mathcal{N}(0, \mathbf{w}^T \mathbf{R}_Z \mathbf{w})$	Burden [25, 26], PrediXcan [6], TWAS [8]
Q-type	$\sum_k w_k Z_k^2$	$\sum_k \lambda_k \chi_{1,k}^2$	SKAT [27], SOCS [28]
M-type	$\max_k Z_k^2$	-	Min-P [29], MOCS [28]
ACAT	$\sum_k w_k F_{\text{Cauchy}(0,1)}^{-1}(1 - p_k)$	$\approx \text{Cauchy}(0, \sum_k w_k)$	[30, 59]
HMP	$\sum_k w_k / (\sum_k w_k / p_k)$	$\approx \text{Landau}(\mu, \pi/2)^{-1}$	[31]

Basic gene-based test forms used in GAMBIT.  $Z_k$  denotes the single-variant z-score association test statistic for variant  $k$ , with p-value  $p_k = 1 - F_{\chi_1^2}(Z_k^2)$ . Under the null hypothesis, each  $Z_k$  is standard normal and  $\mathbf{Z}$  is multivariate normal with correlation matrix  $\mathbf{R}_Z$ .

$w_k$  denotes the weight assigned to variant  $k$ . Any real-valued weights can be used in L-type tests, whereas Q-type, ACAT, and the harmonic mean p-value (HMP) require non-negative weights.

$\lambda_k$  denotes the  $k^{\text{th}}$  eigenvalue of  $\text{diag}(\mathbf{w})^{1/2} \mathbf{R}_Z \text{diag}(\mathbf{w})^{1/2}$ , and each  $\chi_{1,k}^2$  is *i.i.d*  $\chi_1^2$ . The location parameter  $\mu = \log m + 1 + \gamma + \log(\pi/2)$ , where  $m$  is the number of variants and  $\gamma$  is the Euler-Mascheroni constant.

<https://doi.org/10.1371/journal.pgen.1009060.t001>



**Fig 2. Regulatory annotation tracks and gene weights.** Illustration of primary regulatory annotation tracks used in GAMBIT gene-based analysis framework at the *CELSR2* locus on chromosome 1. Top panel: Distance-to-transcription start site (dTSS) weights, calculated as  $w_{jk}(\alpha) = \exp(-\alpha|d_{jk}|)$ , where  $d_{jk}$  is the number of base pairs between variant  $j$  and the TSS of gene  $k$ , shown for  $\alpha = 10^{-5}$  (solid lines),  $\alpha = 5 \times 10^{-5}$  (dashed lines), and  $\alpha = 10^{-4}$  (dotted lines). Gene bodies are indicated by arrows and variant locations are marked in black at  $y = 0$ . Middle panel: enhancer-to-target-gene confidence weights. Weights are shown for enhancer variant and target gene, and unique enhancer elements are marked by black lines at  $y = 0$ . Lower panel: tissue-specific eQTL weights for each gene. eQTL tissues are differentiated by shape.

<https://doi.org/10.1371/journal.pgen.1009060.g002>

was defined by first sampling a single causal protein-coding gene, aggregating all genes within 1 Mbp of the causal gene, and finally aggregating all variants assigned to one or more genes based on functional annotations or within  $\leq 500$  kbp of any gene at the locus. For each of the 2,000 loci, we simulated genetic effects under four causal scenarios: 1) coding variants are causal, 2) eQTL variants are causal, 3) enhancer variants are causal, and 4) UTR variants are causal. For each locus and causal scenario, we varied the proportion of trait variance accounted for by variants at the locus  $h_L^2 = 0.01\%$ ,  $0.025\%$ ,  $0.05\%$ ,  $0.1\%$ ,  $0.25\%$  with constant GWAS sample size  $n = 50,000$ ; and for each locus-scenario- $h_L^2$  combination, we generated 100 independent simulated replicates. To evaluate p-value calibration and Type I error rates of gene-based tests, we further simulated genome-wide summary statistics for 1,000 traits under the null hypothesis. Detailed simulation procedures are provided in Materials and Methods.

### Simulation studies: Power and accuracy identifying causal genes

We compared performance identifying causal genes across 8 gene ranking methods: 1) ranking each gene by distance between its transcription start site (TSS) and the most significant independent single variant at the locus, 2) the Pascal SOCS test  $-\log_{10}p$ -value, which assigns equal weight to all variants within 500 kbp of the gene body, 3) the omnibus test (“GAMBIT”)  $-\log_{10}p$ -value, and 4-8)  $-\log_{10}p$ -values for gene-based tests using each annotation class individually (listed in Table 2 and described in Materials and methods). As expected, test statistics

**Table 2. Single-annotation gene-based tests.**

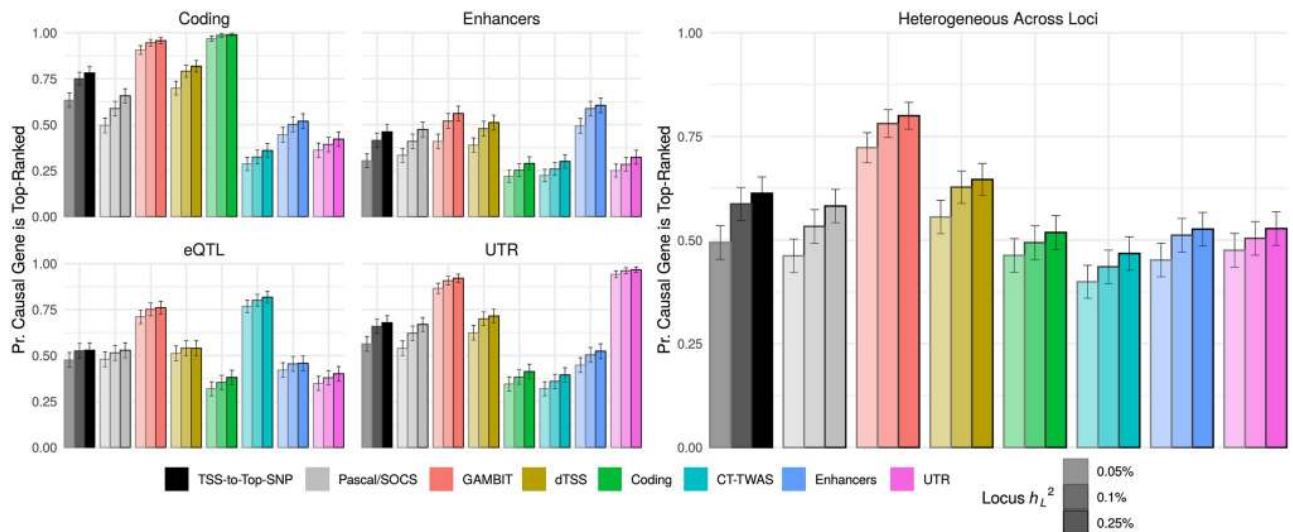
	Test Form	Annotation Subclasses	Annotated Variants
dTSS	ACAT/HMP	dTSS- $\alpha$ value	Variants within 500kbp of TSS
CT-TWAS	L-type	eQTL tissue	eQTL variants across 48 tissues
Enhancers	Q-type; ACAT/HMP	Enhancer region	All enhancer variants
UTR	Q-type; ACAT/HMP	3' and 5' UTR	3' and 5' UTR variants
Coding	Q-type; ACAT/HMP	Variant type (e.g., missense, splice site)	Exonic variants

Summary of variant types, default test methods, and default aggregation procedures for primary annotation classes in GAMBIT. Rationale and further details are provided in Materials and Methods.

<https://doi.org/10.1371/journal.pgen.1009060.t002>

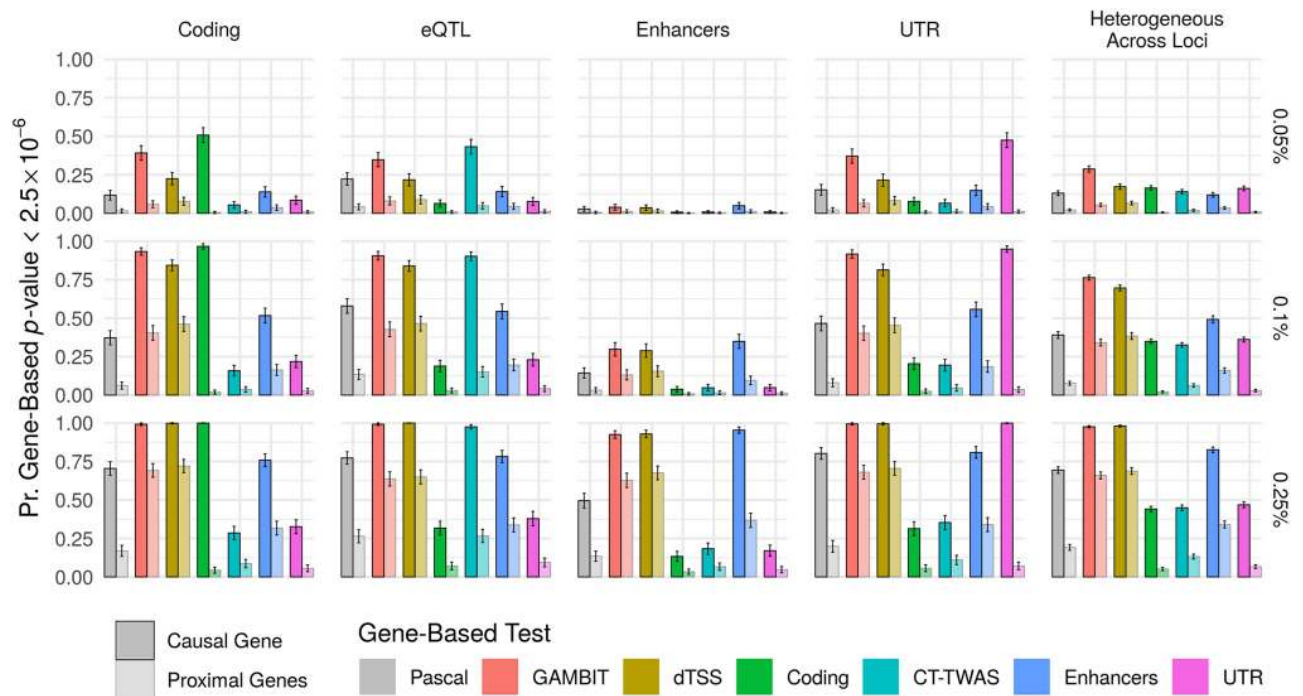
calculated using the known causal annotation class alone were most accurate for identifying the causal gene (e.g., gene-based p-values using coding variants were most accurate when coding variants were causal); however, the GAMBIT omnibus test was nearly as accurate, and had the second-highest performance across simulation settings (Fig 3; S1 Fig). In practical applications, the causal mechanisms underlying associations are unknown and often heterogeneous across loci; in this case, we expect the GAMBIT omnibus testing strategy to be most accurate (Fig 3, right panel).

We also compared statistical power for each of the gene-based test methods at both causal and non-causal proximal genes at each simulated locus (Fig 4). For proximal genes, association signals are driven by LD and pleiotropic regulatory variants shared with the causal gene; thus, gene-based tests should ideally have high power for causal genes but comparatively low power for proximal genes. Similar to the previous analysis, gene-based tests using the causal annotation class alone had the highest power for causal genes and highest specificity (low power for proximal genes) across simulation settings. The omnibus test (“GAMBIT”) generally had the



**Fig 3. Performance identifying causal genes in simulations.** Proportion of simulation replicates in which causal gene is top-ranked at its locus (y-axis) for each gene-based association or gene ranking method (x-axis & bar fill color) stratified by locus heritability  $h_L^2$  (color shade) when either coding, eQTL, enhancer, UTR variants are causal (left panel facets), or a mixture in which either coding, eQTL, enhancer, or UTR variants are causal with equal probability (“heterogeneous across loci”; right panel). TSS-to-top-SNP refers to ranking genes by the distance between their TSS and the most significant single variant at each locus; dTSS-weighted gene-based tests (labeled dTSS) use exponential weight functions to assign higher weight to variants nearer the TSS for each gene (Materials and methods).

<https://doi.org/10.1371/journal.pgen.1009060.g003>



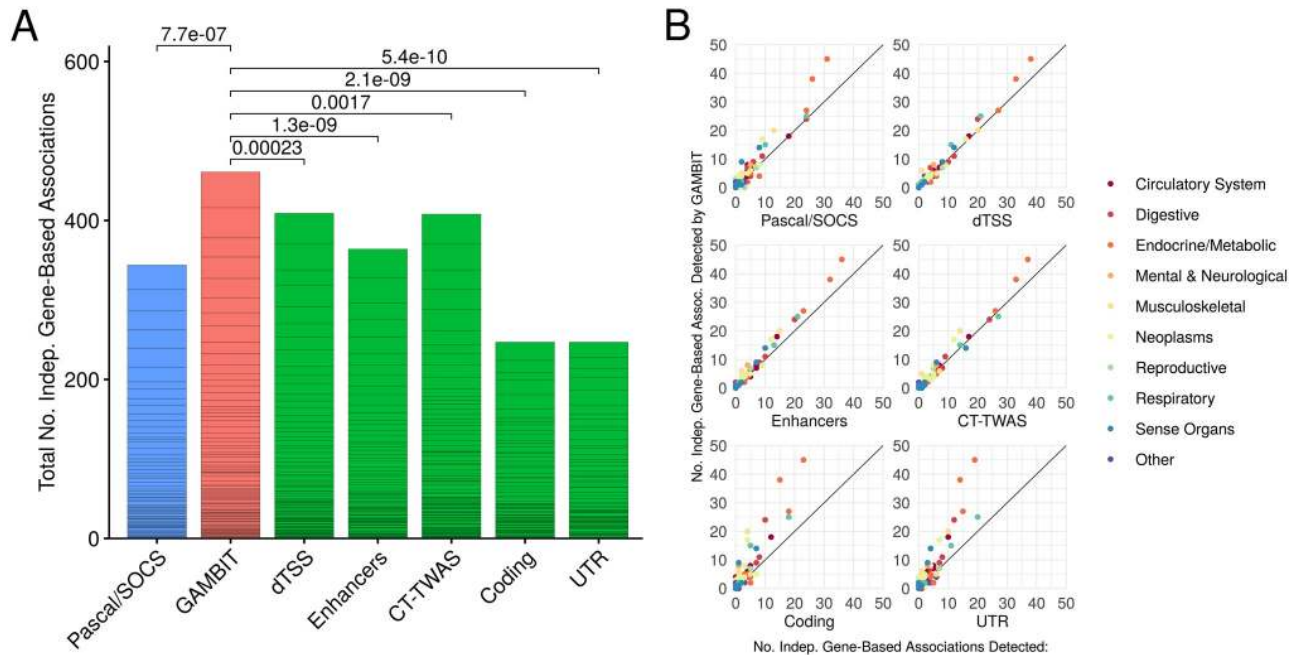
**Fig 4. Statistical power to detect gene-based associations in simulations.** Statistical power (proportion of simulation replicates in which gene-based  $p$ -value  $\leq 2.5 \times 10^{-6}$  across loci;  $y$ -axis) for each gene-based testing approach ( $x$ -axis & color) stratified by locus heritability  $h_L^2$  (plot rows) when coding, eQTL, enhancer, UTR variants, or a mixture of these (“heterogeneous across loci”) are causal (plot columns). In the rightmost column, either coding, eQTL, enhancer, or UTR variants are causal with equal probability (as when the causal annotation class is heterogeneous across loci for a single trait). Power is shown separately for causal genes and proximal genes (non-causal genes that are proximal to a causal gene, as defined in [Materials and methods](#)). Ideally, gene-based tests should have high power for causal genes, and relatively lower power for proximal genes. Error bars show 95% confidence intervals for average power across loci.

<https://doi.org/10.1371/journal.pgen.1009060.g004>

second-highest power for causal genes, and intermediate power for proximal genes. Thus, we expect the omnibus testing approach to be powerful and robust when causal mechanisms are unknown or heterogeneous across loci.

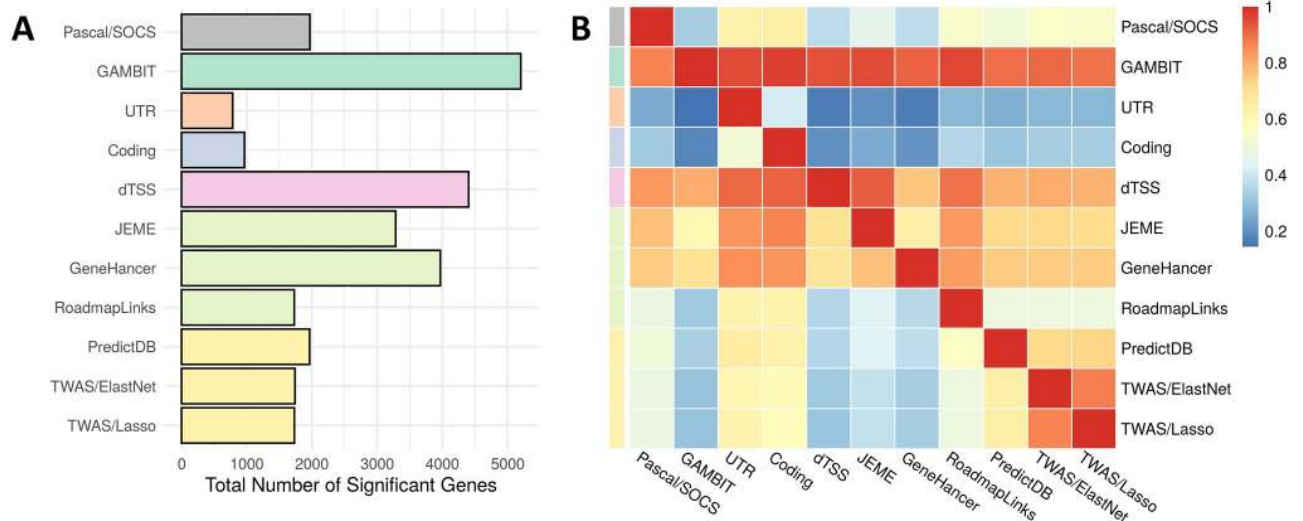
## Analysis of GWAS summary statistics from UK Biobank

**Significant independent associations detected for 128 UK Biobank traits.** To compare the power of gene-based tests in empirical data, we evaluated the numbers of significant independent gene-based associations detected for each method across 128 approximately independent GWAS traits in the UK Biobank (selection procedures are described in [Materials and methods](#)). The number of independent associations is calculated for each trait by selecting the most significant gene-based association  $p$ -value, masking all gene-based tests that include variants within 1 Mbp of variants for the selected gene, and repeating until all genes with Bonferroni-adjusted  $p$ -value  $\leq 5\%$  are either selected or masked. This procedure ensures that all selected genes are separated by at least 1 Mbp, and provides a conservative estimate of the number of significant independent signals. The omnibus test (“GAMBIT”) detected significantly more associations than other gene-based association methods overall ([Fig 5A](#)), and consistently detected more associations than other methods across a wide range of traits ([Fig 5B](#)). We also compared the numbers of significant associations for each method without filtering or LD pruning ([Fig 6](#)). Statistics that incorporate many variants over a broad region for each gene (e.g., dTSS-weighted tests) yield substantially more significant associations, as expected.



**Fig 5. UK Biobank analysis: Numbers of significant independent associations detected.** Numbers of independent gene-based associations (at Bonferroni-corrected 5% significance level) detected by each method across 128 UK Biobank traits. Panel A: Total number of significant independent associations across traits (delineated by horizontal black lines) for each gene-based test; Wilcoxon signed-rank p-values (top) for paired comparisons between no. associations detected by the omnibus test (“GAMBIT”; red) versus Pascal/SOCS (blue) and single-annotation gene-based tests (green). The omnibus test detects significantly more associations than any individual constituent gene-based test or by Pascal/SOCS across UK Biobank traits. Panel B: Comparison of total numbers of genes detected across individual traits for the omnibus test (y-axis) versus single-annotation tests (x-axis).

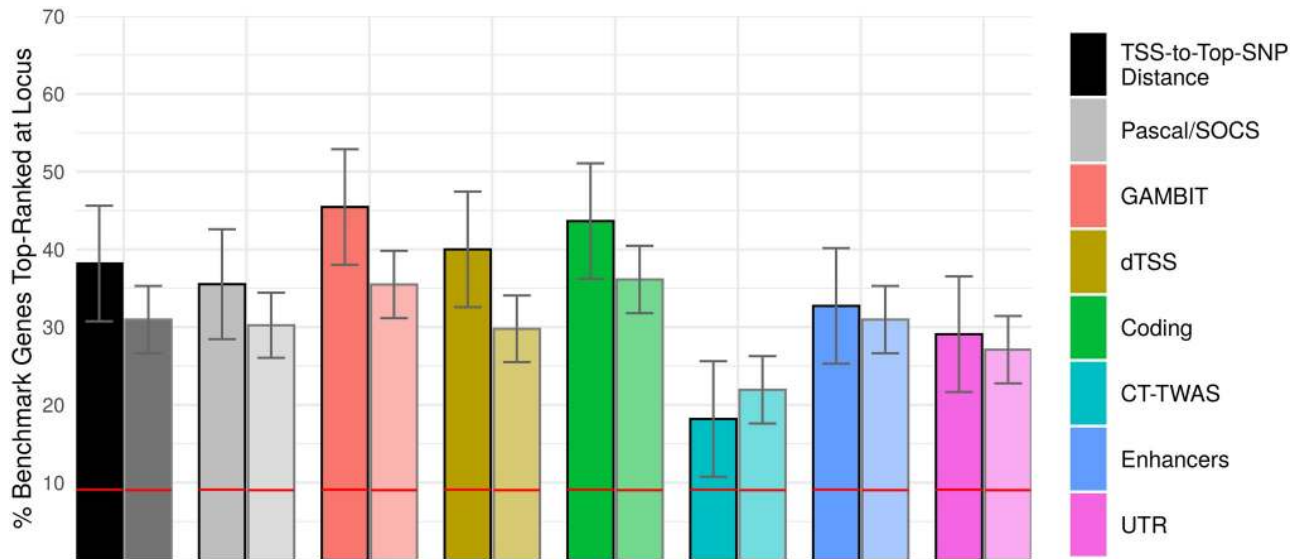
<https://doi.org/10.1371/journal.pgen.1009060.g005>



**Fig 6. UK Biobank analysis: Overlap between gene-based association methods.** Panel A: Total number of significant genes ( $p$ -value  $< 2.5e-6$ ) for each method across all 128 traits. Unlike Fig 5, gene-based associations in Fig 6 are not filtered or LD pruned, and a single significant GWAS variant can produce multiple significant gene-based associations for a given method. Here, a larger number of significant genes does not necessarily suggest greater statistical power. Panel B: The  $i, j^{th}$  heatmap element can be interpreted as the conditional probability that gene-based test  $i$  is significant given that gene-based test  $j$  is significant, which is estimated as the total number of overlapping significant genes between tests  $i$  and  $j$  divided by the total number of significant genes for test  $j$ .

<https://doi.org/10.1371/journal.pgen.1009060.g006>





**Fig 7. UK Biobank analysis: Performance identifying benchmark genes.** Percentage of loci at which the benchmark gene (identified from HPO and/or ClinVar) is top-ranked for each gene-based association or gene ranking method. For each method, bars on the left (outlined in black) are calculated for benchmark loci present in both HPO and ClinVar (54 loci), and bars on the right (faded outline) are calculated using the union of all HPO and ClinVar loci (153 loci). Horizontal red lines indicate the expected percentage of top-ranked benchmark genes under the null hypothesis that gene rank and benchmark labels are independent. Error bars indicate 95% confidence intervals. TSS-to-top SNP refers to ranking genes by the distance between TSS and the most significant single variant at each causal locus; the dTSS-weighted gene-based test (dTSS) uses an exponential weight function to assign higher weight to variants nearer the TSS for each gene ([Methods](#)).

<https://doi.org/10.1371/journal.pgen.1009060.g007>

**Concordance with benchmark genes for 25 UK Biobank traits.** We compiled lists of benchmark genes from the ClinVar database [21] and the Human Phenotype Ontology (HPO) [22] for 25 traits in the UK Biobank to compare the gene-based analysis methods identifying causal genes; procedures and selection criteria are detailed in Materials and Methods. Results are shown separately using the union and intersection of ClinVar and HPO benchmark genes; the latter gene set is expected to have higher specificity, albeit fewer genes. Performance identifying benchmark genes was assessed by ranking genes separately within each benchmark locus for each UK Biobank trait, where a benchmark locus is defined as the set of all genes within 1 Mbp of a genome-wide significant single-variant association that also is within 1 Mbp of a benchmark gene. To compare the performance of gene ranking methods, we calculated the fraction of loci at which the top-ranked gene coincides with a benchmark gene ([Fig 7](#)) and assessed receiver operating characteristic (ROC) and precision-recall curves for each method ([S2 Fig](#)).

GAMBIT omnibus tests had the highest performance identifying benchmark genes among the gene ranking methods considered, particularly for the stricter gene set, although the difference was not statistically significant relative to most other gene ranking methods ([Fig 7](#)). Gene-based tests using coding variants alone had the second-highest performance ([Fig 7](#); [S2 Fig](#)), which may reflect the enrichment for coding associations within the benchmark gene set ([S3 Fig](#)) caused by benchmark gene selection criteria (described in [Materials and methods](#)). Due to the over-representation of coding associations, [Fig 7](#) may underestimate the impact of incorporating heterogeneous regulatory annotations for associated loci without an established benchmark gene.

Further inspection revealed a number of loci of biological or clinical interest. In the analysis of skin cancer in the UK Biobank, three melanin or melanogenesis-related genes (*TYR*, *OCA2*, and *MC1R*) and telomerase reverse transcriptase (*TERT*) were top-ranked by the

omnibus test, but not top-ranked based on TSS-to-top-SNP distance, while all other benchmark genes for skin cancer were top-ranked by both methods or by neither. At the *TERT* locus, the lead GWAS variant was intronic, whereas the lead variants for *TYR*, *OCA2*, and *MC1R* were nonsynonymous. Unsurprisingly, the latter three benchmark genes were also top-ranked based on coding variant gene-based p-values; however, only *TERT* was top-ranked based on CT-TWAS.

Similarly, *APOB*, which encodes an apolipoprotein and is associated with autosomal dominant forms of hypercholesterolemia, was top-ranked by the omnibus test but not by TSS-to-top-SNP distance for disorders of lipid metabolism in the UK Biobank. Despite being >150 Kbp from the intergenic lead GWAS variant, *APOB* was also top-ranked by all single-annotation gene-based tests individually. Conversely, *TSHR*, which encodes a thyroid hormone receptor, was top-ranked based on TSS-to-top-SNP distance but not by the omnibus test for thyrotoxicosis. In this case, the lead GWAS variant was intronic, and CT-TWAS was the only single-annotation gene-based test that ranked *TSHR* as the top gene at its locus; in this example, while the omnibus test for *TSHR* was significant, it was out-ranked by *CEP128* at the locus. A complete table of results for benchmark genes is provided in Supplementary Materials.

## Discussion

Here, we introduced GAMBIT, a statistical framework and software tool for gene-based analysis with heterogeneous annotations. Our work makes several contributions to the field:

First, we conducted extensive simulation studies to systematically compare gene-based test methods across a range of plausible biological scenarios, and demonstrated pitfalls of test methods that use only a single annotation class. When the causal annotation class is misspecified, standard gene-based tests have limited power, and can be confounded by LD and pleiotropic regulatory variants that affect multiple genes. This may lead researchers to misidentify the genes and biological mechanisms that contribute to disease risk. Finemapping, co-localization, and conditional analysis can be applied to refine association signals and mitigate spurious inferences following gene-based analysis (e.g., [41–44]). By contrast, our omnibus testing strategy helps to ameliorate spurious inferences within the context of gene-based testing directly, and also has high power to detect associations across a range of causal mechanisms underlying genetic associations.

Second, we analyzed 128 traits from the UK Biobank to evaluate performance in empirical data across a range of complex traits and genetic architectures, and confirmed that incorporating annotations of many types and across many tissues increases power relative to standard methods. While our analysis of concordance with gold-standard causal genes was limited by the relatively small numbers of benchmark genes identified for UK Biobank traits and the inherent difficulty establishing causal genes underlying regulatory associations, we found suggestive evidence that incorporating diverse annotation types in gene-based analysis can improve performance identifying causal genes relative to standard approaches (e.g., ranking genes by distance to the most significant single variant) and gene based tests using a single annotation type.

Finally, we provide a unifying framework and easy-to-use software tool to incorporate heterogeneous functional annotations in gene-based analysis. From its inception, gene-based analysis was built on the premise that aggregating functional variants at the gene level can increase statistical power and help identify causal genes in GWAS [2]. Early gene based test methods were developed primarily for rare genic variants (e.g., [25, 26]), and early gene-based association analyses often used only deleterious coding variants (e.g., [45, 46]). However,

functional genomics studies have shown that most functional variation is non-coding [37], and most variant associations discovered through GWAS to date occur in non-coding regions [1, 9], highlighting the importance of regulatory annotations for gene-based association analysis. The first gene-based tests developed explicitly for regulatory variation were TWAS and PrediXcan, which aggregate eQTL variants to construct proxy variables for tissue-specific gene expression levels using predictive weights estimated from external eQTL mapping data [7, 8]. However, functional and regulatory genomics projects have introduced a wealth of annotations with potential utility for gene-based analysis (e.g., [11, 35, 38, 47]).

The omnibus testing strategy used here is expected to perform best under sparse alternatives, e.g. when one or few annotation classes harbor causal variants at a given locus. When a larger fraction of annotation classes harbor independent signals at a single gene locus, this omnibus strategy may have less power than one that explicitly accounts for multiple simultaneous signal sources. While we did not explore this possibility in our simulations, it is an interesting question which we defer to future work.

Previous studies have evaluated the performance of gene-based tests under misspecification, e.g. by varying the proportion of causal variants and correlation structure of causal effects [48, 49]. In the present study, we evaluated the performance of single-annotation gene-based tests under misspecified causal mechanisms (for example, TWAS when a mechanism other than gene expression underlies the association signal). The former problem primarily concerns the statistical form of gene-based test and the distribution of causal effects, while the latter is more related to the informativeness of functional annotations and the overlap between classes of functional variation (e.g., Fig 6B). Our simulation studies also included basic forms of misspecification in the distribution of causal effects (e.g., when only a fraction of annotated variants in the causal annotation class have non-zero causal effects) and measurement error in functional annotations (e.g., by including a error term in TWAS weights). However, further research is needed to explicitly address model misspecification and annotation measurement error in gene-based analysis.

The utility of incorporating annotations in gene-based analysis depends crucially on the accuracy and comprehensiveness of the underlying annotation data sets. While we considered the case that causal variants may be misspecified, our simulations assumed that the confidence weights assigned to regulatory elements are well-calibrated, and that causal eQTL variants are annotated. Violations of these assumptions will reduce both power and accuracy in gene-based analysis, and may in part account for differences between our results with empirical versus simulated data. Current transcriptomic and epigenomic studies are generally limited to a subset of human tissues and cell-types, and are derived from data sets of limited sample size (e.g., [37, 47]). Thus, we expect current transcriptomic and epigenomic annotations to be incomplete and imprecise. Looking forward, larger and more comprehensive studies will enable more comprehensive and accurate annotations, increasing the utility of annotation-informed association analysis methods.

In summary, our work builds upon and generalizes previous gene-based association methods, providing a flexible framework for gene-based analysis with heterogeneous annotations that can be readily adapted when new annotation resources are developed and released.

## Materials and methods

We describe 1) gene-based association test statistics, 2) procedures to aggregate variants within each class of functional variation for gene-based analysis, 3) functional annotation data sources 4) procedures to simulate GWAS data using real genotype and functional annotation data, and 5) GWAS data from the UK Biobank to which we applied our methods.

## Multiple-variant association test statistics

Here, we review statistical methods to aggregate multiple variants for gene-based, region-based, or pathway association analysis. For convenience, we assume a quantitative trait and ignore the presence of covariates; however, our results can easily be adapted to other settings.

**Linear-type gene-based tests (L-type).** The oldest and most widely used gene-based tests are linear combinations of genotypes across variants [25, 26, 50], here referred to as L-type tests. We define the L-type test statistic as  $T_L = (\mathbf{w}^\top \mathbf{R}_Z \mathbf{w})^{-1/2} \mathbf{w}^\top \mathbf{Z}$ , where  $\mathbf{w}$  is a vector of single-variant weights,  $\mathbf{Z}$  is a vector of single-variant association statistics (where each  $Z_j$  follows the standard normal distribution under the null hypothesis), and  $\mathbf{R}_Z$  is the correlation matrix of z-scores. Under the null hypothesis of no association,  $T_L$  follows the standard normal distribution. The L-type test statistic  $T_L$  can be computed from GWAS summary statistics (single-variant z-scores, or effect sizes and standard errors) and covariance estimates, and can be written either as linear combinations of single-variant association statistics or as linear combinations of genotypes [4, 51].

Examples of L-type tests include burden tests, which calculate burden scores as a weighted sum of rare, putatively deleterious mutations [25, 50]; the cohort allelic sums test (CAST) [52]; and TWAS/PrediXcan tests [6–8], which aggregate eQTL variants using predictive weights estimated from external data sets, e.g. from the GTEx project [40]. These can be viewed as tests of association between GWAS trait and an explicit proxy variable constructed as a linear combination of genotypes. Importantly, L-type tests rely on prior knowledge regarding the directions of effect across variants [27, 50]. For example, the signed weights used in burden tests often reflect the hypothesis that rare deleterious alleles increase risk for disease, and the predictive weights used in TWAS/PrediXcan reflect the hypothesis that gene expression mediates the associations between genotypes and complex trait.

**Quadratic-type gene-based tests (Q-type).** Variance component tests and quadratic forms of single-variant association statistics comprise another widely used class of gene-based association methods, here referred to as Q-type (quadratic) tests. Q-type tests include VEGAS (or SOCS), defined as the sum of squared single-variant z-scores [28, 53]; the C-alpha test [54]; and SKAT, a weighted quadratic form of single-variant association statistics [27]. We define the Q-type test statistic as  $T_Q = \mathbf{Z}^\top \text{diag}(\mathbf{w}) \mathbf{Z}$ , where  $\text{diag}(\mathbf{w})$  is a diagonal weight matrix and  $\mathbf{Z}$  is a vector of single-variant association z-scores; under the null hypothesis of no association,  $T_Q$  follows a mixture chi-squared distribution with mixture proportions equal to the eigenvalues of  $\text{diag}(\mathbf{w})^{1/2} \mathbf{R}_Z \text{diag}(\mathbf{w})^{1/2}$ , where  $\mathbf{R}_Z$  is the correlation matrix of z-scores. In contrast to L-type tests, Q-type tests aggregate single-variant association statistics without prior knowledge or assumptions pertaining to the directions of effects across variants [27, 50]. While less tractable than L-type, analytical p-values for Q-type tests can be calculated using a variety of techniques to approximate the tail probabilities of multivariate normal quadratic forms (e.g., [55, 56]), which are far more efficient than permutation procedures or Monte Carlo methods [28, 57]. Q-type tests are most appropriate when a sizable proportion of variants are hypothesized to have non-zero effects of unknown and inconsistent direction [50].

**Maximum chi-squared statistic as a gene-based test (M-type).** Perhaps the simplest gene-based test is the maximum chi-squared statistic across variants (or equivalently, the minimum p-value), here referred to as M-type tests. Analytical p-values for M-type tests can be calculated by directly integrating the multivariate normal density of z-scores within the hypercube given by  $\mathbf{x} \in \mathbb{R}^m : \max_k |x_k| \leq \max_j |Z_j|$  where  $m$  is the number of variants, or approximated by adjusting the minimum p-value across variants by the effective number of tests [28, 29]. M-type tests are most appropriate when only one or a small fraction of variants are hypothesized to have non-trivial effects. We note that the M-type test accounts for the

correlation structure across variants, whereas Tippett’s method [58] assumes independent p-values; thus, the M-type test reduces to Tippett’s method when z-scores are uncorrelated.

**Aggregated cauchy association test (ACAT).** The aggregated Cauchy association test (ACAT), a recently proposed method to combine multiple dependent p-values, can be used to construct gene-based tests by transforming single-variant association p-values using the Cauchy quantile and cumulative distribution functions, and computing a p-value

$$p_{ACAT} = F_{Cauchy(0,1)} \left( \frac{1}{\sum_i w_i} \sum_i w_i F_{Cauchy(0,1)}^{-1}(p_i) \right),$$

where  $p_i$  and  $w_i$  are the p-value and weight for the  $i^{th}$  variant and  $F_{Cauchy(0,1)}(t) = \frac{1}{\pi} \arctan(t) + \frac{1}{2}$  is the CDF of the standard Cauchy distribution [30, 59]. ACAT is expected to perform well when only a small fraction of variants are causal [30]. Importantly, ACAT does not require LD computation, and can thus be calculated in  $O(m)$  time where  $m$  is the number of variants.

**Harmonic mean p-value (HMP).** Another recently proposed method to combine multiple dependent p-values, the Harmonic Mean P-value (HMP; [31]), can similarly be used to construct gene-based tests by weighting p-values from single-variant association tests. The unadjusted HMP p-value is defined

$$p_{HMP} = \frac{\sum_k w_k}{\sum_k w_k / p_k}.$$

While this statistic can be anti-conservative when directly interpreted as a p-value, Wilson (2019) showed that  $1/p_{HMP}$  follows a Landau distribution (with scale and location parameters given in Table 1), which can be used to compute an asymptotically exact HMP p-value. The Landau density function is

$$f_{Landau}(x; \mu, \sigma) = \frac{1}{\pi\sigma} \int_0^\infty e^{-u} \cos \{ (x - \mu)u/\sigma + (2u/\pi) \log(u/\sigma) \} du,$$

which can be computed numerically with high precision using asymptotic expansions [60]. To improve p-value calibration, we implemented the asymptotically exact HMP in the GAMBIT software tool. Unlike L-type and Q-type tests, p-values from M-type, ACAT, and HMP tests are greater than or equal to  $\min_i p_i$ . However, these methods can still increase power relative to single-variant analysis by reducing the burden of multiple testing and assigning higher weight to functional variants.

**Generalizations and extensions.** The simple forms of gene-based tests described above can be related and combined through a variety generalizations and extensions. Q-type and M-type can both be viewed as special cases of a statistic  $(\sum_j w_j |Z_j|^p)^{1/p}$ , which is equivalent to Q-type when  $p = 2$  and to M-type when  $p \rightarrow \infty$ ; this generalization has been used, for example, in the aSPU gene-based test [61]. Similarly, Q-type and L-type can both be viewed as special cases of a statistic  $\mathbf{Z}^\top (\pi \text{diag}(\mathbf{w}_1) + (1 - \pi) \mathbf{w}_2 \mathbf{w}_2^\top) \mathbf{Z}$ , which is equivalent to Q-type when  $\pi = 1$  and L-type when  $\pi = 0$ ; this generalization has been used, for example, in the SKAT-O gene-based test [50]. Finally, ACAT and HMP can be used to combine p-values across multiple gene-based test forms [30, 31].

### Integrating functional annotations in gene-based tests

Here we describe methods to aggregate variants within each of the 5 major annotation classes considered in our analysis. Briefly, we use linear (L-type) tests to combine eQTL variants using signed predictive weights reflecting the alternative hypothesis that genotype effects on trait are

mediated by gene expression levels. For coding and UTR variants, we use Q-type tests within each annotation subclass, reflecting the alternative hypothesis that genetic effects follow a symmetric mean-zero distribution. For dTSS-weighted tests, we use weighted dependent p-value combination procedures (ACAT or HMP), which can be viewed as an approximate test for the alternative hypothesis that a variant is causal with prior probability proportional to  $e^{-\alpha d_{TSS}}$ .

**Coding variants.** Gene-based tests for coding variants are calculated by aggregating variants separately within each coding subclass (e.g., missense, nonsense, and synonymous) using Q-type (by default) test statistics. These stratified tests are then combined across subclasses using a p-value combination procedure (HMP or ACAT) to calculate a coding omnibus test for each gene.

**UTR variants.** Gene-based tests for UTR variants are calculated by aggregating variants separately within the 3' and 5' UTR regions using Q-type (by default) test statistics, and applying a p-value combination procedure (HMP or ACAT) to calculate a UTR omnibus test for each gene.

**dTSS weights.** One of the most common heuristics to infer likely causal genes at non-coding GWAS loci is to rank genes by distance between their transcription start site (TSS) and the most significant single GWAS variant. This strategy is appealing given the strong enrichment of regulatory variants near TSS.

To incorporate distance-to-TSS (dTSS) and capture association signals at regulatory variants that are not well-annotated in gene-based analysis, we define the dTSS weights for gene  $k$  as  $w_{jk}(\alpha) = e^{-\alpha |d_{jk}|}$ , where  $d_{jk}$  is the genomic distance (number of base pairs) between variant  $j$  and the TSS for the gene of interest. Larger values of the parameter  $\alpha$  confer more weight to variants nearer the TSS. In practice, we only include variants within a specified window (e.g., 500kbp) of the TSS of the corresponding gene. While dTSS weights can be used in any weighted gene-based test (e.g., Q-type tests), ACAT and HMP are particularly well-suited due to their linear computational complexity, as dTSS-weighted tests often involve thousands of variants per gene.

The optimal  $\alpha$  value is expected to vary across loci, and likely depends on local gene density and other factors. However, ACAT and HMP can be applied again to calculate omnibus p-values by combining dTSS-weighted gene-based test p-values  $p_k(\alpha_i)$  across multiple values  $\alpha_1, \alpha_2, \dots$  [30, 59]. By default, GAMBIT calculates overall dTSS-weighted test statistics by aggregating across  $\alpha$  values  $10^{-4}, 5 \times 10^{-5}, 10^{-5}, 5 \times 10^{-6}$ .

**Enhancer-target gene weights.** To capture association signals across regulatory elements that have been assigned to one or more target gene, we weight variants in regulatory elements by element-to-target-gene confidence scores, and aggregate variants for each gene using either ACAT, HMP, or Q-type gene-based test statistics. For example, we define the regulatory-element weighted Q-type test statistic as  $T_k^R = \sum_i \sum_{j=1}^{m_i} w_{ik} Z_{ij}^2$  where  $m_i$  is the number of variants in the  $i^{\text{th}}$  regulatory element,  $w_{ik}$  is the confidence weight between element  $i$  and gene  $k$ , and  $Z_{ij}$  is the  $j^{\text{th}}$  variant in the  $i^{\text{th}}$  regulatory element.

**eQTL weights.** Given a vector of weights  $\mathbf{b}_{kt}$  to predict expression levels for gene  $k$  in a given tissue or cell type  $t$  as a linear combination of normalized genotypes, we define the z-score TWAS test of association between predicted expression level and GWAS trait as

$S_{kt} = \mathbf{b}_{kt}^T \mathbf{Z} / \sqrt{\mathbf{b}_{kt}^T \mathbf{R}_Z \mathbf{b}_{kt}}$  where  $\mathbf{Z}$  is the vector of single-variant GWAS z-scores and  $\mathbf{R}_Z$  is the correlation matrix of z-scores.

To aggregate test statistics across multiple tissues or cell-types, which we refer to as Cross-Tissue TWAS (CT-TWAS), we considered three approaches:

1. Q-type Cross-tissue Test (CT-Q): Calculating the sum of squared tissue-specific test statistics,  $\sum_t S_{kt}^2$ , which has a mixture chi-squared distribution under the null hypothesis of no association,

2. M-type Cross-tissue Test (CT-M): Calculating an analytic p-value for the maximum absolute test statistic  $\max_t |S_{kt}|$  using the multivariate normal joint density of tissue- or cell-type-specific test statistics  $S_{k1}, S_{k2}, \dots$  under the null hypothesis of no association, and
3. ACAT or HMP Cross-tissue Test (CT-A or CT-H): Combining tissue- or cell-type-specific p-values  $p_{kt} = 2\Phi(-|S_{kt}|)$  using the ACAT method or HMP respectively.

CT-Q and CT-M require the cross-tissue correlation matrix  $\mathbf{R}_S$  with elements

$[r_S]_{tt'} = \text{corr}(\mathbf{b}_{kt}^\top \mathbf{Z}, \mathbf{b}_{kt'}^\top \mathbf{Z}) = \mathbf{b}_{kt}^\top \mathbf{R}_Z \mathbf{b}_{kt'} / \sqrt{(\mathbf{b}_{kt}^\top \mathbf{R}_Z \mathbf{b}_{kt})(\mathbf{b}_{kt'}^\top \mathbf{R}_Z \mathbf{b}_{kt'})}$ , which can be computed in  $O(m^2 n + mn^2)$  time where  $m$  is the number of tissues or cell-types and  $n$  is the number of eQTL variants. By contrast, CT-A and CT-H p-values can be computed in  $O(m)$  time, since ACAT and HMP do not require the correlation matrix to be computed. By default, GAMBIT implements CT-A; in our analysis of UK Biobank data, CT-M, CT-H, and CT-A generally perform similarly, while CT-Q tends to detect fewer significant associations.

**Combining single-annotation test statistics.** In early versions of GAMBIT, we combined gene-based p-values across annotation classes for each gene using standard family-wise error rate (FWER) and false discovery rate (FDR) controlling procedures. However, two recently proposed methods, the Aggregated Cauchy Association Test (ACAT; [59]) and Harmonic Mean P-value (HMP; [31]), provide more powerful approaches to combine multiple dependent p-values, and we have therefore implemented both of these methods in GAMBIT. Unlike gene-based tests such as SKAT, which are formulated as parametric tests in a generalized linear mixed model, these p-value combination methods are essentially non-parametric, assuming only that p-values are uniformly distributed under the null hypothesis. Like the exponential combination (EC) procedure [62], ACAT and HMP are powerful under sparse alternatives; however, unlike EC, they enable efficient analytic p-value calculation, with computation time linear in the number of p-values. We calculated ACAT p-values (defined above) using standard formula for the Cauchy quantile function and CDF. To calculate asymptotically exact HMP p-values, we adapted a C++ routine from the ROOT System [63] for computing the Landau distribution CDF following the derivations of Wilson (2019) for the asymptotic distribution of the HMP.

## Functional annotation data sources

**Enhancer-target annotation sources.** To identify regulatory genetic elements and their putative target genes, we used pre-computed annotation data sets from three existing methods: Joint Effects of Multiple Enhancers (JEME) [11], GeneHancer [35], and RoadmapLinks [10, 34, 64]. GeneHancer provides a global confidence score between each enhancer element and one or more putative target genes, while JEME and RoadmapLinks provide tissue- or cell-type-specific enhancer-target confidence scores. For the latter two data sets, we calculated overall enhancer-target confidence scores across tissues and cell types as the soft maximum (*LogSumExp* function) of tissue- or cell-type-specific scores for each enhancer-target pair. Descriptive statistics for each enhancer annotation dataset are provided in [S2 Table](#).

**eQTL predictive weight annotation sources.** To incorporate eQTL variants in gene-based analysis, we used pre-computed tissue-specific predictive weights for eGene expression estimated using GTEx v7 [47] from TWAS/FUSION (including elastic net and LASSO models) [8] and PredictDB [6, 7]. We generated a GAMBIT eWeight annotation files incorporating all available tissues and cell types for each data resource and predictive model. Descriptive statistics for each eQTL variants weight dataset are provided in [S1 Table](#).

**Coding variant and gene annotation sources.** We annotated coding variants, TSS locations, and UTR variants using TabAnno 419 [32] and EPACTS [33] based on GENCODE v14 [65].

### Simulation procedures

Here, we describe simulation procedures for GWAS summary statistics, configurations of causal genes, and causal variant effects.

**Simulating GWAS summary statistics.** We simulated GWAS traits under the model  $Y = \mathbf{1}_n \beta_0 + \tilde{\mathbf{G}} \boldsymbol{\beta} + \boldsymbol{\varepsilon}$ , where  $Y \in \mathbb{R}^n$  is a quantitative trait for a GWAS sample of size  $n$ ,  $\mathbf{1}_n$  is the  $n \times 1$  vector of 1's and  $\beta_0 \in \mathbb{R}$  is the trait intercept,  $\tilde{\mathbf{G}} \in \mathbb{R}^{n \times m}$  is the centered and scaled genotype matrix where each column has mean 0 and variance 1,  $\boldsymbol{\beta} \in \mathbb{R}^{m \times 1}$  is a vector of causal genetic effects, and  $\boldsymbol{\varepsilon} \in \mathbb{R}^n$  is an *i.i.d.* trait residual with  $\mathbb{E}(\boldsymbol{\varepsilon}_i) = 0$  and  $\text{Var}(\boldsymbol{\varepsilon}_i) = \sigma_\varepsilon^2$ . We scale the parameters  $\sigma_\varepsilon^2$  and  $\boldsymbol{\beta}$  so that  $Y_i$  has unit marginal variance.

We define the vector of single-variant association statistics (equivalent to t-test statistics from simple linear regression) for variants  $k = 1, 2, \dots, m$  as

$$\mathbf{Z} = (n - 1)^{1/2} \hat{\mathbf{D}}^{-1/2} \frac{1}{n} \tilde{\mathbf{G}}^\top \mathbf{Y} = n^{1/2} \hat{\mathbf{D}}^{-1/2} \hat{\mathbf{R}} \boldsymbol{\beta} + (n - 1)^{1/2} \hat{\mathbf{D}}^{-1/2} \frac{1}{n} \tilde{\mathbf{G}}^\top \boldsymbol{\varepsilon}$$

where  $\hat{\mathbf{R}}_n = \frac{1}{n-1} \tilde{\mathbf{G}}^\top \tilde{\mathbf{G}}$  is the sample LD matrix, and  $\hat{\mathbf{D}}$  is an  $m \times m$  diagonal matrix with  $\hat{D}_{kk} = \frac{n^2}{(n-2)(n-1)} (\hat{\sigma}_Y^2 - \hat{\alpha}_k^2)$ . Note that  $\hat{\mathbf{D}} \approx \mathbf{I}_m$  if the proportion of trait variance accounted for by each individual variant is small (e.g.,  $< 1\%$ ).

We simulated GWAS association statistics  $\mathbf{Z}$  by calculating  $\hat{\mathbf{R}}$  from the European subset of the 1000 Genomes Project panel, and replacing  $\hat{\mathbf{D}}$  by its limiting value  $\mathbf{D}$  with elements  $D_{kk} = 1 - \alpha_k^2$ .

**Simulating genetic effects at causal loci.** We used empirical functional annotation data to simulate causal genetic effects  $\boldsymbol{\beta}$ , guided by the intuition that a variant's functional effects ultimately determine its effects on complex traits. While minor allele frequency (MAF) was not explicitly used to select causal variants in simulations, this procedure induces an implicit relationship between MAF and causal status due to the relationship between MAF and functional annotations (S4 Fig). For each simulated causal locus, we selected a causal gene by sampling a single CCDS protein-coding gene, and defined proximal genes as any gene with TSS within 1 Mbp of the causal gene TSS. We then simulated single-variant GWAS summary statistics for all variants associated with any causal and proximal genes by proximity ( $\leq 1$  Mbp) or functional annotations (e.g., eQTL variants).

We simulated causal genetic effects under 5 scenarios: 0) no association (null model), 1) coding association, 2) enhancer association, 3) eGene association, and 4) UTR association. For coding and UTR associations, we first selected the number of causal variants  $M^* = \sum_j I(\beta_j^2 > 0)$  from a Poisson distribution with rate parameter  $\lambda = M/4$  truncated to  $1 \leq M^* \leq M$ , where  $M$  is the total number of coding (or UTR) variants for the causal gene, and randomly selected  $M^*$  causal variants from the total set of  $M$  coding (or UTR) variants for the causal gene. This procedure results in  $\sim 25\%$  of all coding (or UTR variants) having non-zero causal effects, while ensuring that at least one variant is causal. For enhancer associations, we similarly simulated the number of causal enhancers  $M_e^*$  from a Poisson distribution with rate parameter  $\lambda = M_e/4$ , where  $M_e$  is the number of enhancers mapped to the causal gene, and selected causal enhancers using a categorical distribution with probability weights derived from confidence scores between enhancer elements and the causal gene. For eGene associations, we selected a single causal tissue at random, and simulated causal effect sizes proportional to precomputed



eQTL weights for the causal gene and tissue. Because eQTL weights are noisy in practice, we used simulated weights  $\tilde{\mathbf{w}} \sim \mathcal{N}_{M^*}(\mathbf{w}, \frac{9}{10N} \hat{\mathbf{R}}^{-1})$  in place of the original weight vector  $\mathbf{w}$  in TWAS gene-based tests, where  $N$  is the GTEx v7 sample size for the causal tissue.

For non-eQTL effects, we simulated the genetic effect for each causal variant  $\beta_j$  from an iid normal distribution, scaled so that the total genetic variance at each locus is equal to  $h_L^2$ . Because our model has assumed genotypes are scaled with unit variance (and  $\beta$  is scaled accordingly), this simulation approach implicitly assumes that the heritability-scale effect of each causal variant is independent of MAF. This is essentially equivalent to the widely-used model  $\text{Var}(\beta_{j,\text{unscaled}}) = \tau^2 [2MAF_j(1 - MAF_j)]^a$ , where  $a = -1$  [26, 66, 67].

## The UK Biobank resource

We used GWAS summary statistics (single-variant association effect size estimates, standard errors, and p-values) for a set of 1,403 traits in the UK Biobank [20] cohort calculated using SAIGE [68]. Genotype data were imputed using the Haplotype Reference Consortium panel [69], and filtered to include only variants with imputed MAC > 20 in the UK Biobank. We selected a subset of 189 traits for primary analysis by including only traits with effective sample size  $\geq 5,000$ , and  $\geq 1$  single-variant association p-value  $\leq 2.5e-8$ . For our analysis of empirical power, we selected a subset of 128/189 traits by iteratively pruning pairs of correlated traits. Beginning with the most highly correlated pair of traits, we retained the trait with the larger number of significant independent single-variant associations (in the case of ties, we selected the trait with the most detailed description), and repeated this procedure until the maximum pairwise correlation-squared between traits was  $\leq 0.10$ . Trait correlations were estimated from GWAS summary statistics as described in [70]. For our analysis of concordance with benchmark genes, we first selected a subset of 47 traits including only traits with  $\geq 1$  single-variant association p-value  $< 5e-10$ , excluding benign neoplasms, and including at most a single trait within each trait category. We identified  $\geq 1$  relevant benchmark genes for 25 of the original 47 traits.

## Selection of benchmark genes

Benchmark genes for each of the selected UK Biobank traits were identified using the ClinVar [21] and Human Phenotype Ontology (HPO) databases [22]. The HPO database explicitly links genes to traits, while the ClinVar database links traits to variants. To identify benchmark genes from ClinVar, we extracted protein-altering variants (frameshift, missense, nonsense, splice site, or stop-loss variants), and excluded variants with unknown or ambiguous molecular consequence (e.g., intergenic and intronic variants). Despite including only ClinVar genes with coding associations, we expect to capture some genes for which both rare coding variants and common regulatory variants contribute to disease risk. For each UK Biobank trait, we extracted all protein-altering ClinVar variants +/- 1 Mbp of a genome-wide significant UK Biobank variant, and manually selected ClinVar traits equivalent or closely related to the corresponding UK Biobank trait. We then annotated genes associated with one or more relevant ClinVar trait as a ClinVar benchmark gene. We identified benchmark genes from the HPO database by manually matching keywords between UK Biobank and HPO traits. A complete list of HPO/ClinVar traits and benchmark genes for each UK Biobank trait is provided in Supplementary Materials.

## Supporting information

**S1 Table. Descriptive statistics for eQTL annotation data sets.** Descriptive statistics for eQTL variant predictive weights used to calculate TWAS test statistics.  
(TEX)

**S2 Table. Descriptive statistics for enhancer-to-target gene annotation data sets.** Descriptive statistics for regulatory element annotation data sets used to calculate weights between enhancers and target genes.

(TEX)

**S1 Fig. GWAS simulations: ROC and precision-recall curves.** Receiver Operating Characteristic (ROC; top) and Precision-Recall (bottom) curves for each gene-based testing approach (curve color) when either coding, eQTL, enhancer, or UTR variants are causal (plot columns) given locus heritability  $h_L^2 = 0.05\%$ ; similar results were obtained for other  $h_L^2$  values. Detailed description of simulation settings is provided under “GWAS Simulations”, and simulation procedures are described in Materials and Methods. To aggregate results across loci and simulation replicates, we use standardized scores for each method calculated by dividing gene-based scores (e.g.,  $-\log_{10}$ -p-values) by the maximum value at the corresponding locus within each replicate. This procedure ensures that curves reflect performance ranking genes at each locus individually. We obtained similar results using the quantile rank of gene-based scores within each locus for each method rather than dividing by the maximum value.

(TIF)

**S2 Fig. UK Biobank: Sensitivity and specificity of gene ranking materials and methods.**

ROC and Precision-Recall curves for each gene-based association or ranking method across benchmark loci present in both HPO and ClinVar (54 loci in total). To aggregate results across benchmark loci and UK Biobank traits, we use standardized scores for each method calculated by dividing gene-based scores (e.g.,  $-\log_{10}$ -p-values) by the maximum value at the corresponding locus. This procedure ensures that curves reflect performance ranking genes at each locus individually. We obtained similar results using the quantile rank of gene-based scores within each locus for each method rather than dividing by the maximum value.

(TIF)

**S3 Fig. Most significant annotation class for benchmark vs. other genes.** Most significant single-annotation test (x-axis) for genes with one or more gene-based p-value  $\leq 5e-6$ . The proportion of benchmark genes (the union of HPO and ClinVar gene lists) and other genes (not present in either benchmark genes list) for which the indicated annotation class is most significant is shown on the y-axis with 95% confidence intervals. Benchmark genes are strongly enriched for coding associations (odds ratio = 5.03, p-value =  $1.3e-16$ ), which is expected due to the selection criteria used to construct benchmark gene lists (described in [Materials and methods](#)).

(TIF)

**S4 Fig. Comparison of MAF across functional annotation categories.** Empirical cumulative distribution function (ECDF) of minor allele frequency (MAF) in the UK Biobank stratified by stratified by functional annotation. Overall, annotated functional variants tend to have lower MAF than intergenic variants, particularly for nonsense and missense variants, as expected.

(TIF)

**S5 Fig. Comparison of CT-TWAS aggregation methods.** Comparison of Cross-Tissue TWAS (CT-TWAS) p-values, and p-values using only the top single tissue, for disorders of lipid metabolism using GWAS summary statistics from the UK Biobank. The top tissue was defined as the tissue with the largest number of significant genes using FWER threshold  $\alpha = 0.05$  with Bonferroni adjustment for the number of eGenes in each tissue. In this case, the top tissue was “Liver” with 27 significant genes out of 3,314 total eGenes (Bonferroni-adjusted p-value threshold =  $1.5 \times 10^{-5}$ ). Top-Tissue p-values are compared with CT-TWAS p-values

(CT-Q, CT-A, and CT-M), which aggregate across all 47 tissues, restricted to Liver eGenes. CT-Q is calculated using the sum of squared single-tissue TWAS z-scores (similar to SKAT); CT-A is calculated by combining single-tissue TWAS p-values using ACAT; and CT-M is calculated from the minimum single-tissue p-value using the multivariate normal joint density of all single-tissue z-scores (described in [Materials and methods](#)). Here, CT-M detected 51 significant genes, followed by CT-A with 47, CT-Q with 33, and top-tissue-only with 27. (TIF)

**S6 Fig. Comparison TWAS/PrediXcan p-values across software.** Comparison of TWAS/PrediXcan p-values calculated by GAMBIT versus S-PrediXcan (cloned from GitHub on April 10, 2020) using GWAS summary statistics for HDL cholesterol from the Global Lipids Genetics Consortium [71]. Results are shown for 25,691 unique genes across 47 tissues using GTEx v7 HapMap predictive weights from PredictDB [6, 7]. Signed  $-\log_{10}(p)$ -values are shown for  $p \geq 10^{-50}$ ; 10 genes with outlying  $p < 10^{-50}$  are not displayed. The squared Pearson correlation between z-scores is 0.995; differences in z-scores between GAMBIT and S-PrediXcan are presumably due to differences in the LD reference data. S-PrediXcan uses precomputed LD files which are packaged together with predictive weights, whereas GAMBIT calculates LD interactively from a reference panel (here, European individuals in the 1000 Genomes Project). (TIF)

**S1 Data. Gene-based test and ranking results across benchmark loci.** (CSV)

## Acknowledgments

We gratefully acknowledge the participants and investigators of the 1000 Genomes Project and UK Biobank study. We thank Yaowu Liu for helpful discussions on p-value combination procedures; Xihong Lin for helpful comments and suggestions on the manuscript; and Sarah Gagliano, Jonas Billie Nielson, and Wei Zhou for assistance with data sets. This research has been conducted using the UK Biobank Resource under Application Number 24460.

## Author Contributions

**Conceptualization:** Corbin Quick, Gonçalo Abecasis, Michael Boehnke, Hyun Min Kang.

**Data curation:** Corbin Quick.

**Formal analysis:** Corbin Quick.

**Funding acquisition:** Michael Boehnke, Hyun Min Kang.

**Investigation:** Corbin Quick, Gonçalo Abecasis, Hyun Min Kang.

**Methodology:** Corbin Quick, Hyun Min Kang.

**Project administration:** Corbin Quick, Michael Boehnke, Hyun Min Kang.

**Resources:** Corbin Quick, Xiaoquan Wen, Gonçalo Abecasis, Michael Boehnke, Hyun Min Kang.

**Software:** Corbin Quick, Xiaoquan Wen, Hyun Min Kang.

**Supervision:** Corbin Quick, Michael Boehnke, Hyun Min Kang.

**Validation:** Corbin Quick, Hyun Min Kang.

**Visualization:** Corbin Quick, Hyun Min Kang.

**Writing – original draft:** Corbin Quick, Hyun Min Kang.

**Writing – review & editing:** Corbin Quick, Xiaoquan Wen, Gonalo Abecasis, Michael Boehnke, Hyun Min Kang.

## References

1. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic acids research*. 2013; 42(D1):D1001–D1006. <https://doi.org/10.1093/nar/gkt1229> PMID: 24316577
2. Neale BM, Sham PC. The future of association studies: gene-based analysis and replication. *The American Journal of Human Genetics*. 2004; 75(3):353–362. <https://doi.org/10.1086/423901> PMID: 15272419
3. Sham PC, Purcell SM. Statistical power and significance testing in large-scale genetic studies. *Nature Reviews Genetics*. 2014; 15(5):335–346. <https://doi.org/10.1038/nrg3706> PMID: 24739678
4. Liu DJ, Peloso GM, Zhan X, Holmen OL, Zawistowski M, Feng S, et al. Meta-analysis of gene-level tests for rare variant association. *Nature genetics*. 2014; 46(2):200–204. <https://doi.org/10.1038/ng.2852> PMID: 24336170
5. Morrison AC, Voorman A, Johnson AD, Liu X, Yu J, Li A, et al. Whole-genome sequence-based analysis of high-density lipoprotein cholesterol. *Nature genetics*. 2013; 45(8):899. <https://doi.org/10.1038/ng.2671> PMID: 23770607
6. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, et al. A gene-based association method for mapping traits using reference transcriptome data. *Nature Genetics*. 2015; 47(9):1091–8. <https://doi.org/10.1038/ng.3367> PMID: 26258848
7. Barbeira AN, Dickinson SP, Bonazzola R, Zheng J, Wheeler HE, Torres JM, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nature Communications*. 2018; 9(1):1825. <https://doi.org/10.1038/s41467-018-03621-1> PMID: 29739930
8. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BW, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nature Genetics*. 2016; 48(3):245–52. <https://doi.org/10.1038/ng.3506> PMID: 26854917
9. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic acids research*. 2016; 45(D1):D896–D901. <https://doi.org/10.1093/nar/gkw1133> PMID: 27899670
10. Ernst J, Kheradpour P, Mikkelsen TS, Shores N, Ward LD, Epstein CB, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*. 2011; 473(7345):43. <https://doi.org/10.1038/nature09906> PMID: 21441907
11. Cao Q, Anyansi C, Hu X, Xu L, Xiong L, Tang W, et al. Reconstruction of enhancer-target networks in 935 samples of human primary cells, tissues and cell lines. *Nature genetics*. 2017; 49:1428. <https://doi.org/10.1038/ng.3950> PMID: 28869592
12. Wainberg M, Sinnott-Armstrong N, Knowles D, Golan D, Ermel R, Ruusalepp A, et al. Vulnerabilities of transcriptome-wide association studies. *bioRxiv*. 2017; p. 206961.
13. Wainberg M, Sinnott-Armstrong N, Mancuso N, Barbeira AN, Knowles DA, Golan D, et al. Opportunities and challenges for transcriptome-wide association studies. *Nature genetics*. 2019; 51(4):592. <https://doi.org/10.1038/s41588-019-0385-z> PMID: 30926968
14. Schork AJ, Thompson WK, Pham P, Torkamani A, Roddey JC, Sullivan PF, et al. All SNPs are not created equal: genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. *PLoS genetics*. 2013; 9(4):e1003449. <https://doi.org/10.1371/journal.pgen.1003449> PMID: 23637621
15. Lu Q, Powles RL, Wang Q, He BJ, Zhao H. Integrative tissue-specific functional annotations in the human genome provide novel insights on many complex traits and improve signal prioritization in genome wide association studies. *PLoS genetics*. 2016; 12(4):e1005947. <https://doi.org/10.1371/journal.pgen.1005947> PMID: 27058395
16. Kichaev G, Bhatia G, Loh PR, Gazal S, Burch K, Freund MK, et al. Leveraging polygenic functional enrichment to improve GWAS power. *The American Journal of Human Genetics*. 2019; 104(1):65–75. <https://doi.org/10.1016/j.ajhg.2018.11.008> PMID: 30595370
17. Lee D, Gorkin DU, Baker M, Strober BJ, Asoni AL, McCallion AS, et al. A method to predict the impact of regulatory variants from DNA sequence. *Nature genetics*. 2015; 47(8):955. <https://doi.org/10.1038/ng.3331> PMID: 26075791

18. Kelley DR, Snoek J, Rinn JL. Basset: learning the regulatory code of the accessible genome with deep convolutional neural networks. *Genome research*. 2016; 26(7):990–999. <https://doi.org/10.1101/gr.200535.115> PMID: 27197224
19. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic acids research*. 2018; 47(D1):D886–D894. <https://doi.org/10.1093/nar/gky1016>
20. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018; 562(7726):203. <https://doi.org/10.1038/s41586-018-0579-z> PMID: 30305743
21. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic acids research*. 2015; 44(D1):D862–D868. <https://doi.org/10.1093/nar/gkv1222> PMID: 26582918
22. Köhler S, Vasilevsky NA, Engelstad M, Foster E, McMurry J, Aymé S, et al. The human phenotype ontology in 2017. *Nucleic acids research*. 2016; 45(D1):D865–D876.
23. International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature*. 2010; 467(7311):52–58. <https://doi.org/10.1038/nature09298> PMID: 20811451
24. 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*. 2015; 526(7571):68–74. <https://doi.org/10.1038/nature15393> PMID: 26432245
25. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *The American Journal of Human Genetics*. 2008; 83(3):311–321. <https://doi.org/10.1016/j.ajhg.2008.06.024> PMID: 18691683
26. Madsen BE, Browning SR. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genetics*. 2009; 5(2):e1000384. <https://doi.org/10.1371/journal.pgen.1000384> PMID: 19214210
27. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *American Journal of Human Genetics*. 2011; 89(1):82–93. <https://doi.org/10.1016/j.ajhg.2011.05.029> PMID: 21737059
28. Lamparter D, Marbach D, Rueedi R, Kutalik Z, Bergmann S. Fast and Rigorous Computation of Gene and Pathway Scores from SNP-Based Summary Statistics. *PLoS Computational Biology*. 2016; 12(1): e1004714. <https://doi.org/10.1371/journal.pcbi.1004714> PMID: 26808494
29. Conneely KN, Boehnke M. So many correlated tests, so little time! Rapid adjustment of P values for multiple correlated tests. *The American Journal of Human Genetics*. 2007; 81(6):1158–1168. <https://doi.org/10.1086/522036> PMID: 17966093
30. Liu Y, Chen S, Li Z, Morrison AC, Boerwinkle E, Lin X. ACAT: A Fast and Powerful P-value Combination Method for Rare-variant Analysis in Sequencing Studies. *bioRxiv*. 2018; p. 482240.
31. Wilson DJ. The harmonic mean p-value for combining dependent tests. *Proceedings of the National Academy of Sciences of the United States of America*. 2019; 116(4):1195–1200. <https://doi.org/10.1073/pnas.1814092116> PMID: 30610179
32. Zhan X, Liu DJ. TaSer (TabAnno and SeqMiner): a toolset for annotating and querying next-generation sequence data. *arXiv preprint arXiv:13065715*. 2013;.
33. Kang H. Efficient and parallelizable association container toolbox (EPACTS). University of Michigan Center for Statistical Genetics Accessed. 2014; 6:16.
34. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015; 518(7539):317. <https://doi.org/10.1038/nature14248> PMID: 25693563
35. Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Iny Stein T, et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database*. 2017; 2017(1). <https://doi.org/10.1093/database/bax028> PMID: 28605766
36. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. The NIH roadmap epigenomics mapping consortium. *Nature biotechnology*. 2010; 28(10):1045–1048. <https://doi.org/10.1038/nbt1010-1045> PMID: 20944595
37. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012; 489(7414):57–74. <https://doi.org/10.1038/nature11247> PMID: 22955616
38. Lizio M, Harshbarger J, Shimoji H, Severin J, Kasukawa T, Sahin S, et al. Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome biology*. 2015; 16(1):22. <https://doi.org/10.1186/s13059-014-0560-6> PMID: 25723102
39. Marbach D, Lamparter D, Quon G, Kellis M, Kutalik Z, Bergmann S. Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. *Nature methods*. 2016; 13(4):366–370. <https://doi.org/10.1038/nmeth.3799> PMID: 26950747

40. 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> PMID: [25954001](https://pubmed.ncbi.nlm.nih.gov/25954001/)
41. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS genetics*. 2014; 10(5):e1004383. <https://doi.org/10.1371/journal.pgen.1004383> PMID: [24830394](https://pubmed.ncbi.nlm.nih.gov/24830394/)
42. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nature genetics*. 2016; 48(5):481. <https://doi.org/10.1038/ng.3538> PMID: [27019110](https://pubmed.ncbi.nlm.nih.gov/27019110/)
43. Lee Y, Francesca L, Pique-Regi R, Wen X. Bayesian Multi-SNP Genetic Association Analysis: Control of FDR and Use of Summary Statistics. *bioRxiv*. 2018; p. 316471.
44. Mahajan A, Wessel J, Willems SM, Zhao W, Robertson NR, Chu AY, et al. Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. *Nature genetics*. 2018; 50(4):559. <https://doi.org/10.1038/s41588-018-0084-1> PMID: [29632382](https://pubmed.ncbi.nlm.nih.gov/29632382/)
45. Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P, et al. A polygenic burden of rare disruptive mutations in schizophrenia. *Nature*. 2014; 506(7487):185. <https://doi.org/10.1038/nature12975> PMID: [24463508](https://pubmed.ncbi.nlm.nih.gov/24463508/)
46. Majithia AR, Flannick J, Shahinian P, Guo M, Bray MA, Fontanillas P, et al. Rare variants in PPARG with decreased activity in adipocyte differentiation are associated with increased risk of type 2 diabetes. *Proceedings of the National Academy of Sciences*. 2014; 111(36):13127–13132. <https://doi.org/10.1073/pnas.1410428111> PMID: [25157153](https://pubmed.ncbi.nlm.nih.gov/25157153/)
47. Stranger BE, Brigham LE, Hasz R, Hunter M, Johns C, Johnson M, et al. Enhancing GTEx by bridging the gaps between genotype, gene expression, and disease. *Nature genetics*. 2017; 49(12):1664. <https://doi.org/10.1038/ng.3969>
48. Feng S, Pistis G, Zhang H, Zawistowski M, Mulas A, Zoledziewska M, et al. Methods for association analysis and meta-analysis of rare variants in families. *Genetic Epidemiology*. 2015; 39(4):227–38. <https://doi.org/10.1002/gepi.21892> PMID: [25740221](https://pubmed.ncbi.nlm.nih.gov/25740221/)
49. Lee S, Wu MC, Lin X. Optimal tests for rare variant effects in sequencing association studies. *Biostatistics*. 2012; 13(4):762–75. <https://doi.org/10.1093/biostatistics/kxs014> PMID: [22699862](https://pubmed.ncbi.nlm.nih.gov/22699862/)
50. Lee S, Wu MC, Lin X. Optimal tests for rare variant effects in sequencing association studies. *Biostatistics*. 2012; 13(4):762–775. <https://doi.org/10.1093/biostatistics/kxs014> PMID: [22699862](https://pubmed.ncbi.nlm.nih.gov/22699862/)
51. Tang ZZ, Lin DY. MASS: meta-analysis of score statistics for sequencing studies. *Bioinformatics*. 2013; 29(14):1803–1805. <https://doi.org/10.1093/bioinformatics/btt280> PMID: [23698861](https://pubmed.ncbi.nlm.nih.gov/23698861/)
52. Morgenthaler S, Thilly WG. A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: a cohort allelic sums test (CAST). *Mutation Research*. 2007; 615(1-2):28–56. <https://doi.org/10.1016/j.mrfmmm.2006.09.003> PMID: [17101154](https://pubmed.ncbi.nlm.nih.gov/17101154/)
53. Liu JZ, Mcrae AF, Nyholt DR, Medland SE, Wray NR, Brown KM, et al. A versatile gene-based test for genome-wide association studies. *The American Journal of Human Genetics*. 2010; 87(1):139–145. <https://doi.org/10.1016/j.ajhg.2010.06.009> PMID: [20598278](https://pubmed.ncbi.nlm.nih.gov/20598278/)
54. Neale BM, Rivas MA, Voight BF, Altshuler D, Devlin B, Orho-Melander M, et al. Testing for an unusual distribution of rare variants. *PLoS Genetics*. 2011; 7(3):e1001322. <https://doi.org/10.1371/journal.pgen.1001322> PMID: [21408211](https://pubmed.ncbi.nlm.nih.gov/21408211/)
55. Davies RB. Algorithm AS 155: The distribution of a linear combination of  $\chi^2$  random variables. *Journal of the Royal Statistical Society Series C (Applied Statistics)*. 1980; 29(3):323–333.
56. Liu H, Tang Y, Zhang HH. A new chi-square approximation to the distribution of non-negative definite quadratic forms in non-central normal variables. *Computational Statistics & Data Analysis*. 2009; 53(4):853–856. <https://doi.org/10.1016/j.csda.2008.11.025>
57. Mishra A, Macgregor S. VEGAS2: Software for More Flexible Gene-Based Testing. *Twin Research and Human Genetics*. 2015; 18(1):86–91. <https://doi.org/10.1017/thg.2014.79> PMID: [25518859](https://pubmed.ncbi.nlm.nih.gov/25518859/)
58. Tippett LHC. *The methods of statistics; an introduction mainly for workers in the biological sciences*. 2nd ed. London,; Williams & Norgate Ltd.; 1931.
59. Liu Y, Xie J. Cauchy Combination Test: A Powerful Test With Analytic p-Value Calculation Under Arbitrary Dependency Structures. *Journal of the American Statistical Association*. 2020; 115(529):393–402. <https://doi.org/10.1080/01621459.2018.1554485> PMID: [33012899](https://pubmed.ncbi.nlm.nih.gov/33012899/)
60. Kölbig KS, Schorr B. A program package for the Landau distribution. *Computer Physics Communications*. 1983; 31(CERN-DD-83-18):97–111.
61. Kwak IY, Pan W. Adaptive gene-and pathway-trait association testing with GWAS summary statistics. *Bioinformatics*. 2015; 32(8):1178–1184. <https://doi.org/10.1093/bioinformatics/btv719> PMID: [26656570](https://pubmed.ncbi.nlm.nih.gov/26656570/)

62. Chen LS, Hsu L, Gamazon ER, Cox NJ, Nicolae DL. An exponential combination procedure for set-based association tests in sequencing studies. *American Journal of Human Genetics*. 2012; 91(6):977–86. <https://doi.org/10.1016/j.ajhg.2012.09.017> PMID: [23159251](https://pubmed.ncbi.nlm.nih.gov/23159251/)
63. Brun R, Rademakers F. ROOT- an object oriented data analysis framework. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. 1997; 389(1-2):81–86. [https://doi.org/10.1016/S0168-9002\(97\)00048-X](https://doi.org/10.1016/S0168-9002(97)00048-X)
64. Liu Y, Sarkar A, Kheradpour P, Ernst J, Kellis M. Evidence of reduced recombination rate in human regulatory domains. *Genome biology*. 2017; 18(1):193. <https://doi.org/10.1186/s13059-017-1308-x> PMID: [29058599](https://pubmed.ncbi.nlm.nih.gov/29058599/)
65. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome research*. 2012; 22(9):1760–1774. <https://doi.org/10.1101/gr.135350.111> PMID: [22955987](https://pubmed.ncbi.nlm.nih.gov/22955987/)
66. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. *American Journal of Human Genetics*. 2014; 95(1):5–23. <https://doi.org/10.1016/j.ajhg.2014.06.009> PMID: [24995866](https://pubmed.ncbi.nlm.nih.gov/24995866/)
67. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*. 2010; 42(7):565–9. <https://doi.org/10.1038/ng.608> PMID: [20562875](https://pubmed.ncbi.nlm.nih.gov/20562875/)
68. Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nature Genetics*. 2018; 50(9):1335–1341. <https://doi.org/10.1038/s41588-018-0184-y> PMID: [30104761](https://pubmed.ncbi.nlm.nih.gov/30104761/)
69. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nature genetics*. 2016; 48(10):1279. <https://doi.org/10.1038/ng.3643> PMID: [27548312](https://pubmed.ncbi.nlm.nih.gov/27548312/)
70. Zhu X, Feng T, Tayo BO, Liang J, Young JH, Franceschini N, et al. Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. *American Journal of Human Genetics*. 2015; 96(1):21–36. <https://doi.org/10.1016/j.ajhg.2014.11.011> PMID: [25500260](https://pubmed.ncbi.nlm.nih.gov/25500260/)
71. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. *Nature genetics*. 2013; 45(11):1274. <https://doi.org/10.1038/ng.2797> PMID: [24097068](https://pubmed.ncbi.nlm.nih.gov/24097068/)