Partitioning Heritability of Regulatory and Cell-Type-Specific Variants across 11 Common Diseases

Alexander Gusev,[1,*] S. Hong Lee,[2] Gosia Trynka,[3,4,5,6,16] Hilary Finucane,[7] Bjarni J. Vilhjálmsson,[1] Han Xu,[8] Chongzhi Zang,[8] Stephan Ripke,[9,10] Brendan Bulik-Sullivan,[9,10] Eli Stahl,[11] Schizophrenia Working Group of the Psychiatric Genomics Consortium, SWE-SCZ Consortium, Anna K. Kähler,[12] Christina M. Hultman,[12] Shaun M. Purcell,[9,10,11] Steven A. McCarroll,[10] Mark Daly,[6,9,10] Bogdan Pasaniuc,[13] Patrick F. Sullivan,[12,14] Benjamin M. Neale,[6,9,10] Naomi R. Wray,[2] Soumya Raychaudhuri,[3,4,5,6,15] and Alkes L. Price[1,6,*]

Regulatory and coding variants are known to be enriched with associations identified by genome-wide association studies (GWASs) of complex disease, but their contributions to trait heritability are currently unknown. We applied variance-component methods to imputed genotype data for 11 common diseases to partition the heritability explained by genotyped SNPs (h_g^2) across functional categories (while accounting for shared variance due to linkage disequilibrium). Extensive simulations showed that in contrast to current estimates from GWAS summary statistics, the variance-component approach partitions heritability accurately under a wide range of complex-disease architectures. Across the 11 diseases DNaseI hypersensitivity sites (DHSs) from 217 cell types spanned 16% of imputed SNPs (and 24% of genotyped SNPs) but explained an average of 79% (SE = 8%) of h_g^2 from imputed SNPs (5.1× enrichment; $p = 3.7 \times 10^{-17}$) and 38% (SE = 4%) of h_g^2 from genotyped SNPs (1.6× enrichment, $p = 1.0 \times 10^{-4}$). Further enrichment was observed at enhancer DHSs and cell-type-specific DHSs. In contrast, coding variants, which span 1% of the genome, explained <10% of h_g^2 despite having the highest enrichment. We replicated these findings but found no significant contribution from rare coding variants in independent schizophrenia cohorts genotyped on GWAS and exome chips. Our results highlight the value of analyzing components of heritability to unravel the functional architecture of common disease.

Introduction

Recent work by ENCODE and other projects[1,2] has shown that specific classes of variants can have complex and diverse impacts on cell function and phenotype.[3–10] Although the importance of coding variation has long been understood, these projects identified other genomic regions that can contribute to function and highlighted the role of regulatory variants. With many potentially informative functional categories and competing biological hypotheses, quantifying the contribution of variants in these categories to heritability of complex traits would inform trait biology and focus genetic mapping.

The availability of significantly associated variants from hundreds of genome-wide association studies (GWASs)[11] has opened one avenue for quantifying enrichment. Indeed, 11% of GWAS hits lie in coding regions,[11] 57% of noncoding GWAS hits lie in broadly defined DNaseI hypersensitivity sites (DHSs; spanning 42% of the genome),[3] and still additional GWAS hits tag these regions. The full distribution of GWAS association statistics exhibits enriched p values in coding regions and UTRs.[12] Analysis of DHS subclasses and other histone marks has revealed a complex pattern of cell-type-specific relationships with known disease associations.[6] Recent work has also shown that functional enrichment can be leveraged for increasing association mapping power.[13]

Although relative enrichment has been documented, the question of how much each category contributes to disease heritability remains unanswered.[14,15] Recently, investigators have used variance-component methods to estimate the total additive variance explained by all genotyped SNPs (h_g^2),[16,17] and to estimate the h_g^2 of many quantitative and dichotomous traits.[18–22] We propose joint estimation of h_g^2 from functional-category-specific variance components for assessing enrichment. In contrast to analyses of top GWAS hits, the variance-component approach leverages the entire polygenic architecture of each trait and accounts for pervasive linkage disequilibrium (LD) across functional categories. Indeed, our

[1]Harvard School of Public Health, Boston, MA 02115, USA; [2]The University of Queensland, Queensland Brain Institute, Brisbane, QLD 4072, Australia; [3]Division of Genetics, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA; [4]Division of Rheumatology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA; [5]Partners Center for Personalized Genetic Medicine, Boston, MA 02115, USA; [6]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; [7]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02142, USA; [8]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard School of Public Health, Boston, MA 02115, USA; [9]Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA; [10]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; [11]Department of Psychiatry, Mount Sinai School of Medicine, New York, NY 10029, USA; [12]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm 17177, Sweden; [13]Department of Pathology and Laboratory Medicine, Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA; [14]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA; [15]Faculty of Medical and Human Sciences, University of Manchester, Manchester M13 9PL, UK
[16]Present address: Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge CB10 1SA, UK
*Correspondence: agusev@hsph.harvard.edu (A.G.), aprice@hsph.harvard.edu (A.L.P.)
http://dx.doi.org/10.1016/j.ajhg.2014.10.004.

simulations showed that this approach provides accurate genome-wide estimates of functional enrichment in diverse genetic architectures. We applied variance-component methods to functional categories in GWAS- and exome-chip data from over 100,000 samples in 11 traits.

Material and Methods

Estimating Enrichment of h_g^2 with Variance Components

For a single component of genotyped (or imputed) SNPs, we define h_g^2, an underlying parameter in the population, as the r^2 between the true phenotype and the best linear prediction over those SNPs. With multiple components, the goal of the partitioned analysis is to quantify the h_g^2 directly explained by SNPs in each functional category while excluding tagging of SNPs in other categories. We thus define the h_g^2 for each functional category as the r^2 between the true phenotype and the prediction only from SNPs in that functional category when all functional categories are jointly analyzed for a best linear prediction. When SNPs are in LD, this definition remains valid as long as the individual causal effect sizes are independent, as we would expect in highly polygenic traits. For disease traits, we model the phenotype (and corresponding h_g^2) by using the liability-threshold model, in which individuals whose underlying unobserved continuous liability exceeds a threshold are labeled as disease case subjects.[19,23]

We estimate h_g^2 jointly across multiple variance components, each constructed from variants belonging to nonoverlapping functional categories. The underlying model assumes that SNP effect sizes are drawn from a normal distribution with category-specific variance. (We note that the normality assumption is unrealistic; previous work in the single-variance-component case has indicated that this does not introduce bias, although modeling a more realistic mixture distribution can increase precision.[24] Because of computational constraints, we do not consider mixture distributions here.) The model relates the observed phenotypic covariance to a weighted sum of genetic relationship matrices computed from SNPs in each category. The joint estimate allows all components to compete for shared variance due to LD.

Formally, for a functional categories each containing the set of SNPs S_i (of size M_i), we model the phenotype as a sum of individual SNP effect sizes:

$$y = \sum_{i=1}^{a} \sum_{s \in S_i} W_s \beta_s^i + e,$$

where W_s is the genotype at SNP s, β_s^i is the effect size at SNP s in category i and is drawn from category-specific normal distribution $\beta^i \sim N(0, \sigma_i^2)$, and e is the residual effect $e \sim N(0, \sigma_e^2)$. We assume that for each annotation i, SNPs normalized to have mean 0 and variance 1 are contained in the matrix W_i. The variance of the phenotype is then modeled as

$$V(y) = \sum_{i=1}^{a} K_i \sigma_i^2 + e,$$

where each K_i represents a genetic-relationship matrix (GRM) computed directly from the SNPs in annotation i as

$$K_i = W_i W_i' / M_i.$$

The corresponding σ are then jointly inferred with the REML algorithm in GCTA (Genome-wide Complex Trait Analysis),[16,17] yielding

$$h_{gi}^2 = \frac{\sigma_{gi}^2}{\sum_{j=1}^{a} \sigma_{gi}^2 + \sigma_e^2}.$$

The inverse of the final average-information matrix yields an estimate of the corresponding error-covariance matrix of the variance-component estimates.[25] We use the error-covariance matrix and delta method[26] to compute SEs on h_g^2 and the percentage of h_g^2 while accounting for error correlations (referred to here as the analytical SE[27]). All estimates of h_g^2 were transformed to the liability scale[19] with the prevalence values in Table S1 (available online). We evaluated the accuracy of the analytical SE for both quantitative and ascertained traits and found it to correspond well to the true SD under reasonable polygenicity (see Appendix A). Meta-analysis estimates were computed with inverse-variance weighting:[28] given individual study estimates h_{gi}^2, analytical SE_i, and corresponding weight $w_i = 1/\mathrm{SE}_i^2$, the meta-analysis mean is equal to

$$\frac{\sum_i w_i \times h_{gi}^2}{\sum_i w_i},$$

and the meta-analysis SE is equal to $\sqrt{1/\sum_i w_i}$.

Enrichment is computed for each category i as the ratio of the percentage of h_{gi}^2 (the percentage of h_g^2 in category i) to the percentage of SNP_i (the percentage of SNPs in category i) and is tested for significance by Z score relative to a null of 1:0 with the (likewise-rescaled) analytical SE. Under the assumption that all causal variants are typed, this statistic is equivalent to the relative risk that a SNP in category i is causal in comparison to an average SNP. To achieve unbiasedness, the estimate of h_g^2 is not constrained to lie inside the plausible 0–1 bound, which can lead to negative estimates in rare instances.

Estimating Enrichment from Summary Statistics

We considered alternative methods for estimating functional enrichment from summary association statistics. The simplest approach is to directly count the number of individual genome-wide-significant variants in each functional annotation and compare to the null expectation from all SNPs (or random SNPs matched on certain features). This approach can either include all significant markers or restrict to the most significant variant in each locus. The genome-wide-significant-SNP approach has been extended to the full distribution of association statistics for quantifying overall p value enrichment.[3] Over increasingly restrictive p value thresholds, the fraction of SNPs passing a given threshold and belonging to each category is computed and normalized by the category-specific genome-wide fraction. The distributions are then inspected visually for enrichment or assessed by permutation. For completeness, we considered two additional methods—stratified quantile-quantile (Q-Q) plots[12] and Bayesian hierarchical modeling (fgwas)[13]—which assess functional enrichment but are primarily focused on improving association mapping power (see Discussion).

Data Sets Analyzed

11 Diseases from WTCCC1 and WTCCC2

We analyzed seven traits from Wellcome Trust Case Control Consortium 1 (WTCCC1) and four traits from WTCCC2 for a total

47,000 samples (Table S1). Estimates of h_g^2 are particularly sensitive to individually small artifacts or batch effcts,[19,29] and we followed the rigorous quality-control (QC) protocol outlined previously[21] by removing any SNPs that were below a minor allele frequency (MAF) of 0.01, were above 0.002 missingness, or deviated from Hardy-Weinberg equilibrium at a p value below 0.01. For each case-control cohort, we removed SNPs that had differential missingness with a p value below 0.05. We excluded one of any pair of samples with kinship entries ≥ 0.05[19] and performed five rounds of outlier removal whereby all individuals more than 6 SDs away from the mean along any of the top 20 eigenvectors were removed and all eigenvectors were recomputed[30] (Figure S1). For all autoimmune diseases analyzed (rheumatoid arthritis [RA], Crohn disease [CD], type 1 diabetes [T1D], ulcerative colitis [UC], multiple sclerosis [MS], and ankylosing spondylitis [AS]), we also excluded from the analysis any SNPs in the well-studied major histocompatibility complex (MHC) locus (chr6: 26–34 Mb), which is known to have a complex LD structure, and many heterogeneous variants of strong effect for these traits.

The WTCCC1 samples were phased and imputed as described in Gusev et al.[21] The WTCCC2 samples were split into two cohorts by platform, and each cohort was imputed separately according to the following protocol. All samples in a cohort were phased together in 10 Mb blocks with HAPI-UR (Haplotype Inference for Unrelated Samples)[31] (see Web Resources) and three rounds of phasing and consensus voting. All phased samples in a cohort were then imputed in 1 Mb blocks with IMPUTE2[32] (see Web Resources) and the 1000 Genomes[33] Phase I integrated haplotypes (September 2013 release; see Web Resources) with no singletons. Where relevant, the Oxford recombination map[34] was used. Markers with an information (info) score greater than 0.5 were retained. Finally, SNPs were excluded if they met any of the following criteria in any case or control population: Hardy-Weinberg p value < 0.05, per-locus missingness > 0.05, MAF < 0.01, or case-control differential missingness p value < 0.05.

Schizophrenia Cohort from the Psychiatric Genomics Consortium

We analyzed 24,926 schizophrenia (SP) subjects and 33,271 control individuals from 33 cohorts from the Psychiatric Genomics Consortium (PGC2); they were typed on a variety of platforms, quality controlled, and imputed to the 1000 Genomes reference panel as previously described[35] (Tables S1 and S2). Because of computational constraints, we split the cohort into four subgroups of individuals typed on similar platforms; each contained roughly 10,000–20,000 samples. We performed all analyses on the intersection of well-imputed SNPs within each subgroup, ranging from four to five million, and reported meta-analyzed results. Individual study identifiers and 20 multidimensional-scaling components were included as fixed-effect covariates in all analyses.

Swedish SP Exome Chip

We analyzed 12,674 Swedish samples typed on GWAS and exome chips (Tables S1 and S3). The exome chip yielded 238,652 SNPs (including monomorphic sites), of which 10,567 were also typed on a mix of Affymetrix GWAS chips (exome-chip calls were retained). The GWAS-chip data contained an intersection of 163,051 SNPs typed on all platforms in addition to per-platform imputation from 1000 Genomes for a total of 5,053,934 SNPs imputed on all platforms. Principal-component analysis (PCA) of the GWAS data revealed a large cluster of "homogenous" Swedish samples and clines related to Northern Swedish and Finnish admixture (Figure S2). After excluding samples that (1) were not typed on both GWAS and exome chips, (2) failed QC, (3) were PCA outliers by 6 SDs, or (4) were in a pair with GRM values $>$ 0.05 (close relatives), we retained a total of 8,967 samples, of which 6,375 were of "homogenous" Swedish ancestry. In all of our analyses, rare variants had a MAF < 0.01, and common variants had a MAF ≥ 0.01. Simulations were performed on the homogenous samples (without principal components). We performed analyses of real phenotypes on the homogeneous samples and included the top 20 principal components as covariates (to account for any residual population structure; analyses on the full cohort are reported in Tables S23 and S25).

Functional Annotations

We annotated the genome by using six primary categories (Table S4): (1) coding, (2) UTR, (3) promoter, (4) DHS in any of 217 cell types, (5) intronic, and (6) intergenic. Each SNP was then assigned a unique annotation defined by the first of these categories with which it was annotated, resulting in six nonoverlapping variance components (the DHS category was thus restricted to distal regions). Each resulting category exhibited similar average allele frequency and imputation accuracy, although the DHS category had systematically lower LD[36] (Table S5). We also computed the "effective" number of SNPs in each category by using an LD-based metric that does not depend on sample size.[36,37] Table S6 shows that this metric was not substantially different from the actual percentage of SNPs used in imputed data, given that DHSs harbored slightly more effective SNPs (15.7% SNPs versus 18.9% effective SNPs) as a result of lower LD. For the imputed categories analyzed here, the differences in the percentage of SNPs, percentage of effective SNPs, and percentage of physical size were relatively minor. A greater difference was observed for genotyped SNPs: 23.6% of DHS SNPs corresponded to 33.6% of effective SNPs, suggesting that DHS enrichments from genotyped data might be indicative of better tagging.

For the DHS annotation, we used DNase sequencing libraries downloaded from ENCODE and Epigenome Roadmap projects in May 2012 and merged biological replicates into a single library (GEO accession numbers are available in Table S7). We used BOWTIE v.1.0[38] to align raw read sequences to UCSC Genome Browser hg19 and used MACS v.2.0 with false-discovery rate $<$ 0.01 (the default cutoff) and Benjamini–Hochberg correction[39] to call DHS peaks. For the primary analysis, all peaks were merged into a single DHS annotation spanning 16% of the genome. We note that 98% of the primary DHS annotation was covered by the DHSs released by Maurano et al.[3] (spanning 37% of the genome), and 67% of the primary DHS annotation was covered by the DHSs analyzed in Thurman et al.[4] (spanning 15% of the genome). For the cell-type-specific analysis, duplicate lines were merged to form a final set of 83 unique cell types. The resulting annotations are available for download (see Web Resources).

Segway-chromHMM combined genome segmentations[40] were downloaded for six cell lines (see Web Resources). All regions classified as enhancers or weak enhancers were then combined into a single enhancer annotation. DNaseI digital genomic footprinting (DGF) regions were downloaded for 57 cell lines (see Web Resources). All regions from the narrow-peak classification were then merged into a single DGF annotation.

Simulation Framework

The goal of our simulations was to demonstrate that the partitioned h_g^2 properly recovers the heritability explained by causal variants in a given functional category under a variety of disease architectures. We performed simulations in genotyped and

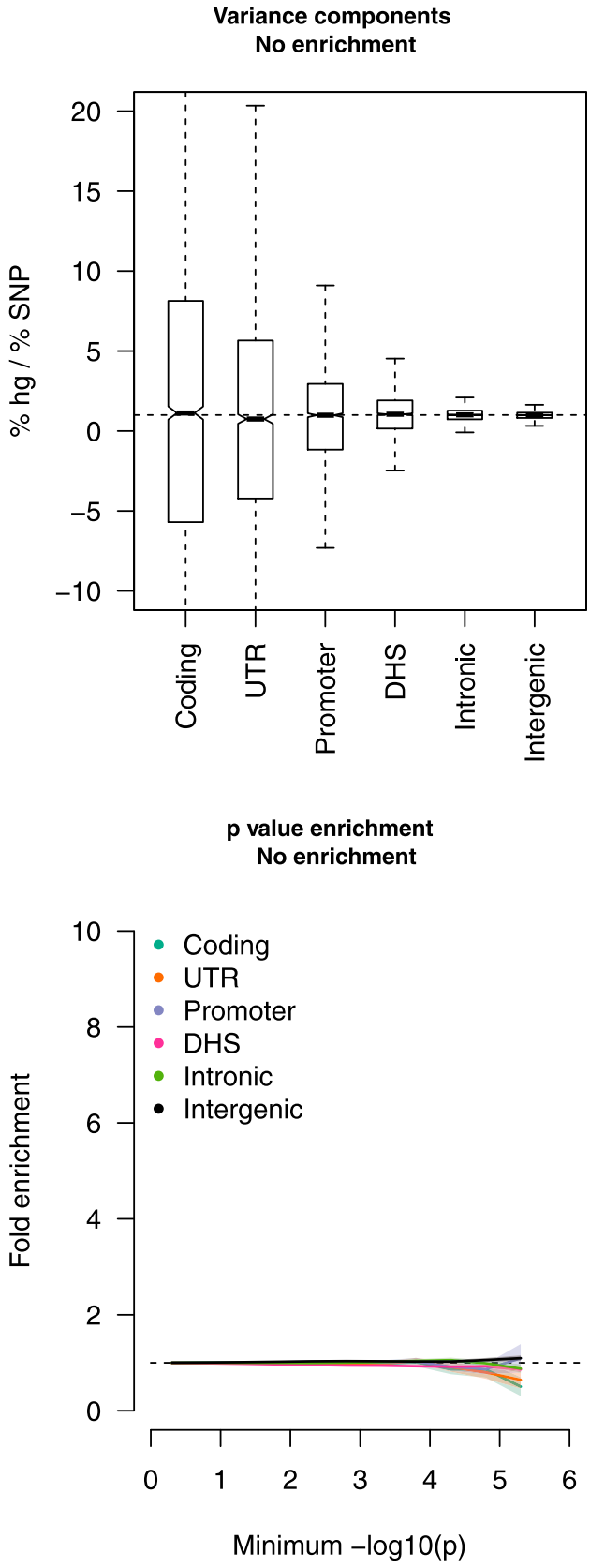

Figure 1. Estimates of Functional Enrichment under the Null
We simulated a polygenic disease architecture in imputed data with no functional enrichment (see text). Simulated phenotypes were tested with the variance-component method (top) from 3,000 simulations and with p value enrichment (bottom) from 100 simulations. In the variance-component subplot, the thin line represents the median, boxes represent the first and third quartiles, and whiskers represent the 1.5× interquartile range from the first to the third quartile. A subplot of p value enrichment shows 1.96× SE as shaded regions.

imputed data in 4,414 samples from the WTCCC1 coronary artery disease (CAD) case-control cohort together with the six main functional annotations to evaluate robustness and accuracy of the proposed variance-component method and the p-value-enrichment approach; we note that the genome-wide-significant-SNP approach is subsumed by the latter and is not reported separately in most analyses. For each simulation, 10% of the (genotyped or imputed) SNPs were randomly sampled to be causal, and normally distributed effect sizes were assigned to each SNP such that each explained equal variance in expectation. Additive phenotypes were then constructed, and random noise was added for an overall h_g^2 of 0.50. Except when evaluating h_g^2 between genotyped and imputed SNPs, we did not hide causal variants from the analyses, corresponding to the assumption that all causal variants are typed. We evaluated the variance-component model by using multiple components with GCTA in the unconstrained mode. For approaches based on summary statistics, we computed *Z* scores, SEs, and p values for the univariate regression of each SNP to a simulated phenotype.

Results

Simulations

We first evaluated the calibration of the methods in simulations of no enrichment by assuming a MAF-independent architecture where causal variants were uniformly sampled from the genome (see Material and Methods). We observed no significant deviations from the null for any categories estimated by variance components or p value enrichment (Figure 1). To evaluate possible biases due to MAF-dependent architectures,[21,41,42] we also considered a low-frequency architecture where only SNPs with a MAF below 0.05 can be causal and a DHS-low-frequency architecture where causal DHS variants are drawn from MAF below 0.05 and all other variants are drawn from any MAF (Figures S3 and S4). Results were generally similar to the MAF-independent architecture, although variance-component estimates exhibited slight but statistically significant deviations for the promoter and UTR categories, which were very small and in tight LD with each other.

We next considered simulations with maximal enrichment, where all causal variants were drawn from a single functional category. MAF-independent results for the coding and DHS categories are shown in Figure 2 (see Figure S5 for other results). The variance-component estimate of the percentage of h_g^2 was again around 100% for the true causal category and 0% for all others. The plots of p value enrichment correctly demonstrated significant enrichment for five of the categories, but not the DHS category, which, when causal, was not significantly different from the null. This lack of enrichment at DHSs and not at other large categories was most likely due to the uniquely lower

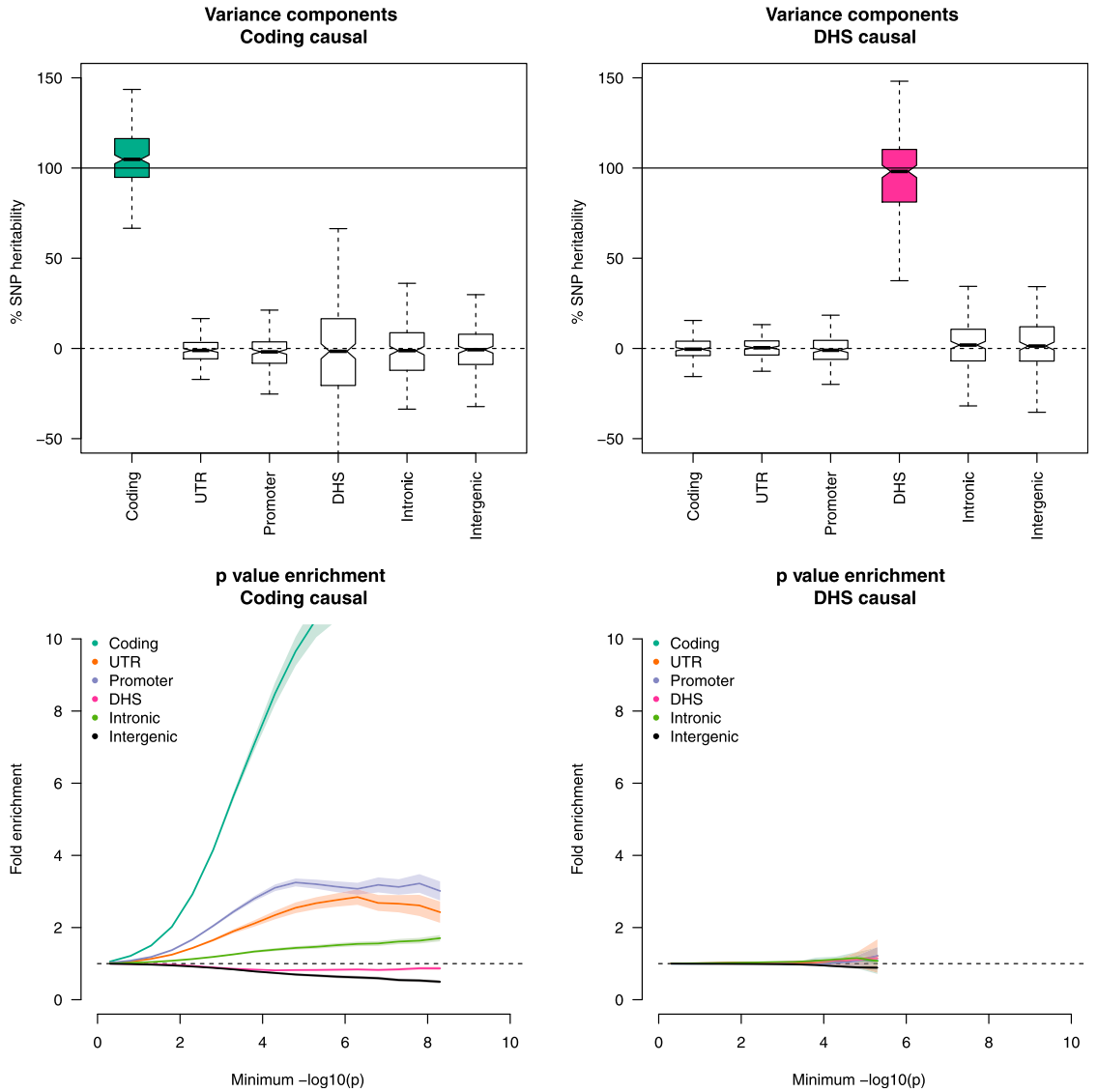

Figure 2. Estimates of Functional Enrichment from a Single Causal Category
We simulated a polygenic disease architecture in imputed data with causal SNPs drawn from a single functional category, corresponding to complete enrichment. Simulated phenotypes were tested with the variance-component method (top) from 200 simulations and with p value enrichment (bottom) from 100 simulations. In the variance-component subplot, the thin line represents the median, boxes represent the first and third quartiles, and whiskers represent the 1.5× interquartile range from the first to the third quartile. Subplots of p value enrichment show 1.96× SE across simulations as shaded regions. For each method, only the coding-causal and DHS-causal scenarios are shown (additional simulations appear in Figures S6 and S7).

LD of DHS SNPs (Table S5). For the small categories (coding, UTR, and promoter), true causals in one category always yielded false p value enrichment in the others because of their close proximity and high LD (Figure 2; Figures S6 and S7). In the MAF-dependent scenarios, the variance-component estimate of h_g^2 was nearly unbiased: it had slight but significant inflation at the coding and UTR categories when they contained 100% of h_g^2 (Figure S8). Plots of p value enrichment exhibited similar patterns as in the MAF-independent simulations, and the DHS category was further falsely depleted (Figure S7).

To investigate the differences between genotype- and imputation-based estimates, we partitioned h_g^2 of category-specific phenotypes simulated from imputed SNPs by using components constructed from genotyped SNPs only. If the genotypes are reasonable proxies for imputed variants, 100% of h_g^2 should again be partitioned into each truly causal category. Instead, we observed significant deviations for all of the categories, and h_g^2 was partitioned into nearby categories as a result of incomplete tagging (Figures S9 and S10). In particular, less than half of the h_g^2 at imputed DHSs was partitioned into the DHS category in genotype data. Thus, estimates produced with only genotyped SNPs can severely underestimate enrichment. The difference between genotyped and imputed simulations suggests that estimates from imputed SNPs could also underestimate the true enrichments or depletions for rare causal variants that are absent from 1000 Genomes or are

poorly imputed. We investigated this possibility by using the exome-chip SP data (see below). We separately assessed the impact of imputation error by simulating phenotypes with induced genotype noise proportional to the per-SNP imputation quality score (info score; Supplementary information S3 in Marchini et al.[43]) but observed no significant biases in null or causal simulations (Tables S8 and S9), most likely as a result of the stringent postimputation QC.

We evaluated multiple other complex architectures with respect to LD (see Appendix A) but observed significant bias in only one deliberately severe scenario: causal variants sampled from intronic and intergenic regions either directly adjacent to or proximal to a DHS (within 1 kb of a DHS boundary). Although no substantial false DHS heritability was observed in genotyped SNPs, the imputed DHS component picked up 50% (0–500 bp) and 20% (500–1,000 bp) of the non-DHS h_g^2 (Figure S11). Given our findings that genotyped SNPs are expected to greatly underestimate DHS enrichment, we consider genotyped and imputed estimates to be lower and upper bounds, respectively, on the true causal enrichment.

Heritability of Functional Categories across 11 Diseases

We analyzed a total of 11 WTCCC1 and WTCCC2 phenotypes.[44–46] After QC,[21] the seven WTCCC1 traits each included an average of 1,700 affected subjects and a set of 2,700 shared control subjects; the four WTCCC2 traits included 1,800–9,300 affected subjects and 5,300 shared control subjects (see Material and Methods; Table S1). In all analyses of autoimmune traits, SNPs in the well-studied MHC region were excluded, although inclusion of the MHC as a separate component did not significantly affect the results. Each cohort was imputed to the 1000 Genomes reference panel, yielding four to six million SNPs per trait after QC (see Material and Methods; Table S1). This analysis is expected to be skewed toward the autoimmune traits, which composed 6/11 traits analyzed and 20,461/30,158 affected subjects analyzed. We computed meta-analysis values by using inverse-variance weighting with the analytical SE to account for different levels of error across h_g^2 estimates. After meta-analysis, resulting SEs were adjusted for the use of shared controls by genomic control (unless otherwise stated), and p values were computed by a simple Z score comparing the mean enrichment and adjusted SE to a null of 1.0 enrichment. Estimating enrichment from shifted functional annotations yielded null enrichments and p values (Tables S10 and S11), confirming that this null is comparable to random SNP comparisons used in previous work.[3,11,40,47]

Combined results meta-analyzed across all traits are reported in Figure 3 (Tables S10, S12, and S13). In genotyped data, DHS variants (spanning 24% of genotyped SNPs) were the most significantly enriched and explained an average of 38% (SE = 4%) of the total h_g^2, a 1.6× enrichment ($p = 1.0 \times 10^{-4}$). Coding variants were the only other category significantly enriched (after six tests were accounted for) and explained 4% (SE = 1%; $p = 1.1 \times 10^{-3}$). All enrichments or depletions were stronger when imputed SNPs were analyzed in terms of both significance and information content, consistent with our previous simulations (Figures S9 and S10; Table S16). Variants in DHSs again exhibited the greatest h_g^2 and most significant enrichment: imputed DHS SNPs explained an average of 79% (SE = 8%) of the total h_g^2, a 5.1× enrichment ($p = 3.7 \times 10^{-17}$). The enrichment varied across traits (Figure S12; Table S14), and there was a nominally significant difference between the six autoimmune traits (AS, CD, MS, RA, T1D, and UC) and the five nonautoimmune traits (SP, bipolar disorder, CAD, hypertension, and type 2 diabetes [T2D]) at 5.5× and 3.3×, respectively ($p = 0.01$ for difference without accounting for shared control subjects). Coding variants exhibited the greatest overall enrichment at 13.8× ($p = 1.8 \times 10^{-3}$) but accounted for 8% of h_g^2 because of the much smaller category size. Correspondingly, we observed a significant depletion for both intronic regions (0.1×; $p = 4.9 \times 10^{-9}$) and intergenic regions (−0.1×; $p < 10^{-20}$) and h_g^2 that was not significantly different from 0. We note that compared to genotyped SNPs, imputation in these traits generally does not explain additional h_g^2,[21] but it can more precisely partition heritability into functional categories. We performed additional simulations mimicking the enrichment observed in imputed data with 8,300 causal variants (as inferred in a large GWAS of a polygenic trait[48]) and found that 79% of heritability was explained by imputed DHS SNPs, 8% was explained by imputed coding SNPs, and the remainder was uniformly drawn from the other variant categories. This "realistic" scenario yielded much weaker estimates of enrichment from genotyped SNPs, and they were similar to estimates from genotyped SNPs in real data (Figure 3).

We considered alternative estimation procedures to rule out potential biases. Although we allowed individual values of h_g^2 to fluctuate outside the 0–1 bound on variance to achieve unbiased estimates prior to averaging across traits,[49] a constrained analysis yielded similar results (see Table S15). Individual point estimates escaping the 0–1 bound were consistent with our imputed simulations under realistic enrichment, which showed that the percentage of h_g^2 for DHSs exceeded 1.0 10% of the time, whereas the percentage of h_g^2 for intronic and intergenic regions fell below 0.0 30% and 23% of the time, respectively, for a typical 7,000-sample cohort. Using flat instead of inverse-variance weighting yielded a comparable estimate such that DHS SNPs explained an average of 85% (SE = 15%) of h_g^2. With the flat weighting, the SD of imputed DHS estimates across different traits was 48%, which corresponds to a SD of 32% in the true unobserved values after the analytical SE of each estimate is accounted for (Table S14). We further evaluated the robustness of these estimates and found that biases arising from analytical SEs, ancestry, or case-control ascertainment were unlikely to significantly affect the enrichment (see Appendix A).

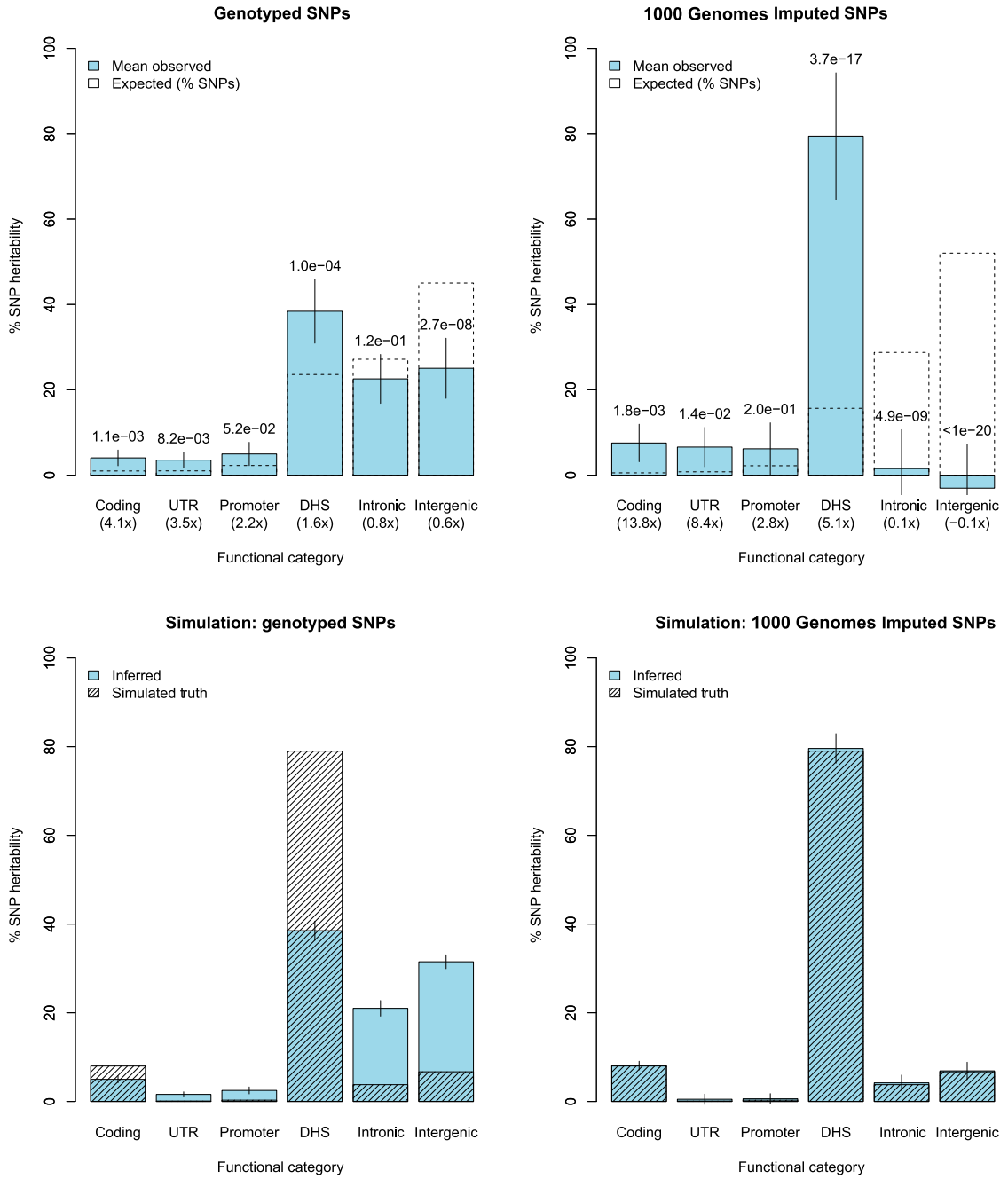

Figure 3. Functional Partitioning of SNP Heritability across 11 Traits

(Top panels) Joint estimates of the percentage of h_g^2 from six functional components are shown in filled bars (meta-analyzed over 11 traits). The null expectation, equal to the percentage of SNPs in each category, is shown by dashed, unfilled bars, and p values report the difference from this expectation. Fold enrichment relative to the null expectation is shown in parentheses below each category. The left panel shows results from analyses of genotyped SNPs only, and the right panel shows analysis of genotyped and 1000 Genomes imputed SNPs. Error bars show 1.96× SE after adjustment for shared controls.

(Bottom panels) Partitioned h_g^2 in simulations of a "realistic" trait where DHS and coding variants explained 79% and 8% of h_g^2, respectively (with no enrichment elsewhere). Filled bars show the mean inferred percentage of h_g^2 from genotyped (left) and imputed (right) SNPs over 100 simulations. Patterned bars show the simulated true partition. Error bars show 1.96× SE (on average, SEs on imputed data were 2.2× higher than SEs on genotype data as a result of the abundance of new variants).

To investigate whether enrichment in h_g^2 from all SNPs at known loci was consistent with the genome-wide estimates, we partitioned the h_g^2 explained by SNPs within 1 Mb of published GWAS loci for each trait (NHGRI GWAS catalog;[11] see Web Resources) (Figure S13). Because some traits had a small number of loci, the DHS component was jointly analyzed with only a single other component containing all non-DHS SNPs. We again observed a highly significant DHS enrichment in imputed data and a significant difference between the genotyped and imputed

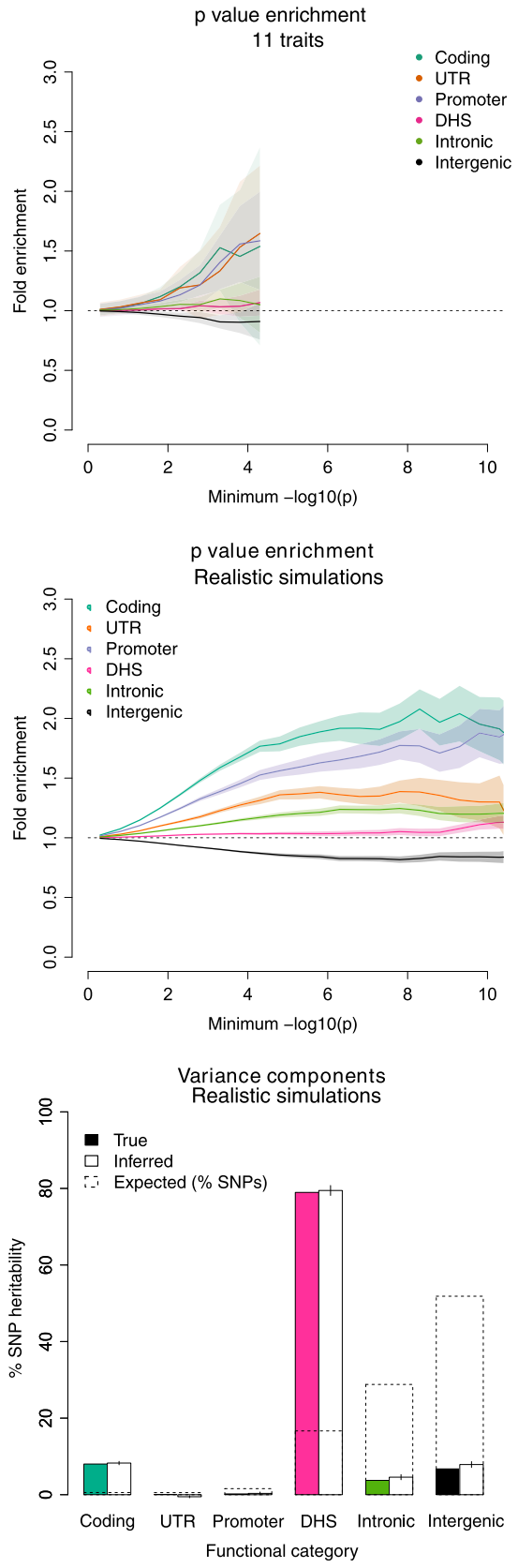

Figure 4. Enrichment from GWAS Summary Statistics
(Left panel) Estimates of p value enrichment are averaged over 11 analyzed traits and are restricted to minimum p value thresholds (x axis) for which at least one association meeting the threshold was observed in every trait.

estimates ($p = 7.3 \times 10^{-14}$). We observed a marginally significant difference between the DHS enrichment at known loci versus genome-wide in the imputed data (3.6× versus 5.5×, p = 0:003). Although it does not pass multiple-test correction, this p value suggests that genome-wide-significant SNPs of large effects might be less enriched with DHS variants than the rest of the genome.

We have shown by simulation that estimates from genotyped SNPs are expected to provide a lower bound on enrichment or depletion and that estimates from imputed SNPs are biased upward only when causal variants are very close to the annotation boundary. For brevity, subsequent results focus primarily on the analysis of imputed SNPs.

Comparison to Estimates of Enrichment from Summary Statistics

We compared our imputed variance-component estimates of 5.1× DHS enrichment for the 11 traits to the DHS enrichment of genome-wide-significant variants identified in these data or from published loci (NHGRI GWAS catalog;[11] see Web Resources). The enrichments from genome-wide-significant variants were much smaller (0.91× and 1.74× for variants in these data and published loci, respectively; Table S17). This is roughly consistent with previous results indicating that 57% of noncoding GWAS hits (from any trait) lie in broadly defined DHSs spanning 42% of the genome (1.4× noncoding enrichment; 1.2× overall enrichment) and that this percentage increases to 77% of noncoding GWAS hits when SNPs in perfect LD with a DHS SNP are included (1.8× noncoding enrichment; 1.6× overall enrichment).[3] Similarly, 30% of the noncoding GWAS hits analyzed in Maurano et al.[3] lay in our DHS annotation, yielding a comparable 1.8× noncoding enrichment. Extending to the full distribution of association statistics did not reveal significant DHS enrichment in any of these traits (Figure 4, left panel; Figure S14). This is consistent with our previous simulations showing the variance-component approach to be more effective than the p-value-enrichment approach at identifying DHS enrichment from complex-disease architectures (Figure 2).

We sought to further confirm this observation by extending our simulations to a single large cohort with realistic levels of enrichment on the basis of the above results. We simulated the "realistic" level of enrichment (see above) in 33,000 combined WTCCC2 samples, corresponding to a large GWAS. We then conducted a standard GWAS on the simulated traits and plotted functional enrichment by using p value enrichment (see Material and Methods). The strategy yielded enrichment at coding

(Middle panel) p value enrichment from a "realistic" simulation. (Right panel) Variance-component enrichment from a "realistic" simulation. Realistic traits were simulated with DHS and coding variants explaining 79% and 8% of h_g^2, respectively, and with computed GWAS statistics in a cohort of 32,000 samples. Shaded regions and error bars represent the SE from meta-analysis (left) and 50 replicates (middle and right).

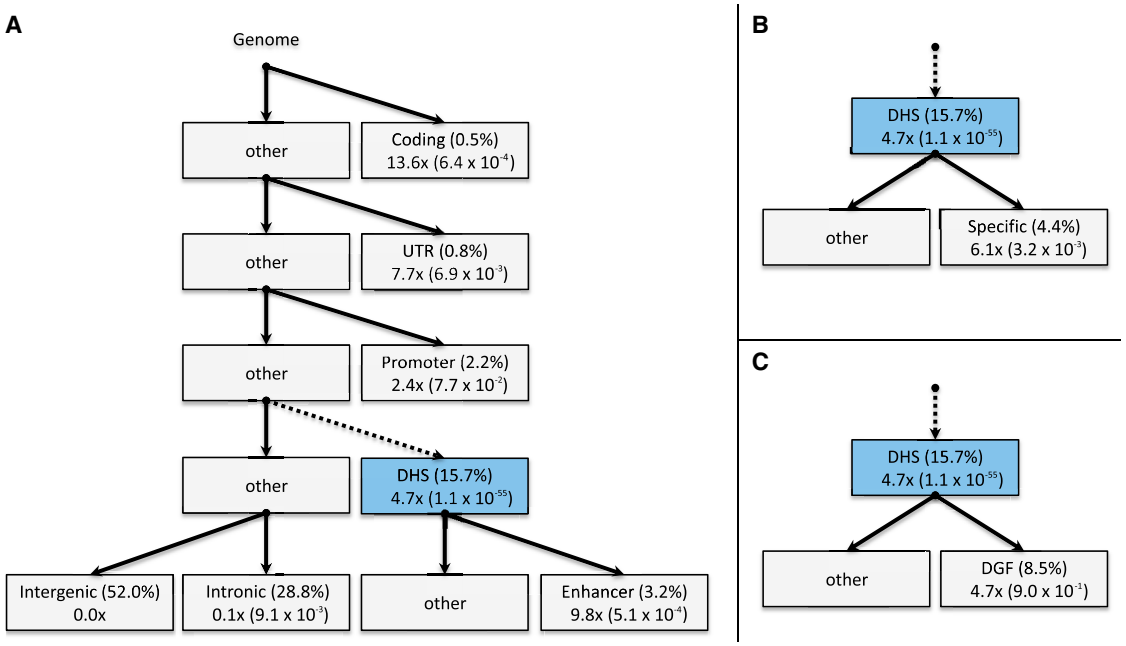

Figure 5. Hierarchical Analysis of Functional Enrichment
DHS variants were further partitioned into three subcategories: predicted enhancers (A), cell-type-specific DHSs (B), and DGF targets (C). Each block contains (on the top line) the functional category and fraction of the genome (in parentheses) and (on the bottom line) the fraction enriched in relation to the rest of the genome and the p value of enrichment in relation to the parent category (in parentheses). DHS enrichment of 4.7× nonsignificantly differed from 5.1× in Figure 3 as a result of additional free parameters.

variants through the full distribution of association statistics (Figure 4, middle panel). However, proximal categories such as UTR and promoter, which were truly depleted, also appeared enriched through tagging of significant coding variants. DHS variants were the least-enriched nonintergenic category, even though they made the single largest contribution to heritability. This was likely due to lower power to detect DHS SNPs as a result of their lower average effect size (relative to that of coding SNPs) and less LD. On the other hand, applying the variance-component strategy to the simulated cohorts correctly recovered the enrichment factors (Figure 4, right panel). These simulations further demonstrate that GWAS p values, although partially informative, can yield false-positive and false-negative enrichment to make functional interpretation difficult, motivating further development of methods that can produce robust estimates of partitioned heritability from summary statistics.

Analysis of PGC2 SP Data

We replicated our functional-enrichment results in an independent cohort of 58,197 samples from PGC2 (Tables S1 and S2). In the PGC2 data, the imputed DHS enrichment was significant at 3.2× (SE = 0.29, $p = 1.4 \times 10^{-13}$), and the intergenic category was significantly depleted at 0.3× (SE = 0.06, depletion $p < 1 \times 10^{-20}$; Table S18). For comparison, the WTCCC2 analysis restricted to SP produced a nonsignificant DHS enrichment of 2.6× (SE = 1.47, p = 0.28) and intergenic h_g^2 of 0.4× (SE = 0.27, depletion p = 0.02; Table S14). The consistency of WTCCC2 and PGC2 estimates indicates that platform artifacts are unlikely to be a major confounder. Moreover, the substantially lower SE in this large cohort demonstrates the effectiveness of our methods at characterizing a single complex trait. As in our previous simulations, p value enrichment did not identify substantial enrichment at DHS variants (Figure S15).

Partitioning h_g^2 within DHSs

We further partitioned DHS enrichment in the WTCCC1 data into functional subcategories to assess significance in relation to all DHSs. We used Segway-chromHMM combined classifications of enhancer regions[40] to partition DHSs (15.7% of the genome) into those that overlapped predicted enhancers (3.2% of the genome) and those that did not (Figure 5A). The enhancer DHSs explained 31.7% (SE = 3.3%) of the total h_g^2, yielding an enrichment of 9.8× versus all SNPs (1.9× versus all DHSs; $p = 5.1 \times 10^{-4}$). We also partitioned DHSs into regions that were called in two or fewer cell types ("specific"; after merging similar tissues) and those that were not (Figure 5B). We observed a significant enrichment for cell-type-specific DHSs (6.1× versus all SNPs; 1.3× versus all DHSs; $p = 3.2 \times 10^{-3}$). The enrichment was not significant when we repeated this analysis for enhancer and nonenhancer DHSs separately. We next split the DHSs into SNPs overlapping and not overlapping the ENCODE database of DGF regions (8.5% of the genome), which are expected to precisely map sites where regulatory factors bind to the genome[50] (Figure 5C). We observed no difference in h_g^2

Table 1. Cell-Type- and Phenotype-Specific DHS Enrichment

		Autoimmune		Nonautoimmune	
Tissue Type	**Cell Type**	**Genotyped**	**Imputed**	**Genotyped**	**Published**
Blood	Primary T helper 1 cells	5.8 (4.2×10^{-6})	10.2 (1.3×10^{-12})	2.1 (3.5×10^{-1})	Maurano et al.[3] (CD)
	leukemia cells	3.5 (6.7×10^{-6})	4.7 (5.3×10^{-10})	1.0 (9.8×10^{-1})	–
	lymphoblastoid cells	3.3 (1.1×10^{-5})	4.9 (5.4×10^{-11})	1.0 (9.4×10^{-1})	Maurano et al.[3] (MS)
	$CD8^+$ primary cells	3.0 (3.0×10^{-4})	5.4 (1.8×10^{-10})	1.0 (9.6×10^{-1})	Trynka et al.[6] (RA)
Fetal kidney	fetal right renal pelvic cells	5.4 (1.4×10^{-4})	8.2 (5.7×10^{-8})	1.5 (7.4×10^{-1})	–
Bone marrow	$CD14^+$ monocytes	4.1 (1.6×10^{-4})	5.7 (2.2×10^{-7})	1.3 (7.6×10^{-1})	Maurano et al.[3] (MS)
Fetal thymus	Fetal thymus cells	2.6 (4.0×10^{-4})	4.5 (3.2×10^{-9})	0.8 (6.6×10^{-1})	–

Fold enrichment of h_g^2 reported for cell-type-specific DHSs observed as significant in genotype data (after adjustment for 83 cell types tested). We measured enrichment in comparison to h_g^2 at DHSs to account for the background DHS enrichment. Results are shown separately from meta-analyses of six autoimmune traits and five nonautoimmune traits. Instances where enrichment was also observed in Trynka et al.[6] or Maurano et al.[3] are indicated.

between these DHSs and other DHSs (1.0×, $p = 0.90$). However, DGF annotations were collected for only a subset of DHS cell types analyzed, and analysis in additional cell types is needed. Lastly, we partitioned the h_g^2 by using an expanded DHS annotation (including regions overlapping coding regions, UTRs, and promoters) into the remaining five major categories (Table S19), which yielded 34.4× enrichment at DHS coding variants versus all SNPs (5.3× versus all DHSs, $p = 1.35 \times 10^{-3}$) and 13.2× enrichment at DHS promoter variants versus all SNPs (2.3× versus all DHSs, $p = 7.90 \times 10^{-3}$). Notably, unlike the non-DHS introns, DHS introns did not show substantial depletion (0.9× versus all DHSs, $p = 0.037$).

To investigate the role of specific cell types, we separately estimated enrichment in h_g^2 for DHSs in each of 83 unique cell types (see Material and Methods). For each trait and cell type, we estimated h_g^2 jointly from three components corresponding to DHSs observed in that cell type, other DHSs not observed in that cell type, and all other SNPs; we assessed enrichment in relation to all DHSs. On the basis of our previous observation of heterogeneity, we performed meta-analyses across the six autoimmune traits (excluding the MHC) and across the five nonautoimmune traits. We observed seven cell types that were significantly enriched in autoimmune traits in genotype data (we conservatively adjusted for 83 tests, although the cell types are highly correlated), and none were significantly enriched in nonautoimmune traits (Table 1). Four of these seven cell types have previously been implicated in autoimmune diseases: Trynka et al.[6] found that GWAS hits for RA were enriched within H3K36me3 peaks from $CD8^+$ primary cells (at $p = 0.0042$), and Maurano et al.[3] found that nominally significant SNPs in a GWAS of CD were enriched within DHS peaks from primary T helper 1 cells and that nominally significant SNPs in a GWAS of MS were enriched in DHS peaks from lymphoblastoid and monocyte $CD14^+$ cells. The remaining three significant cell types were leukemia cells, fetal pelvis cells, and fetal thymus cells (additional nominally significant cell types are listed in Table S20). The enrichment was typically observed in all autoimmune traits individually; CD was the least enriched on average (2.8×), and UC was the most enriched on average (5.1×; Table S21). As before, the signal was stronger and more significant when we included imputed SNPs (Table 1).

On the basis of the hypothesis that most regulatory sites lie at the center of the called DHS peaks, we considered the enrichment after progressively narrowing the DHS annotations. Specifically, we trimmed the ends of each DHS peak (without removing any individual peaks) to a maximum length set such that the resulting overall DHS annotation covered 1%, 5%, or 10% of the physical genome. We then tested these three narrowed annotations in two models: (1) a univariate model in which h_g^2 was inferred from only the narrowed DHS component, thereby including any tagged heritability from other functional categories; and (2) a six-component model in which the full DHS component was replaced with the narrowed DHS component and the remaining DHS SNPs were distributed into the intron and other components. We found the DHS centers to be particularly strongly enriched (Table S22); the 1% annotation explained 19.8% of the total h_g^2 in the multivariate model ($p = 2.6 \times 10^{-6}$) and 61.0% of the total h_g^2 in the univariate model. For comparison, the coding component covering roughly 1% of the genome explained 30.0% of the total h_g^2 in the univariate model. The monotonic increase in h_g^2 from narrowed annotations is further evidence of enrichment at the DHS centers. We caution that this experiment might have been particularly susceptible to bias from causal variants very close to the annotation boundary.

Unbiased Estimates of h_g^2 with Rare and Common Variants

We separately analyzed a cohort of 2,500 SP subjects and 3,875 control subjects who were of homogenous Swedish origin and had been typed on both GWAS and exome chips (see Material and Methods; Tables S1 and S3) to

Table 2. h_g^2 of SP from Exome Chip

Variant Class	h_g^2	Percentage of h_g^2
Separately		
All	0.370 ± 0.040	–
Noncoding	0.317 ± 0.042	–
Coding	0.158 ± 0.034	–
Jointly		
Noncoding	0.291 ± 0.028	79% ± 8%
Coding	0.079 ± 0.034 ($p = 1.2 \times 10^{-2}$)	21% ± 6%
Coding (rare)	0.037 ± 0.029 ($p = 1.0 \times 10^{-1}$)	10% ± 7%
Coding (common)	0.042 ± 0.017 ($p = 7.7 \times 10^{-3}$)	11% ± 4%

Estimates of h_g^2 (adjusted for biases due to LD; see Figure S17 and Table S23) are reported from variance components in the homogenous Swedish subpopulation. The top section shows estimates that include tagging of variants in other classes. The bottom section shows joint estimates accounting for tagged variance due to LD. The p values from a likelihood-ratio test are shown in parentheses.

investigate the possible contribution of rare coding variants to missing heritability,[51] defined as the gap between our genome-wide estimates of h_g^2 and the total narrow-sense heritability. The exome-chip variants were primarily rare and consisted of 18% singletons and 64% nonsingletons with a MAF below 0.01. A concern is that h_g^2 estimates from exome-chip data can be substantially biased as a result of the abundance of rare variants.[21,41,42] To address this, we performed simulations across the full causal-allele frequency spectrum and found that joint estimates from two frequency-stratified[42] components computed from rare (MAF $\leq$ 0.01) and common (MAF $>$ 0.01) SNPs eliminated most of the observed bias. Subsequently adjusting each component for LD completely eliminated bias for normalized effect sizes (Figure S17) and yielded the most accurate estimate for standard effect sizes (Figure S18). We report estimates from joint components with (Table 2) and without (Table S23) LD adjustment.

We partitioned the heritability explained by GWAS-chip and exome-chip data into three separate variance components: noncoding, rare coding (MAF $<$ 0.01), and common coding variants. This partitioned analysis identified a total h_g^2 of 0.079 (SE = 0.034) from all coding variants (Table 2); only the h_g^2 of 0.042 (SE = 0.017) from common coding variants was significantly different from 0 ($p = 7.7 \times 10^{-3}$; rare coding p = 0:10). Moreover, the estimate of DHS enrichment from common SNPs was unaffected by the inclusion of rare coding variants (Table S24), confirming that DHS enrichment was not an artifact of untagged coding variation in this cohort. The h_g^2 from rare variants remained nonsignificant even after we partitioned according to PolyPhen-2 scores,[52] restriction to putative SP-associated genes (see Appendix A), or gene collapsing (Tables S25 and S41–S43). This does not invalidate the use of collapsed-gene burden tests for association and genetic mapping because the individual collapsed gene is still a fundamentally informative unit of association. It does, however, demonstrate that the maximum variance that can be explained by such methods is guaranteed to be substantially lower than that of association with the full model, as has been shown in previous analyses of burden tests.[53] For singleton variants, we can place a 95% upper bound on collapsed h_g^2 at 0.014. We caution that our exome-chip results pertain to rare variants included in the chip design (ascertained from 12,000 samples) but do not extend to extremely rare variants. However, our findings are consistent with a recent analysis of SP exome sequencing data, which identified a significant but modest rare-coding burden (0.4%–0.6% of total variance) in a subset of ~2,500 genes.[54]

Fine Mapping with Functional Priors

Estimates of functional h_g^2 enrichment can guide fine-mapping analysis, where the goal is to identify a minimal set of SNPs that include the underlying causal variant(s).[55] To investigate the potential benefits of fine mapping on the basis of our estimates of functional enrichment, we applied these estimates as priors for fine mapping in four traits (RA, T2D, CAD, and SP) with publicly available imputed summary statistics (Table S26; see Web Resources). We used corresponding estimates of functional enrichment in the WTCCC1 data for RA, T2D, and CAD (while implicitly assuming a best-case scenario in which functional enrichment was accurately estimated for each trait) and used estimates of functional enrichment in PGC2 data for SP. Given that SNPs at genome-wide-significant loci explain only a small proportion of the trait variance, we do not expect partial sample overlap to be a significant confounder. Although fine-mapping analysis ideally involves targeted sequencing or genotyping, Maller et al.[55] observed that the latter had little impact on their fine-mapping analysis in comparison to imputed data, so we expect imputed markers to be a reasonable proxy. Each locus was defined as the union of 1 Mb windows around any SNP with a p value $< 5 \times 10^{-8}$. Association statistics consisting of individual SNP effect sizes and SEs were converted to Bayes factors as described in Pickrell[13] and Wakefield[56] and were multiplied by either a flat prior or the genome-wide functional prior (computed as the estimated h_g^2 per SNP of the SNP category in the corresponding trait). We then computed the credible set for each locus for each scenario by including SNP Bayes factors from highest to lowest until the sum of the Bayes factors in the set was at least 95% of the sum of the Bayes factors at the locus. On average, we found that the six main functional priors reduced the credible set of causal variants by 30% across the four traits (Table 3). The largest reduction of 63% was observed in RA, where the total credible set for five loci (excluding the MHC) was reduced from 69 SNPs to 26. For comparison, including only coding-variant enrichment as a prior reduced the credible sets by 5% on average and had no reduction for RA. We showed by simulation that the credible sets were well calibrated with the correct priors and miscalibrated by less than 10% when the priors were at

Table 3. Credible Sets of Causal SNPs at Known Associated Loci

Phenotype	No. of Loci	Total SNPs	Flat Prior	Coding Prior	Main Functional Priors	Main and Enhancer Priors
RA	5	8,393	69	69	26	26
T2D	13	24,799	101	90	84	83
CAD	16	27,685	112	112	90	86
CAD (metabo-chip)	34	7,498	325	325	264	260
PGC2	146	582,401	5,696	5,660	4,756	not available

For each trait, genome-wide-significant loci from meta-analysis association statistics were reduced to 95% credible sets with and without functional priors. The right-most four columns describe the number of SNPs in the credible set obtained from each prior type. "Flat prior" corresponds to standard analysis with no functional information. "Coding prior" uses only enrichment at coding variants. "Main functional priors" include all six priors from the main functional analysis. "Main and enhancer priors" include all six main priors and the enhancer-DHS prior.

the extremes of the meta-analysis estimates (Table S27), demonstrating that this functional fine-mapping strategy might become robust and effective as individual trait sample sizes reach the current meta-analysis sample size. However, we caution that our estimates of functional enrichment for individual traits, except SP, are not tight enough for this strategy to be actionable at the current time.

Discussion

The importance of regulatory and cell-type-specific variation in common disease has previously been recognized,[3–10] but in contrast to previous work, we provide a quantification of this contribution to disease heritability. We have demonstrated by extensive simulations that our variance-component strategy yields robust estimates that account for LD between categories and complex-disease architecture. Across 11 traits, we found that regulatory regions marked by DHSs explained an average of 79% of imputed h_g^2 and 38% of genotyped h_g^2. We replicated our results in a large SP cohort, yielding a single-trait estimate of 3.2× (SE = 0.29, $p = 1.4 \times 10^{-13}$) from imputed SNPs, and found that the contribution from rare, exome-chip variants was nonsignificant and did not affect the enrichment.

Given that GWASs primarily identify noncoding variants, many hypotheses have been developed to explain the architecture of complex traits, including noncoding RNA, DNA methylation, alternative splicing, and unannotated transcripts.[14,57] Several previous studies have demonstrated an excess of significant GWAS associations in regulatory categories.[5,6,11,58] In particular, Ernst et al.[59] observed 2× enrichment in cell-type-relevant enhancers, Schaub et al.[8] identified 1.12× enrichment at DHSs, and Maurano et al.[3] identified 1.4×–1.8× enrichment at DHSs (relative to noncoding SNPs) and enrichment at cell-type-relevant DHSs. In our analyzed cohorts, known variants were1.7× enriched with DHSs, but there was less enrichment at variants identified only in these cohorts. In contrast, our findings constrain most of h_g^2 to the 16% of SNPs that lie in the DHS marks tested (or to SNPs that lie very close to DHSs; see below), particularly in those that overlap enhancers, and suggest that the other proposed mechanisms are unlikely to make substantial independent contributions. A deeper analysis of DHSs narrowed to cover 1% of the genome still explained 20% of h_g^2 directly (and 61% in total), potentially motivating a DHS-targeted genotyping chip analogous to the exome chip.[60] More generally, our approach provides a means of assessing biological hypotheses of contributions to disease heritability.

Unlike previous methods, our approach infers disease-relevant biological function from all SNPs simultaneously instead of one GWAS hit at a time. Over multiple simulated disease architectures, we show that variance-component methods are more accurate in partitioning heritability than summary-statistic-based approaches, such as p value enrichment, despite the appeal of analyses of summary statistics in many contexts.[61–64] For completeness, we also considered two additional methods, stratified Q-Q plots[12] and Bayesian hierarchical modeling (fgwas),[13] which assess functional enrichment but are primarily focused on strong associations and improving mapping power. These methods did not produce consistent estimates of h_g^2 enrichment either in simulations or in real data (Figure S19–S29), although we note that they have different objectives. In addition to having implications for mapping power,[12,13,65–68] functional enrichment has direct implications for fine mapping[55,69,70] and risk prediction. Enrichments at the level we observed could substantially reduce the set of potential causal variants in the four traits we tested by downweighing SNPs in low-heritability categories. On the other hand, the improvement in polygenic risk prediction was limited because of pervasive LD across categories (Table S28).

Several limitations of our approach remain as avenues for future work. The variance-component method might still be subject to subtle biases[21,41,42] under disease architectures or annotations with complex LD structure, although our analyses indicate that it is generally less biased than published methods. In particular, we found that imputed data might lead to an overestimate of category enrichment from causal variation very close to that category. For computational reasons, we did not make use of the mixture of the normal-effect-size approach, which has been shown to increase precision.[24] The method also requires

individual-level genotype data and is computationally infeasible for extremely large cohorts or a very large number of components, motivating further work on methods that analyze summary statistics. A limitation of assessing enrichment from GWAS platforms is that we cannot account for untagged causal variation, which represents roughly half of total narrow-sense heritability.[71] Although we have shown that rare coding variants are unlikely to alter the DHS enrichment, the missing heritability could lie in other categories. The precision of inferred enrichment is also limited by the underlying annotations and variants. It is possible that certain biological features could be subject to systematically poorer variant calling or imputation and exhibit decreased h_g^2 as a result of artifacts,[72] although we did not observe substantial differences in the categories we analyzed. Because of the data available, our meta-analysis estimates were weighted toward autoimmune traits both in the number of individual studies and in total sample size; estimates of DHS enrichment were higher in autoimmune than in nonautoimmune traits, which could be partly due to the abundance of hematopoietic cell types in available DHS annotations. Except for SP, for which many samples are available, we could not provide precise estimates for single traits. However, we have shown by simulation that the individual estimates and errors were well calibrated, justifying meta-analysis of estimates that are not constrained to the plausible 0–1 range (an established strategy[49]). Further partitioning of DHSs can yield additional enrichment, and it is likely that other functional categories—including additional chromatin marks, histone modifications, formaldehyde-assisted isolation of regulatory elements, transcription factor binding sites,[73] gene expression,[58,74,75] and measures of conservation[7]—will be highly informative.

Appendix A

LD

We further interrogated the role of LD and violations of model assumptions in the variance-component estimate. We considered two contrived annotations constructed from either the 16% of SNPs with the most LD partners or the 16% of SNPs with the fewest LD partners to mimic a high or low LD category, respectively, approximately equal in SNP number to the DHS category. Testing the uniformly drawn MAF-independent architecture, we again observed no enrichment for either the high-LD (1.02×, SE = 0.01) or the low-LD (1.02×, SE = 0.03) annotations over 1,000 trials. Finally, we considered a disease architecture in which causal variants were strongly enriched at the centers of DHSs such that variants in the middle 7% of the DHS (1% of the genome) explained 25% of the h_g^2 and the remaining DHS variants explained 75% of the h_g^2. We observed a slight deflation of the DHS estimate, but no significant false enrichment, at the neighboring categories (Figure S20).

Jackknife Estimates of SEs

The analytical SE used for significance testing was accurate in our simulations (Table S29) and has previously been shown to be robust in real data[21,27] but can be biased when the number of causal variants is very small.[41] We assessed this directly with a weighted-block jackknife estimate[76] of the enrichment in the real traits by dropping each chromosome in turn, constructing new GRMs, and recomputing the percentage of h_g^2 for each functional category (and the corresponding enrichment). The jackknife estimate of the enrichment and its variance was then computed as described in Busing et al.[76] Although there is a demonstrable relationship between chromosome length and h_g^2, we do not expect to observe such a relationship with respect to the percentage of h_g^2 because of the normalization. However, this estimate of the variance does capture true biological variation in enrichment across chromosomes and is therefore conservative. Although we observed little difference between the jackknife and standard estimates in genotyped data (Table S30), the jackknife estimate of the imputed percentage of h_g^2 (71%, SE = 7.7%; Table S31) was indeed more conservative than the analytical estimate (79%, SE = 6.6%), but the enrichment was still highly significant ($p = 5.5 \times 10^{-13}$), and the overall results were not substantially affected. Because the jackknife makes no assumptions about the underlying distribution of enrichment, this consistency with the analytical estimate supports the use of REML SEs for case-control data (see also simulations below).

Ancestry

We found little population structure in all of the traits except for MS and SP (Figure S1), which have been previously reported as structured. For the MS cohort, we have shown previously[21] that rigorous ancestry matching did not substantially change the total or partitioned h_g^2. For the SP cohort, we relied on the consistently replicated enrichment across the PGC2 and Swedish SP cohorts, which have been rigorously quality controlled for the avoidance of population stratification. Recently, Janss et al.[77] demonstrated that h_g^2 can vary significantly when principal components are also included as fixed effects as a function of the number of included eigenvectors. To assess the presence of this bias in our Swedish SP data, we recomputed the joint variance-component estimates of $\widehat{h}_g^2$ while including an increasing number of eigenvectors as fixed effects. We observed no significant fluctuation of $\widehat{h}_g^2$ such that the estimates over 1–20 eigenvector covariates had a SD of 0.002, suggesting a tight estimate unbiased by the fixed effects.

Case-Control Ascertainment

Recent work[37,78,79] has shown that liability-scale estimates of h_g^2 from REML can be biased downward in dichotomous traits with strong case-control ascertainment. Golan and Rosset[78] and Hayeck et al.[79] propose an alternative estimator based on Haseman-Elston (H-E) regression[80] and show that it eliminates bias. In brief, this approach

regresses the product of normalized phenotypes on the genetic covariance (off-diagonal GRM entries) for all unique pairs of samples; the resulting slope is used as an estimate of the observed-scale h_g^2 and is converted to the liability scale. This method can be extended naturally to multiple components, where the product of phenotypes is regressed onto GRM entries from each analyzed component in a multiple linear regression. Here, we compared the method and transformation of Golan and Rosset[78] to the REML estimator described in the main text. We also evaluated the impact of incorporating principal components as fixed effects to account for genetic ancestry. This is particularly important for the SP and MS cohorts (see below), which were ascertained in a way that induces correlations between ancestry and phenotype. All analyses were performed with the same set of GRMs computed from 1000 Genomes imputed data, and the H-E regression (and H-E regression with fixed effects) was implemented as described in Golan and Rosset.[78] In all instances, we used analytical error-covariance estimates and rescaled them with the delta method to compute SEs. (We note that the SE for H-E regression makes strongly violated assumptions about independence, and they are therefore only presented for completeness). We observed little difference between variance-component methods and H-E regression methods, and H-E regression yielded an average estimate 1.05× greater than that of REML and an overall $r^2 = 0.95$ between the two methods (across 11 traits; Table S32). The relative performance was similar when we considered only the percentage of h_g^2 from the DHS component (Table S33) such that H-E regression yielded average estimates 1.04× higher than those of REML and an overall $r^2 = 0.94$. When principal components where included as fixed effects, meta-analysis across traits within each method did not yield significant differences (Table S34); H-E regression identified DHS enrichment of 5.8× (SE = 0.45), and REML identified DHS enrichment of 5.1× (SE = 0.42). When we did not include principal components as fixed effects, we observed a large difference between variance components and H-E regression in the SP and MS cohorts, where liability-scale H-E regression estimates of liability-scale h_g^2 were 10.00 and 2.91, respectively (Table S32), outside the plausible 0–1 bound and vastly larger than REML estimates without fixed effects. This suggests that H-E-regression-based estimates might be particularly sensitive to the confounding effects of ancestry.

Lastly, we repeated our null simulations by using the merged WTCCC2 cohort of ~33,000 samples, allowing us to simulate a case-control ascertainment (327 case and 654 control subjects) at a prevalence of 0.01 (see Table S35 for simulation details). When we generated ~1,000 samples on chromosome 1 only, this simulated cohort had an effective SNP-sample ratio (the key quantity driving the effects of case-control ascertainment[37]) corresponding to that of ~10,000 samples genome-wide. We tested a "polygenic" scenario where causal variants were sampled uniformly, as well as a "high-effect" scenario where DHS variants had 10× the effect of other SNPs, and found no significant deviation from the null estimate (Table S35) or the analytical SE (Table S36). Although ascertainment has previously been shown to induce correlation between causal variants, our simulations indicate that this does not bias estimates of enrichment for the prevalence and sample size simulated here.

Detailed Analyses of Rare-Variant h_g^2

Having identified no significant rare-variant h_g^2 at any coding regions, we were interested in quantifying this phenomenon at the set of loci known to be associated with SP. To do so, we constructed six variance components only from SNPs at the 22 loci identified by the PGC1 in a large meta-analysis[48] and estimated h_g^2 jointly with a component for the remaining noncoding variants genome-wide (to account for tagging). As expected, we found the union of all noncoding GWAS variants at these loci to harbor significant heritability of 0.018 (SE = 0.004) (Table S37). However, we did not see any significant heritability from the coding variants at these classes when they were modeled jointly with the other component. This is consistent with our genome-wide finding that common noncoding variants explained a substantial fraction of trait heritability and tagged nearly half of the common coding variation. We also partitioned h_g^2 at the set of 1,796 "composite" genes reported by Purcell et al.[54] to exhibit enrichment of rare disruptive mutations, modeled jointly with exome-chip variants in the remaining genes and noncoding GWAS-chip variants as separate components. However, no significant h_g^2 was observed at either the entire set of composite variants ($h_g^2 = 0.014$, SE = 0.012) or the rare composite variants ($h_g^2 = 0.008$, SE = 0.012).

We observed a significant enrichment in h_g^2 at 4,919 (nonsingleton) loss-of-function variants, which collectively accounted for 6.0% of (nonsingleton) exonic SNPs but explained 24.3% of the exonic h_g^2 (permuted p = 0.02 after MAF matching). We saw no significant enrichment of h_g^2 at coding sites that were predicted to be functionally important by PolyPhen-2.[52] Comparing likelihoods between the model where variants were split into (1) probably damaging and damaging, (2) benign and other, and (3) noncoding components and the model with only (1) coding and (2) noncoding components yielded no significant difference by a 1-degree-of-freedom likelihood-ratio test (p = 0.13).

Supplemental Data

Supplemental Data include 29 figures and 42 tables and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2014.10.004.

Consortia

The members of the Schizophrenia Working Group of the Psychiatric Genomics Consortium are Stephan Ripke, Benjamin M.

Neale, Aiden Corvin, James T.R. Walters, Kai-How Farh, Peter A. Holmans, Phil Lee, Brendan Bulik-Sullivan, David A. Collier, Hailiang Huang, Tune H. Pers, Ingrid Agartz, Esben Agerbo, Margot Albus, Madeline Alexander, Farooq Amin, Silviu A. Bacanu, Martin Begemann, Richard A. Belliveau, Jr., Judit Bene, Sarah E. Bergen, Elizabeth Bevilacqua, Tim B. Bigdeli, Donald W. Black, Anders D. Børglum, Richard Bruggeman, Nancy G. Buccola, Randy L. Buckner, William Byerley, Wiepke Cahn, Guiqing Cai, Dominique Campion, Rita M. Cantor, Vaughan J. Carr, Noa Carrera, Stanley V. Catts, Kimberly D. Chambert, Raymond C.K. Chan, Ronald Y.L. Chen, Eric Y.H. Chen, Wei Cheng, Eric F.C. Cheung, Siow Ann Chong, C. Robert Cloninger, David Cohen, Nadine Cohen, Paul Cormican, Nick Craddock, James J. Crowley, David Curtis, Michael Davidson, Kenneth L. Davis, Franziska Degenhardt, Jurgen Del Favero, Lynn E. DeLisi, Ditte Demontis, Dimitris Dikeos, Timothy Dinan, Srdjan Djurovic, Gary Donohoe, Elodie Drapeau, Jubao Duan, Frank Dudbridge, Naser Durmishi, Peter Eichhammer, Johan Eriksson, Valentina Escott-Price, Laurent Essioux, Ayman H. Fanous, Martilias S. Farrell, Josef Frank, Lude Franke, Robert Freedman, Nelson B. Freimer, Marion Friedl, Joseph I. Friedman, Menachem Fromer, Giulio Genovese, Lyudmila Georgieva, Elliot S. Gershon, Ina Giegling, Paola Giusti-Rodrguez, Stephanie Godard, Jacqueline I. Goldstein, Vera Golimbet, Srihari Gopal, Jacob Gratten, Jakob Grove, Lieuwe de Haan, Christian Hammer, Marian L. Hamshere, Mark Hansen, Thomas Hansen, Vahram Haroutunian, Annette M. Hartmann, Frans A. Henskens, Stefan Herms, Joel N. Hirschhorn, Per Hoffmann, Andrea Hofman, Mads V. Hollegaard, David M. Hougaard, Masashi Ikeda, Inge Joa, Antonio Julià, René S. Kahn, Luba Kalaydjieva, Sena Karachanak-Yankova, Juha Karjalainen, David Kavanagh, Matthew C. Keller, Brian J. Kelly, James L. Kennedy, Andrey Khrunin, Yunjung Kim, Janis Klovins, James A. Knowles, Bettina Konte, Vaidutis Kucinskas, Zita Ausrele Kucinskiene, Hana Kuzelova-Ptackova, Anna K. Kähler, Claudine Laurent, Jimmy Lee Chee Keong, S. Hong Lee, Sophie E. Legge, Bernard Lerer, Miaoxin Li, Tao Li, Kung-Yee Liang, Jeffrey Lieberman, Svetlana Limborska, Carmel M. Loughland, Jan Lubinski, Jouko Lnnqvist, Milan Macek, Jr., Patrik K.E. Magnusson, Brion S. Maher, Wolfgang Maier, Jacques Mallet, Sara Marsal, Manuel Mattheisen, Morten Mattingsdal, Robert W. McCarley, Colm McDonald, Andrew M. McIntosh, Sandra Meier, Carin J. Meijer, Bela Melegh, Ingrid Melle, Raquelle I. Mesholam-Gately, Andres Metspalu, Patricia T. Michie, Lili Milani, Vihra Milanova, Younes Mokrab, Derek W. Morris, Ole Mors, Preben B. Mortensen, Kieran C. Murphy, Robin M. Murray, Inez Myin-Germeys, Bertram Mller-Myhsok, Mari Nelis, Igor Nenadic, Deborah A. Nertney, Gerald Nestadt, Kristin K. Nicodemus, Liene Nikitina-Zake, Laura Nisenbaum, Annelie Nordin, Eadbhard O'Callaghan, Colm O'Dushlaine, F. Anthony O'Neill, Sang-Yun Oh, Ann Olincy, Line Olsen, Jim Van Os, Psychosis Endophenotypes International Consortium, Christos Pantelis, George N. Papadimitriou, Sergi Papiol, Elena Parkhomenko, Michele T. Pato, Tiina Paunio, Milica Pejovic-Milovancevic, Diana O. Perkins, Olli Pietilinen, Jonathan Pimm, Andrew J. Pocklington, John Powell, Alkes Price, Ann E. Pulver, Shaun M. Purcell, Digby Quested, Henrik B. Rasmussen, Abraham Reichenberg, Mark A. Reimers, Alexander L. Richards, Joshua L. Roffman, Panos Roussos, Douglas M. Ruderfer, Veikko Salomaa, Alan R. Sanders, Ulrich Schall, Christian R. Schubert, Thomas G. Schulze, Sibylle G. Schwab, Edward M. Scolnick, Rodney J. Scott, Larry J. Seidman, Jianxin Shi, Engilbert Sigurdsson, Teimuraz Silagadze, Jeremy M. Silverman, Kang Sim, Petr Slominsky, Jordan W. Smoller, Hon-Cheong So, Chris C.A. Spencer, Eli A. Stahl, Hreinn Stefansson, Stacy Steinberg, Elisabeth Stogmann, Richard E. Straub, Eric Strengman, Jana Strohmaier, T. Scott Stroup, Mythily Subramaniam, Jaana Suvisaari, Dragan M. Svrakic, Jin P. Szatkiewicz, Erik Sderman, Srinivas Thirumalai, Draga Toncheva, Paul A. Tooney, Sarah Tosato, Juha Veijola, John Waddington, Dermot Walsh, Dai Wang, Qiang Wang, Bradley T. Webb, Mark Weiser, Dieter B. Wildenauer, Nigel M. Williams, Stephanie Williams, Stephanie H. Witt, Aaron R. Wolen, Emily H.M. Wong, Brandon K. Wormley, Jing Qin Wu, Hualin Simon Xi, Clement C. Zai, Xuebin Zheng, Fritz Zimprich, Naomi R. Wray, Kari Stefansson, Peter M. Visscher, Wellcome Trust Case Control Consortium, Rolf Adolfsson, Ole A. Andreassen, Douglas H.R. Blackwood, Elvira Bramon, Joseph D. Buxbaum, Anders D. Brglum, Sven Cichon, Ariel Darvasi, Enrico Domenici, Hannelore Ehrenreich, Tõnu Esko, Pablo V. Gejman, Michael Gill, Hugh Gurling, Christina M. Hultman, Nakao Iwata, Assen V. Jablensky, Erik G. Jönsson, Kenneth S. Kendler, George Kirov, Jo Knight, Todd Lencz, Douglas F. Levinson, Qingqin S. Li, Jianjun Liu, Anil K. Malhotra, Steven A. McCarroll, Andrew McQuillin, Jennifer L. Moran, Preben B. Mortensen, Bryan J. Mowry, Markus M. Nthen, Roel A. Ophoff, Michael J. Owen, Aarno Palotie, Carlos N. Pato, Tracey L. Petryshen, Danielle Posthuma, Marcella Rietschel, Brien P. Riley, Dan Rujescu, Pak C. Sham, Pamela Sklar, David St. Clair, Daniel R. Weinberger, Jens R. Wendland, Thomas Werge, Mark J. Daly, Patrick F. Sullivan, and Michael C. O'Donovan.

The members of the SWE-SCZ Consortium are Stephan Ripke, Colm O'Dushlaine, Kimberly Chambert, Jennifer L. Moran, Anna K. Kähler, Susanne Akterin, Sarah Bergen, Patrik K.E. Magnusson, Benjamin M. Neale, Douglas Ruderfer, Edward Scolnick, Shaun Purcell, Steve McCarroll, Pamela Sklar, Christina M Hultman, and Patrick F. Sullivan.

Acknowledgments

This study made use of data generated by the Wellcome Trust Case Control Consortium (WTCCC) and the Wellcome Trust Sanger Institute. A full list of the investigators who contributed to the generation of the WTCCC data is available at www.wtccc.org.uk. Funding for the WTCCC project was provided by the Wellcome Trust under award 076113. We thank Manolis Kellis, Abhishek Sarkar, Joe Pickrell, X. Shirley Liu, Nick Patterson, Sara Lindstrom, Peter Kraft, and Shamil Sunyaev for helpful discussions and Amy Williams for assistance with HAPI-UR. This research was funded by NIH grants R01 MH101244, R03 HG006731, and 1U01HG0070033, the Doris Duke Clinical Scientist Development Award, and NIH fellowship F32 GM106584. G.T. was supported by the Rubicon grant from the Netherlands Organization for Scientific Research. H.F. was supported by the Fannie and John Hertz Foundation. We also acknowledge grant funding from the Australian Research Council (DE130100614 and FT0991360) and the National Health and Medical Research Council (613602 and 1050218).

Received: July 28, 2014
Accepted: October 2, 2014
Published: November 6, 2014

Web Resources

The URLs for data presented herein are as follows:

1000 Genomes Phase 1 reference panels, https://mathgen.stats.ox.ac.uk/impute/impute_v2.html#reference

CARDIoGRAM CAD summary statistics, http://www.cardiogramplusc4d.org/downloads/

DIAGRAM T2D summary statistics, http://diagram-consortium.org/downloads.html

DNaseI Digital Genomic Footprinting (DGF) annotations, http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeUwDgf/

Exome Chip Design, http://genome.sph.umich.edu/wiki/Exome_Chip_Design

fgwas, https://github.com/joepickrell/fgwas

Functional annotations, http://www.hsph.harvard.edu/alkes-price/software/

Genome-wide Complex Trait Analysis (GCTA), http://www.complextraitgenomics.com/software/gcta/

HAPI-UR, http://genetics.med.harvard.edu/reich/Reich_Lab/Software.html

IMPUTE2, https://mathgen.stats.ox.ac.uk/impute/impute_v2.html

NHGRI GWAS catalog, http://www.genome.gov/gwastudies/

Oxford recombination map, http://hapmap.ncbi.nlm.nih.gov/downloads/recombination/

Psychiatric Genomic Consortium, Sweden+SCZ1 schizophrenia summary statistics, http://www.med.unc.edu/pgc/downloads

RA summary statistics, http://www.broadinstitute.org/ftp/pub/rheumatoid_arthritis/Stahl_etal_2010NG/

Segway-chromHMM combined enhancer annotations, ftp://ftp.ebi.ac.uk/pub/databases/ensembl/encode/integration_data_jan2011/byDataType/segmentations/jan2011

References

1. Bernstein, B.E., Stamatoyannopoulos, J.A., Costello, J.F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M.A., Beaudet, A.L., Ecker, J.R., et al. (2010). The nih roadmap epigenomics mapping consortium. Nat. Biotechnol. *28*, 1045–1048.
2. Bernstein, B.E., Birney, E., Dunham, I., Green, E.D., Gunter, C., and Snyder, M.; ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.
3. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190–1195.
4. Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75–82.
5. Degner, J.F., Pai, A.A., Pique-Regi, R., Veyrieras, J.-B., Gaffney, D.J., Pickrell, J.K., De Leon, S., Michelini, K., Lewellen, N., Crawford, G.E., et al. (2012). DNase I sensitivity QTLs are a major determinant of human expression variation. Nature *482*, 390–394.
6. Trynka, G., Sandor, C., Han, B., Xu, H., Stranger, B.E., Liu, X.S., and Raychaudhuri, S. (2013). Chromatin marks identify critical cell types for fine mapping complex trait variants. Nat. Genet. *45*, 124–130.
7. Ward, L.D., and Kellis, M. (2012). Evidence of abundant purifying selection in humans for recently acquired regulatory functions. Science *337*, 1675–1678.
8. Schaub, M.A., Boyle, A.P., Kundaje, A., Batzoglou, S., and Snyder, M. (2012). Linking disease associations with regulatory information in the human genome. Genome Res. *22*, 1748–1759.
9. Stergachis, A.B., Haugen, E., Shafer, A., Fu, W., Vernot, B., Reynolds, A., Raubitschek, A., Ziegler, S., LeProust, E.M., Akey, J.M., and Stamatoyannopoulos, J.A. (2013). Exonic transcription factor binding directs codon choice and affects protein evolution. Science *342*, 1367–1372.
10. McVicker, G., van de Geijn, B., Degner, J.F., Cain, C.E., Banovich, N.E., Raj, A., Lewellen, N., Myrthil, M., Gilad, Y., and Pritchard, J.K. (2013). Identification of genetic variants that affect histone modifications in human cells. Science *342*, 747–749.
11. Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S., and Manolio, T.A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc. Natl. Acad. Sci. USA *106*, 9362–9367.
12. Schork, A.J., Thompson, W.K., Pham, P., Torkamani, A., Roddey, J.C., Sullivan, P.F., Kelsoe, J.R., O'Donovan, M.C., Furberg, H., Schork, N.J., et al.; Tobacco and Genetics Consortium; Bipolar Disorder Psychiatric Genomics Consortium; Schizophrenia Psychiatric Genomics Consortium (2013). All SNPs are not created equal: genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. PLoS Genet. *9*, e1003449.
13. Pickrell, J.K. (2014). Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. Am. J. Hum. Genet. *94*, 559–573.
14. Mudge, J.M., Frankish, A., and Harrow, J. (2013). Functional transcriptomics in the post-ENCODE era. Genome Res. *23*, 1961–1973.
15. Kellis, M., Wold, B., Snyder, M.P., Bernstein, B.E., Kundaje, A., Marinov, G.K., Ward, L.D., Birney, E., Crawford, G.E., Dekker, J., et al. (2014). Defining functional DNA elements in the human genome. Proc. Natl. Acad. Sci. USA *111*, 6131–6138.
16. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. *42*, 565–569.
17. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. *88*, 76–82.
18. Yang, J., Manolio, T.A., Pasquale, L.R., Boerwinkle, E., Caporaso, N., Cunningham, J.M., de Andrade, M., Feenstra, B., Feingold, E., Hayes, M.G., et al. (2011). Genome partitioning of genetic variation for complex traits using common SNPs. Nat. Genet. *43*, 519–525.
19. Lee, S.H., Wray, N.R., Goddard, M.E., and Visscher, P.M. (2011). Estimating missing heritability for disease from genome-wide association studies. Am. J. Hum. Genet. *88*, 294–305.
20. Lee, S.H., DeCandia, T.R., Ripke, S., Yang, J., Sullivan, P.F., Goddard, M.E., Keller, M.C., Visscher, P.M., and Wray, N.R.; Schizophrenia Psychiatric Genome-Wide Association Study Consortium (PGC-SCZ); International Schizophrenia Consortium (ISC); Molecular Genetics of Schizophrenia Collaboration (MGS) (2012). Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. Nat. Genet. *44*, 247–250.
21. Gusev, A., Bhatia, G., Zaitlen, N., Vilhjalmsson, B.J., Diogo, D., Stahl, E.A., Gregersen, P.K., Worthington, J., Klareskog, L.,

Raychaudhuri, S., et al. (2013). Quantifying missing heritability at known GWAS loci. PLoS Genet. *9*, e1003993.
22. Lee, S.H., Ripke, S., Neale, B.M., Faraone, S.V., Purcell, S.M., Perlis, R.H., Mowry, B.J., Thapar, A., Goddard, M.E., Witte, J.S., et al.; Cross-Disorder Group of the Psychiatric Genomics Consortium; International Inflammatory Bowel Disease Genetics Consortium (IIBDGC) (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat. Genet. *45*, 984–994.
23. Falconer, D.S. (1965). The inheritance of liability to certain diseases, estimated from the incidence among relatives. Ann. Hum. Genet. *29*, 51–76.
24. Zhou, X., Carbonetto, P., and Stephens, M. (2013). Polygenic modeling with bayesian sparse linear mixed models. PLoS Genet. *9*, e1003264.
25. Fischer, T.M., Gilmour, A.R., and van der Werf, J.H. (2004). Computing approximate standard errors for genetic parameters derived from random regression models fitted by average information REML. Genet. Sel. Evol. *36*, 363–369.
26. Lynch, M., and Walsh, B. (1998). Genetics and Analysis of Quantitative Traits (Sunderland: Sinauer Associates).
27. Visscher, P.M., Hemani, G., Vinkhuyzen, A.A.E., Chen, G.-B., Lee, S.H., Wray, N.R., Goddard, M.E., and Yang, J. (2014). Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. PLoS Genet. *10*, e1004269.
28. Nakaoka, H., and Inoue, I. (2009). Meta-analysis of genetic association studies: methodologies, between-study heterogeneity and winner's curse. J. Hum. Genet. *54*, 615–623.
29. Clayton, D.G., Walker, N.M., Smyth, D.J., Pask, R., Cooper, J.D., Maier, L.M., Smink, L.J., Lam, A.C., Ovington, N.R., Stevens, H.E., et al. (2005). Population structure, differential bias and genomic control in a large-scale, case-control association study. Nat. Genet. *37*, 1243–1246.
30. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. *38*, 904–909.
31. Williams, A.L., Patterson, N., Glessner, J., Hakonarson, H., and Reich, D. (2012). Phasing of many thousands of genotyped samples. Am. J. Hum. Genet. *91*, 238–251.
32. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., and Abecasis, G.R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat. Genet. *44*, 955–959.
33. Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A.; 1000 Genomes Project Consortium (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56–65.
34. Myers, S., Bottolo, L., Freeman, C., McVean, G., and Donnelly, P. (2005). A fine-scale map of recombination rates and hotspots across the human genome. Science *310*, 321–324.
35. Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421–427.
36. Bulik-Sullivan, B., Loh, P.-R., Finucane, H., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M.; Schizophrenia Working Group Psychiatric Genomics Consortium (2014). LD score regression distinguishes confounding from polygenicity in genome-wide association studies. bioRxiv http://dx.doi.org/10.1101/002931.
37. Yang, J., Zaitlen, N.A., Goddard, M.E., Visscher, P.M., and Price, A.L. (2014). Advantages and pitfalls in the application of mixed-model association methods. Nat. Genet. *46*, 100–106.
38. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.
39. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.
40. Hoffman, M.M., Ernst, J., Wilder, S.P., Kundaje, A., Harris, R.S., Libbrecht, M., Giardine, B., Ellenbogen, P.M., Bilmes, J.A., Birney, E., et al. (2013). Integrative annotation of chromatin elements from ENCODE data. Nucleic Acids Res. *41*, 827–841.
41. Speed, D., Hemani, G., Johnson, M.R., and Balding, D.J. (2012). Improved heritability estimation from genome-wide SNPs. Am. J. Hum. Genet. *91*, 1011–1021.
42. Lee, S.H., Yang, J., Chen, G.-B., Ripke, S., Stahl, E.A., Hultman, C.M., Sklar, P., Visscher, P.M., Sullivan, P.F., Goddard, M.E., and Wray, N.R. (2013). Estimation of SNP heritability from dense genotype data. Am. J. Hum. Genet. *93*, 1151–1155.
43. Marchini, J., and Howie, B. (2010). Genotype imputation for genome-wide association studies. Nat. Rev. Genet. *11*, 499–511.
44. Wellcome Trust Case Control Consortium (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature *447*, 661–678.
45. Sawcer, S., Hellenthal, G., Pirinen, M., Spencer, C.C., Patsopoulos, N.A., Moutsianas, L., Dilthey, A., Su, Z., Freeman, C., Hunt, S.E., et al.; International Multiple Sclerosis Genetics Consortium; Wellcome Trust Case Control Consortium 2 (2011). Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature *476*, 214–219.
46. Barrett, J.C., Lee, J.C., Lees, C.W., Prescott, N.J., Anderson, C.A., Phillips, A., Wesley, E., Parnell, K., Zhang, H., Drummond, H., et al.; UK IBD Genetics Consortium; Wellcome Trust Case Control Consortium 2 (2009). Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region. Nat. Genet. *41*, 1330–1334.
47. Karczewski, K.J., Dudley, J.T., Kukurba, K.R., Chen, R., Butte, A.J., Montgomery, S.B., and Snyder, M. (2013). Systematic functional regulatory assessment of disease-associated variants. Proc. Natl. Acad. Sci. USA *110*, 9607–9612.
48. Ripke, S., O'Dushlaine, C., Chambert, K., Moran, J.L., Kähler, A.K., Akterin, S., Bergen, S.E., Collins, A.L., Crowley, J.J., Fromer, M., et al.; Multicenter Genetic Studies of Schizophrenia Consortium; Psychosis Endophenotypes International Consortium; Wellcome Trust Case Control Consortium 2 (2013). Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nat. Genet. *45*, 1150–1159.
49. Wright, F.A., Sullivan, P.F., Brooks, A.I., Zou, F., Sun, W., Xia, K., Madar, V., Jansen, R., Chung, W., Zhou, Y.-H., et al. (2014). Heritability and genomics of gene expression in peripheral blood. Nat. Genet. *46*, 430–437.
50. Neph, S., Vierstra, J., Stergachis, A.B., Reynolds, A.P., Haugen, E., Vernot, B., Thurman, R.E., John, S., Sandstrom, R., Johnson, A.K., et al. (2012). An expansive human regulatory lexicon encoded in transcription factor footprints. Nature *489*, 83–90.

51. Gibson, G. (2011). Rare and common variants: twenty arguments. Nat. Rev. Genet. *13*, 135–145.
52. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat. Methods *7*, 248–249.
53. Liu, D.J., and Leal, S.M. (2012). Estimating genetic effects and quantifying missing heritability explained by identified rare-variant associations. Am. J. Hum. Genet. *91*, 585–596.
54. Purcell, S.M., Moran, J.L., Fromer, M., Ruderfer, D., Solovieff, N., Roussos, P., O'Dushlaine, C., Chambert, K., Bergen, S.E., Kähler, A., et al. (2014). A polygenic burden of rare disruptive mutations in schizophrenia. Nature *506*, 185–190.
55. Maller, J.B., McVean, G., Byrnes, J., Vukcevic, D., Palin, K., Su, Z., Howson, J.M.M., Auton, A., Myers, S., Morris, A., et al.; Wellcome Trust Case Control Consortium (2012). Bayesian refinement of association signals for 14 loci in 3 common diseases. Nat. Genet. *44*, 1294–1301.
56. Wakefield, J. (2009). Bayes factors for genome-wide association studies: comparison with P-values. Genet. Epidemiol. *33*, 79–86.
57. Edwards, S.L., Beesley, J., French, J.D., and Dunning, A.M. (2013). Beyond GWASs: illuminating the dark road from association to function. Am. J. Hum. Genet. *93*, 779–797.
58. Nicolae, D.L., Gamazon, E., Zhang, W., Duan, S., Dolan, M.E., and Cox, N.J. (2010). Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet. *6*, e1000888.
59. Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature *473*, 43–49.
60. Huyghe, J.R., Jackson, A.U., Fogarty, M.P., Buchkovich, M.L., Stančáková, A., Stringham, H.M., Sim, X., Yang, L., Fuchsberger, C., Cederberg, H., et al. (2013). Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. Nat. Genet. *45*, 197–201.
61. Lee, S., Teslovich, T.M., Boehnke, M., and Lin, X. (2013). General framework for meta-analysis of rare variants in sequencing association studies. Am. J. Hum. Genet. *93*, 42–53.
62. Lee, D., Bigdeli, T.B., Riley, B.P., Fanous, A.H., and Bacanu, S.-A. (2013). DIST: direct imputation of summary statistics for unmeasured SNPs. Bioinformatics *29*, 2925–2927.
63. Pasaniuc, B., Zaitlen, N., Shi, H., Bhatia, G., Gusev, A., Pickrell, J., Hirschhorn, J., Strachan, D.P., Patterson, N., and Price, A.L. (2014). Fast and accurate imputation of summary statistics enhances evidence of functional enrichment. Bioinformatics *30*, 2906–2914.
64. Liu, D.J., Peloso, G.M., Zhan, X., Holmen, O.L., Zawistowski, M., Feng, S., Nikpay, M., Auer, P.L., Goel, A., Zhang, H., et al. (2014). Meta-analysis of gene-level tests for rare variant association. Nat. Genet. *46*, 200–204.
65. Sun, L., Craiu, R.V., Paterson, A.D., and Bull, S.B. (2006). Stratified false discovery control for large-scale hypothesis testing with application to genome-wide association studies. Genet. Epidemiol. *30*, 519–530.
66. Eskin, E. (2008). Increasing power in association studies by using linkage disequilibrium structure and molecular function as prior information. Genome Res. *18*, 653–660.
67. Darnell, G., Duong, D., Han, B., and Eskin, E. (2012). Incorporating prior information into association studies. Bioinformatics *28*, i147–i153.
68. Carbonetto, P., and Stephens, M. (2013). Integrated enrichment analysis of variants and pathways in genome-wide association studies indicates central role for IL-2 signaling genes in type 1 diabetes, and cytokine signaling genes in Crohn's disease. PLoS Genet. *9*, e1003770.
69. Gaffney, D.J., Veyrieras, J.-B., Degner, J.F., Pique-Regi, R., Pai, A.A., Crawford, G.E., Stephens, M., Gilad, Y., and Pritchard, J.K. (2012). Dissecting the regulatory architecture of gene expression QTLs. Genome Biol. *13*, R7.
70. Gagliano, S.A., Barnes, M.R., Weale, M.E., and Knight, J. (2014). A Bayesian method to incorporate hundreds of functional characteristics with association evidence to improve variant prioritization. PLoS ONE *9*, e98122.
71. Visscher, P.M., Brown, M.A., McCarthy, M.I., and Yang, J. (2012). Five years of GWAS discovery. Am. J. Hum. Genet. *90*, 7–24.
72. Green, P., and Ewing, B. (2013). Comment on "Evidence of abundant purifying selection in humans for recently acquired regulatory functions". Science *340*, 682.
73. Pique-Regi, R., Degner, J.F., Pai, A.A., Gaffney, D.J., Gilad, Y., and Pritchard, J.K. (2011). Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. Genome Res. *21*, 447–455.
74. Lappalainen, T., Sammeth, M., Friedländer, M.R., 't Hoen, P.A., Monlong, J., Rivas, M.A., Gonzàlez-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G., et al.; Geuvadis Consortium (2013). Transcriptome and genome sequencing uncovers functional variation in humans. Nature *501*, 506–511.
75. Davis, L.K., Yu, D., Keenan, C.L., Gamazon, E.R., Konkashbaev, A.I., Derks, E.M., Neale, B.M., Yang, J., Lee, S.H., Evans, P., et al. (2013). Partitioning the heritability of Tourette syndrome and obsessive compulsive disorder reveals differences in genetic architecture. PLoS Genet. *9*, e1003864.
76. Busing, F.M., Meijer, E., and Van Der Leeden, R. (1999). Delete-m jackknife for unequal m. Stat. Comput. *9*, 3–8.
77. Janss, L., de Los Campos, G., Sheehan, N., and Sorensen, D. (2012). Inferences from genomic models in stratified populations. Genetics *192*, 693–704.
78. Golan, D. and Rosset, S. (2013). Narrowing the gap on heritability of common disease by direct estimation in case-control GWAS. arXiv, arXiv:1305.5363, http://arxiv.org/abs/1305.5363.
79. Hayeck, T., Zaitlen, N., Loh, P.-R., Vilhjalmsson, B., Pollack, S., Gusev, A., Yang, J., Chen, G.-B., Goddard, M.E., Visscher, P.M., et al. (2014). Mixed model with correction for case-control ascertainment increases association power. bioRxiv http://dx.doi.org/10.1101/008755.
80. Haseman, J.K., and Elston, R.C. (1972). The investigation of linkage between a quantitative trait and a marker locus. Behav. Genet. *2*, 3–19.

The American Journal of Human Genetics, Volume *95*

Supplemental Data

Partitioning Heritability of Regulatory and Cell-Type-Specific Variants across 11 Common Diseases

Alexander Gusev, S. Hong Lee, Gosia Trynka, Hilary Finucane, Bjarni J. Vilhjálmsson, Han Xu, Chongzhi Zang, Stephan Ripke, Brendan Bulik-Sullivan, Eli Stahl, Schizophrenia Working Group of the Psychiatric Genomics Consortium, SWE-SCZ Consortium, Anna K. Kähler, Christina M. Hultman, Shaun M. Purcell, Steven A. McCarroll, Mark Daly, Bogdan Pasaniuc, Patrick F. Sullivan Benjamin M. Neale, Naomi R. Wray, Soumya Raychaudhuri, and Alkes L. Price

Supplemental Figures

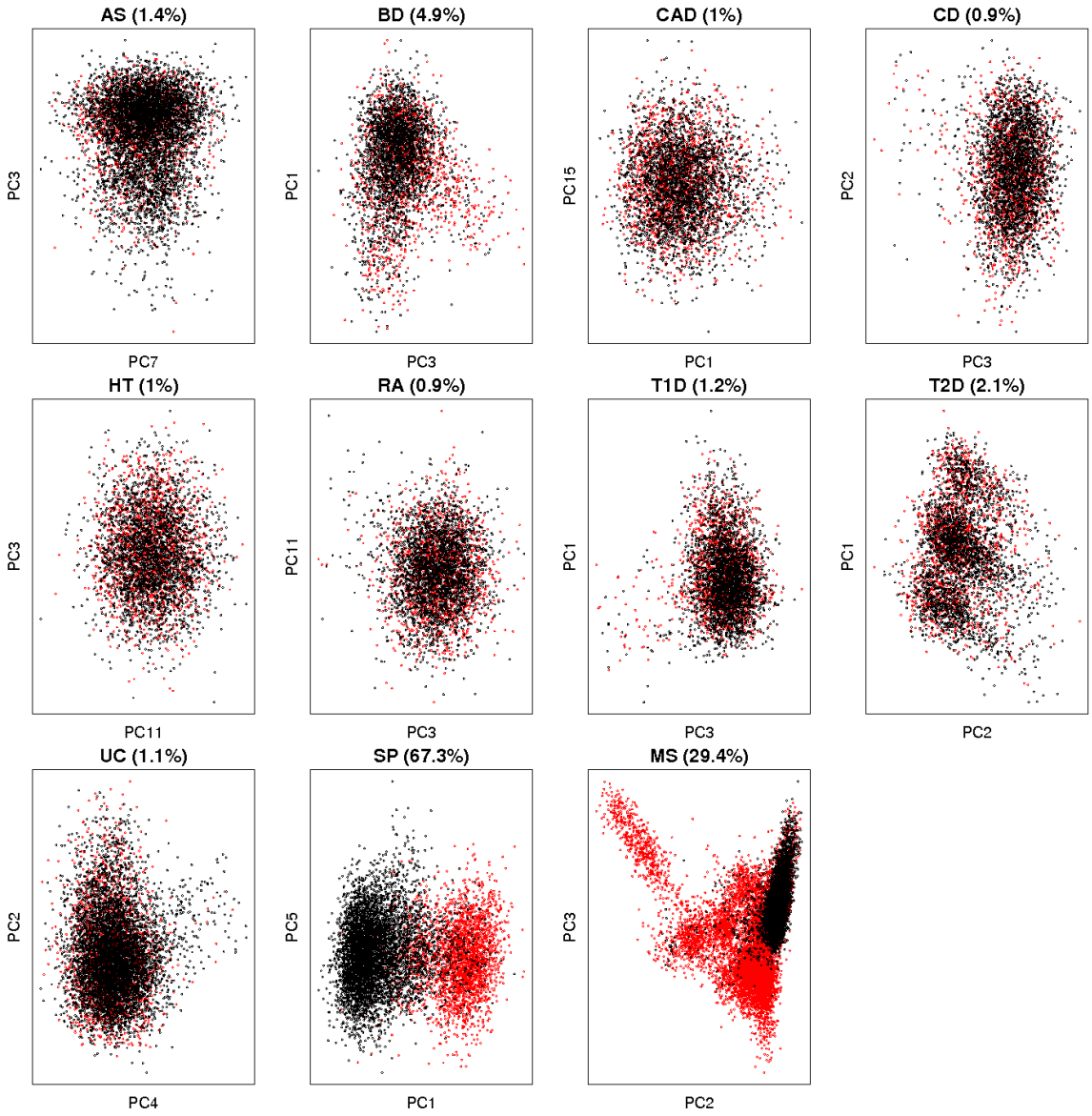

Figure S1. Principal components analysis of WTCCC samples. Two most significant principal components are plotted for each disease cohort, with cases and controls color coded red and black respectively. Each sub-panel label specifies the variance in phenotype explained by all 20PCs in parentheses. MS and SP cohorts are known to be highly structured due to environment and ascertainment.

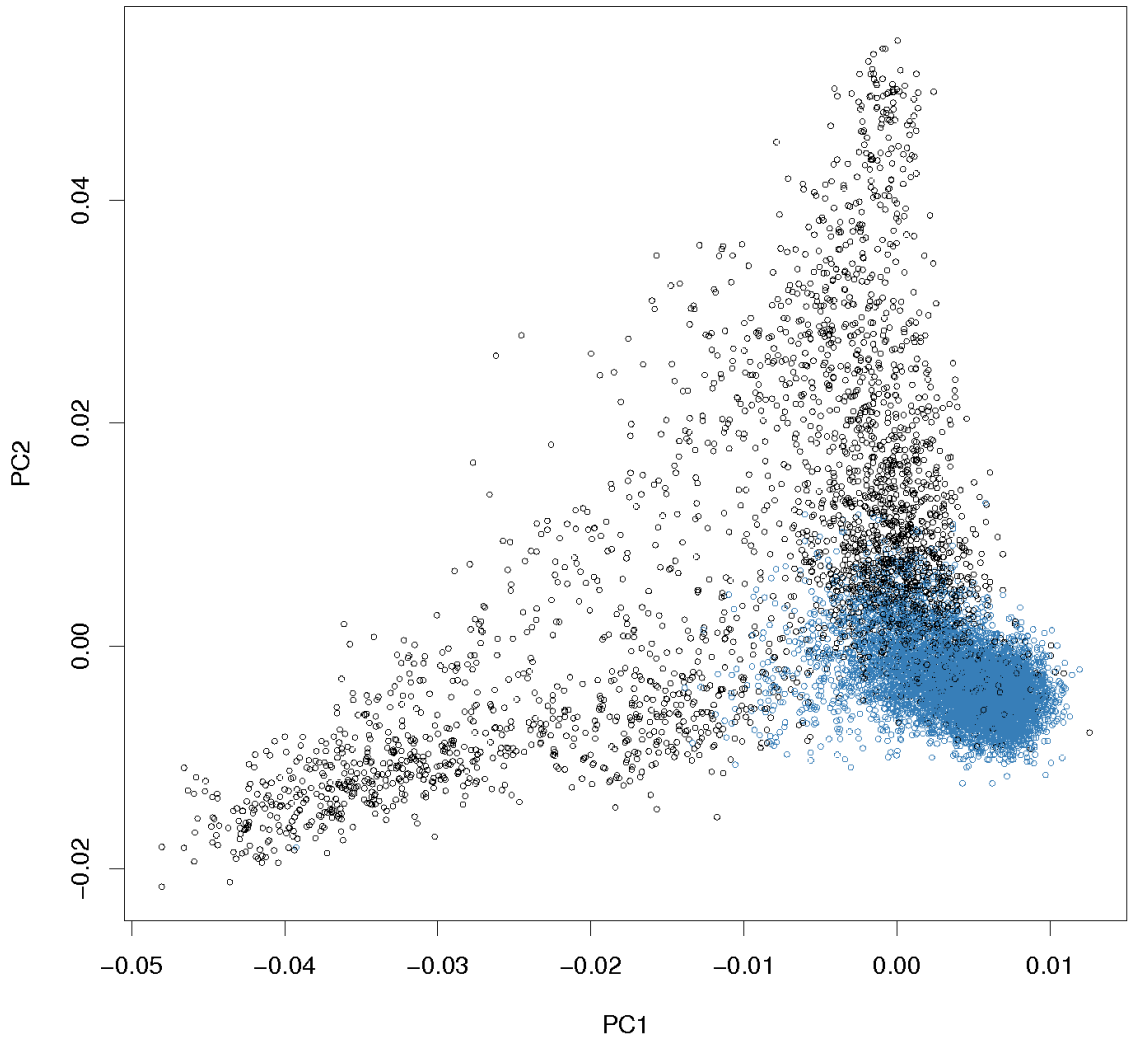

Figure S2. Principal components analysis of Swedish samples. Two main principal components are shown for analysis of GWAS data from the full Swedish Schizophrenia cohort. Homogenous Swedish samples are highlighted in blue.

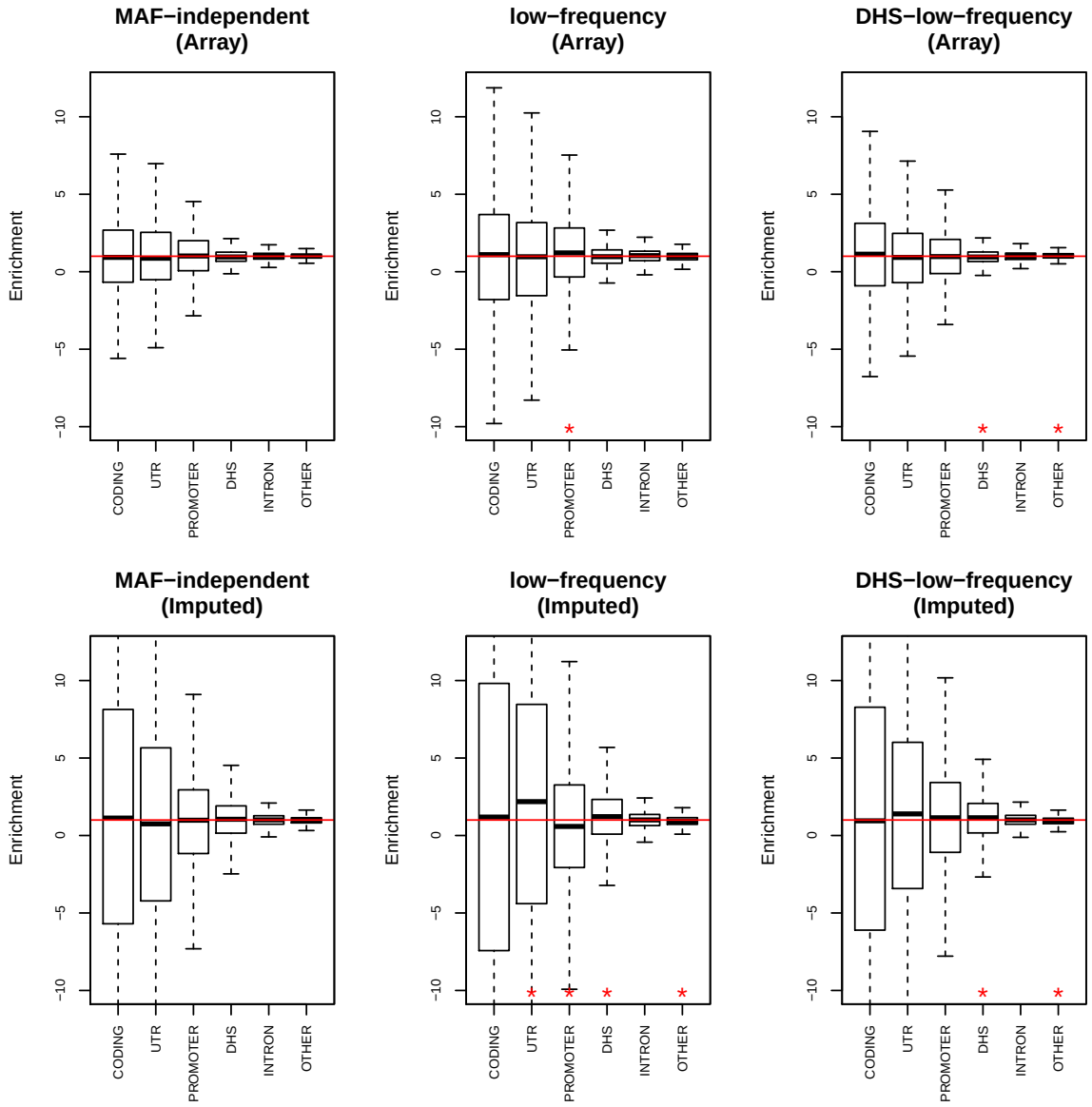

Figure S3. Observed variance-component enrichment from simulated null architecture. Distribution of enrichment estimate over 1000 simulations with three different disease architectures performed in genotyped SNPs (top) and imputed SNPs (bottom). All phenotypes simulated without category-specific enrichment, red line showing expected enrichment of 1.0×. Red asterisk indicates significant difference from expectation (by z-test, accounting for 36 comparisons).

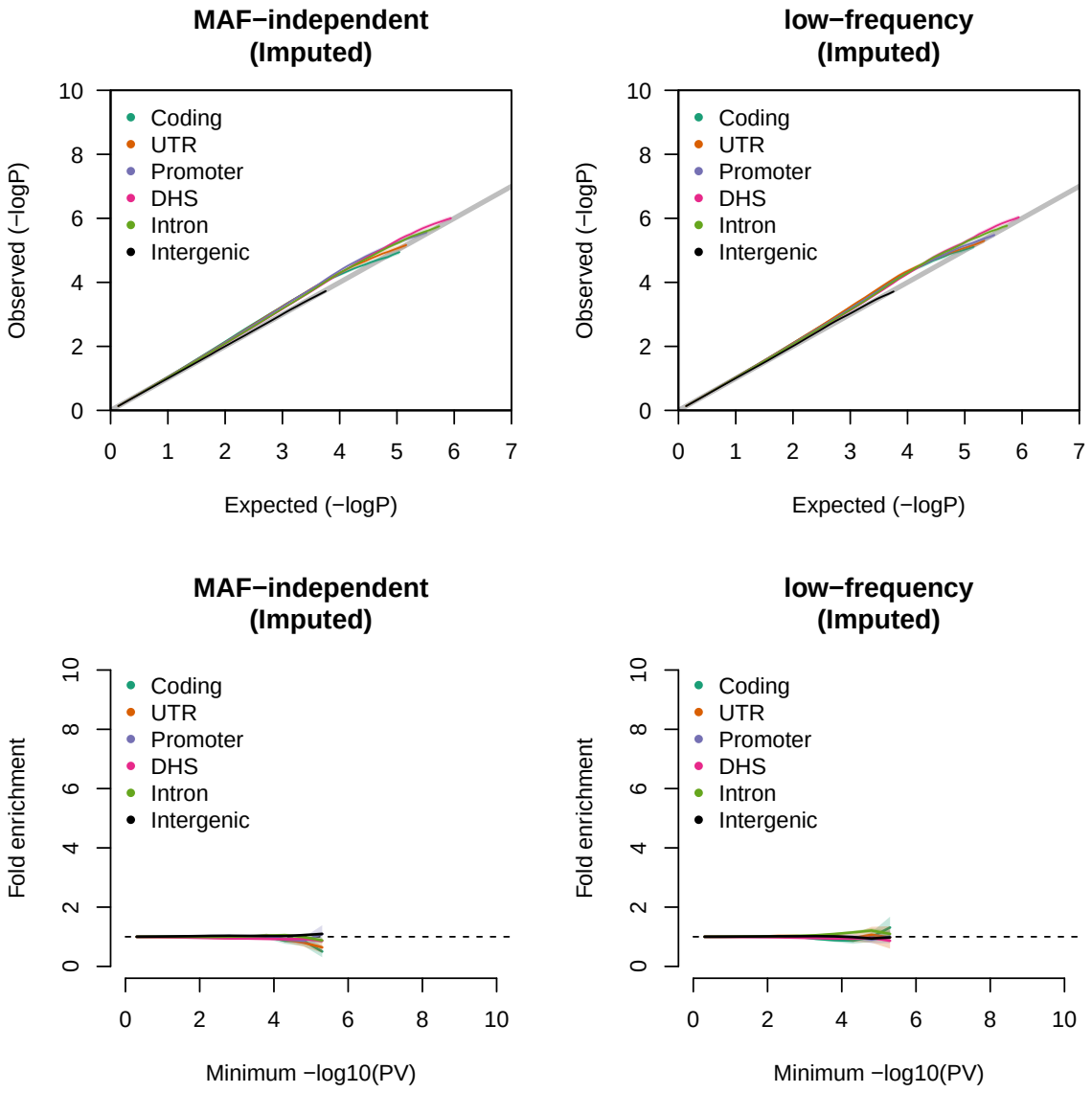

Figure S4. Enrichment of summary statistics under the null. Results for MAF-independent (left) and low-frequency (right) architectures shown for stratified QQ-plots (top) and P-value enrichment plots (bottom).

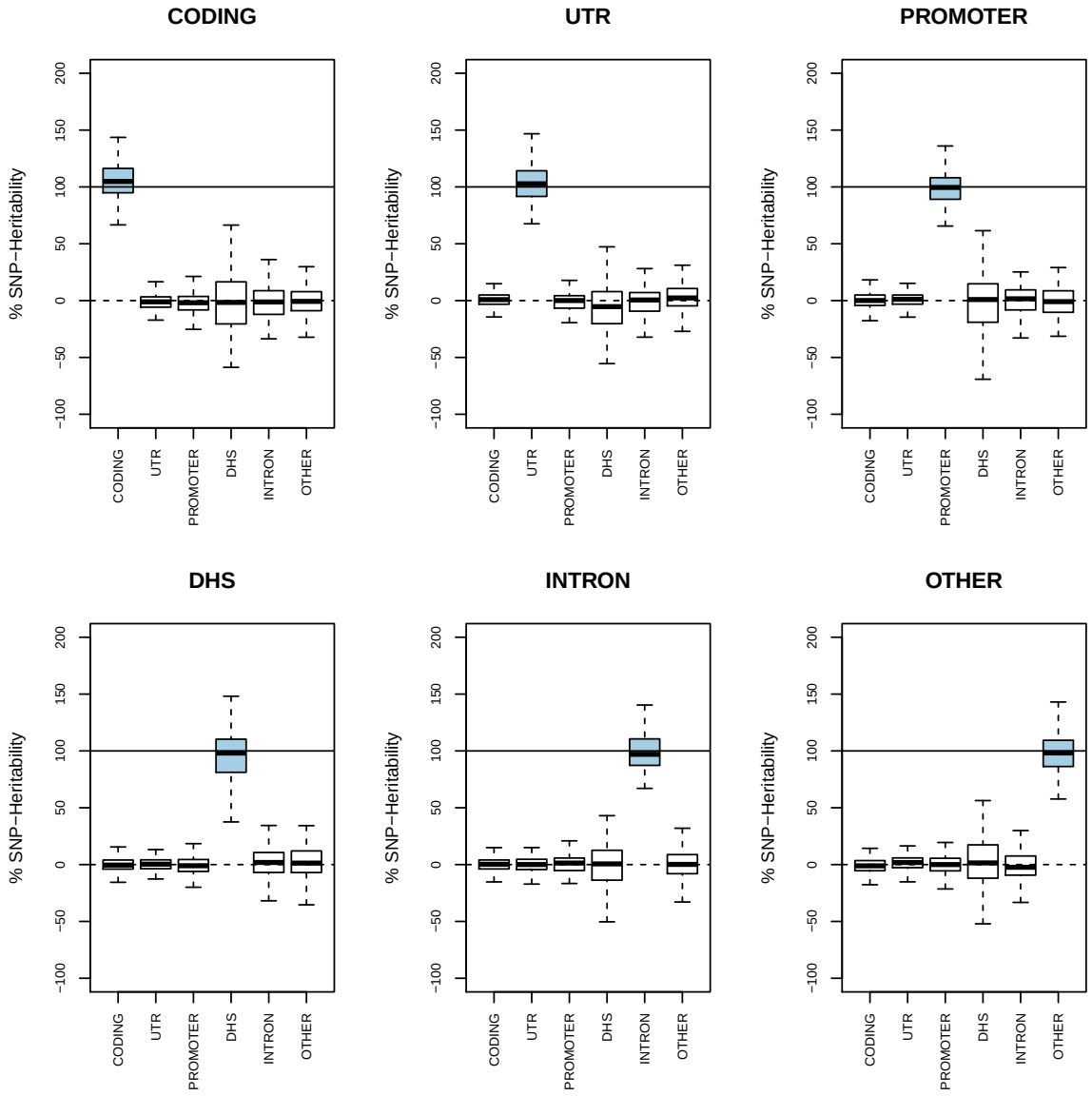

Figure S5. Partitioning of $h^2_{\mathbf{g}}$ with imputed SNPs and MAF-independent causals. Estimate of h^2_{g} from imputed SNPs in each functional category for phenotypes simulated from imputed SNPs with any frequency. Each section of the figure describes results from 200 simulations where all h^2_{g} was induced in the titular functional category (highlighted in blue).

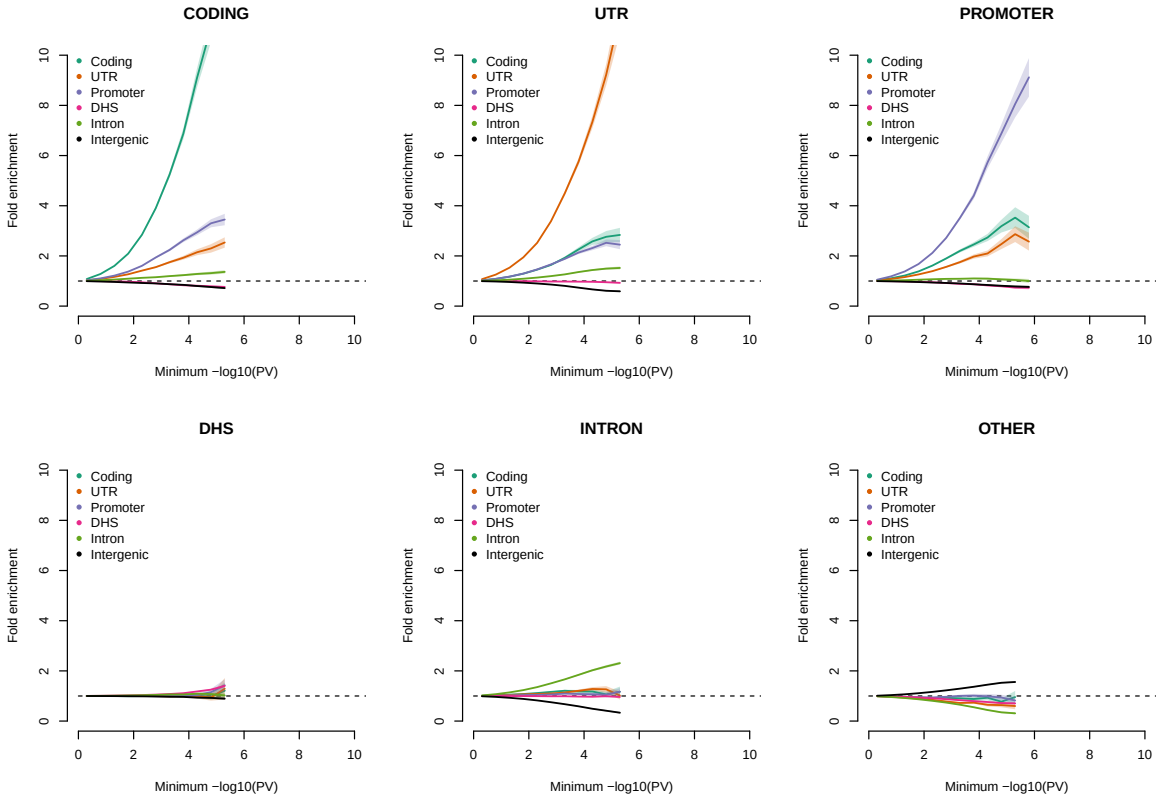

Figure S6. Observed P-value enrichment from simulated enrichment (MAF-independent).

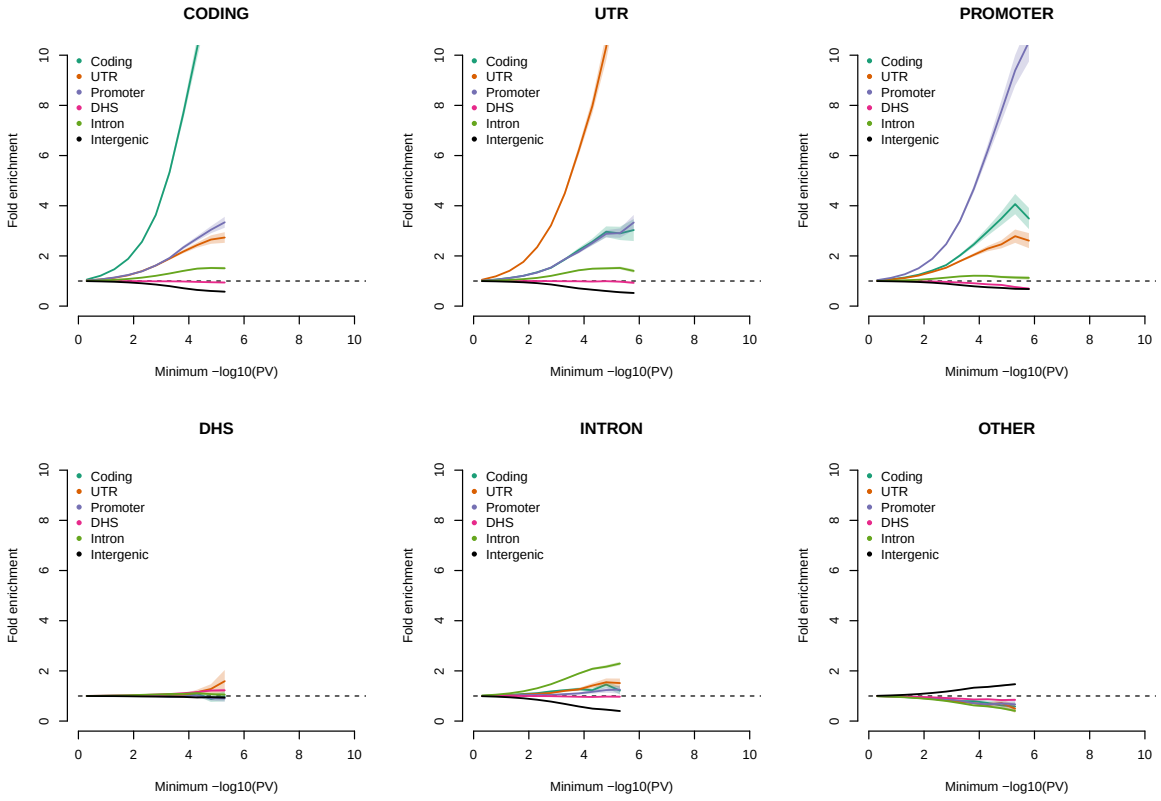

Figure S7. Observed P-value enrichment from simulated enrichment (low-frequency).

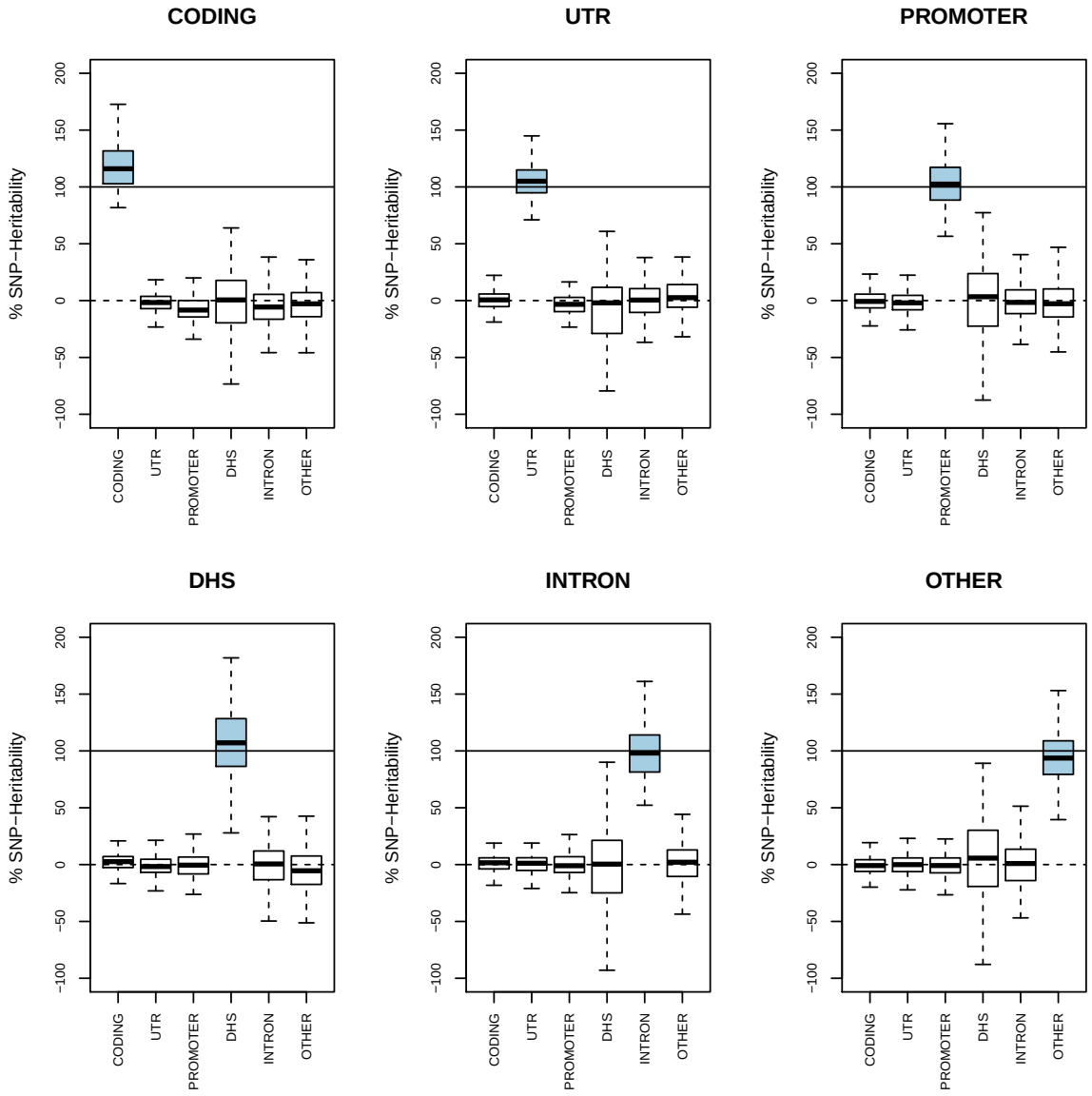

Figure S8. Partitioning of $h^2_{\mathbf{g}}$ with imputed SNPs and low-frequency causals. Estimate of h^2_{g} from imputed SNPs in each functional category for phenotypes simulated from imputed SNPs with MAF < 0.05. Each section of the figure describes results from 200 simulations where all h^2_{g} was induced in the titular functional category (highlighted in blue).

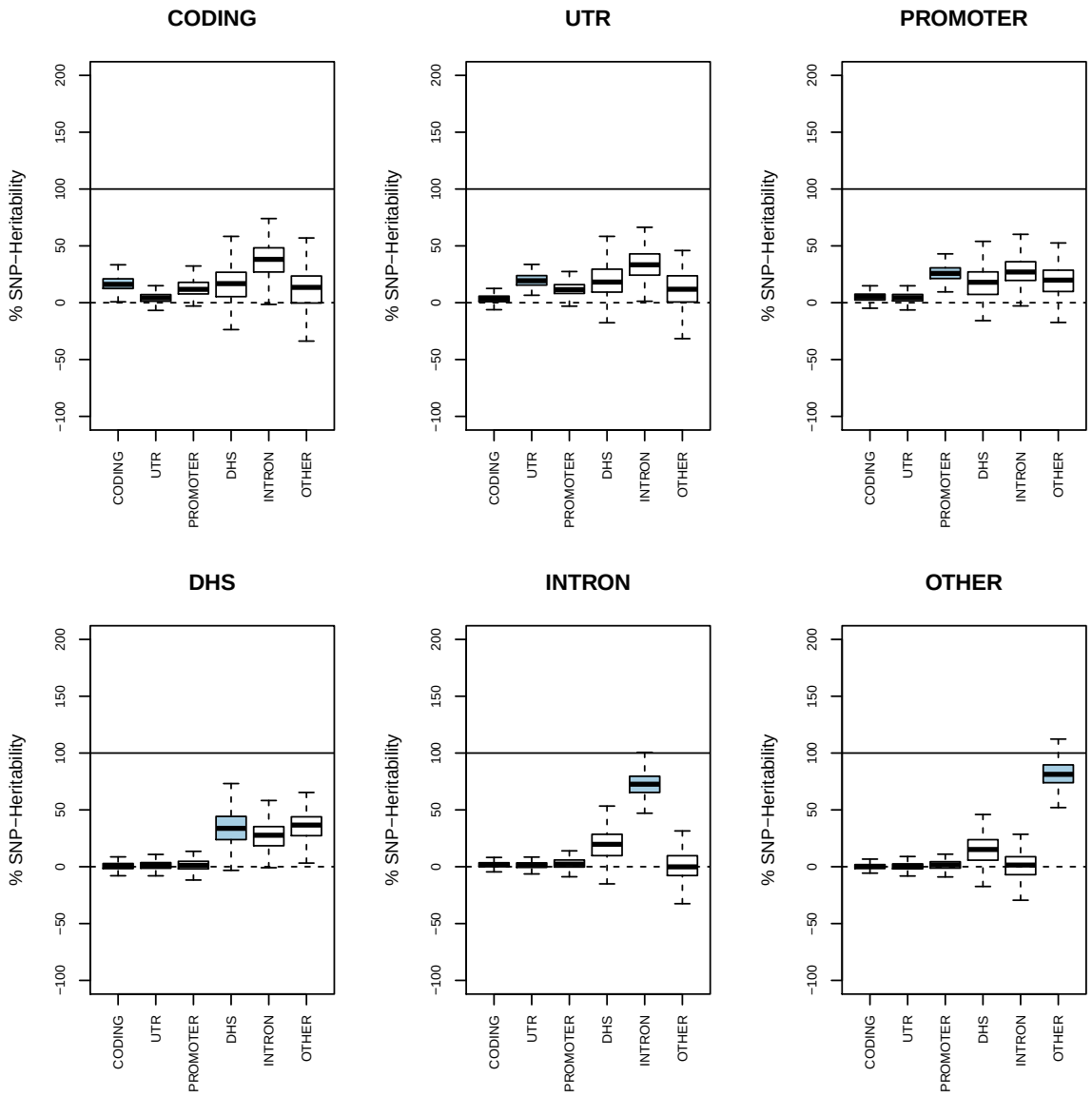

Figure S9. Partitioning of $h_{\mathbf{g}}^2$ simulated in MAF-independent imputed data. Estimate of h_{g}^2 from genotyped SNPs in each functional category for phenotypes simulated from imputed SNPs from any MAF. Each section of the figure describes average results from simulations where all h_{g}^2 was induced in the titular functional category (highlighted in blue). Estimate h_{g}^2 is spread across multiple functional categories due to incomplete tagging. Error-bars indicate standard error from 200 simulations.

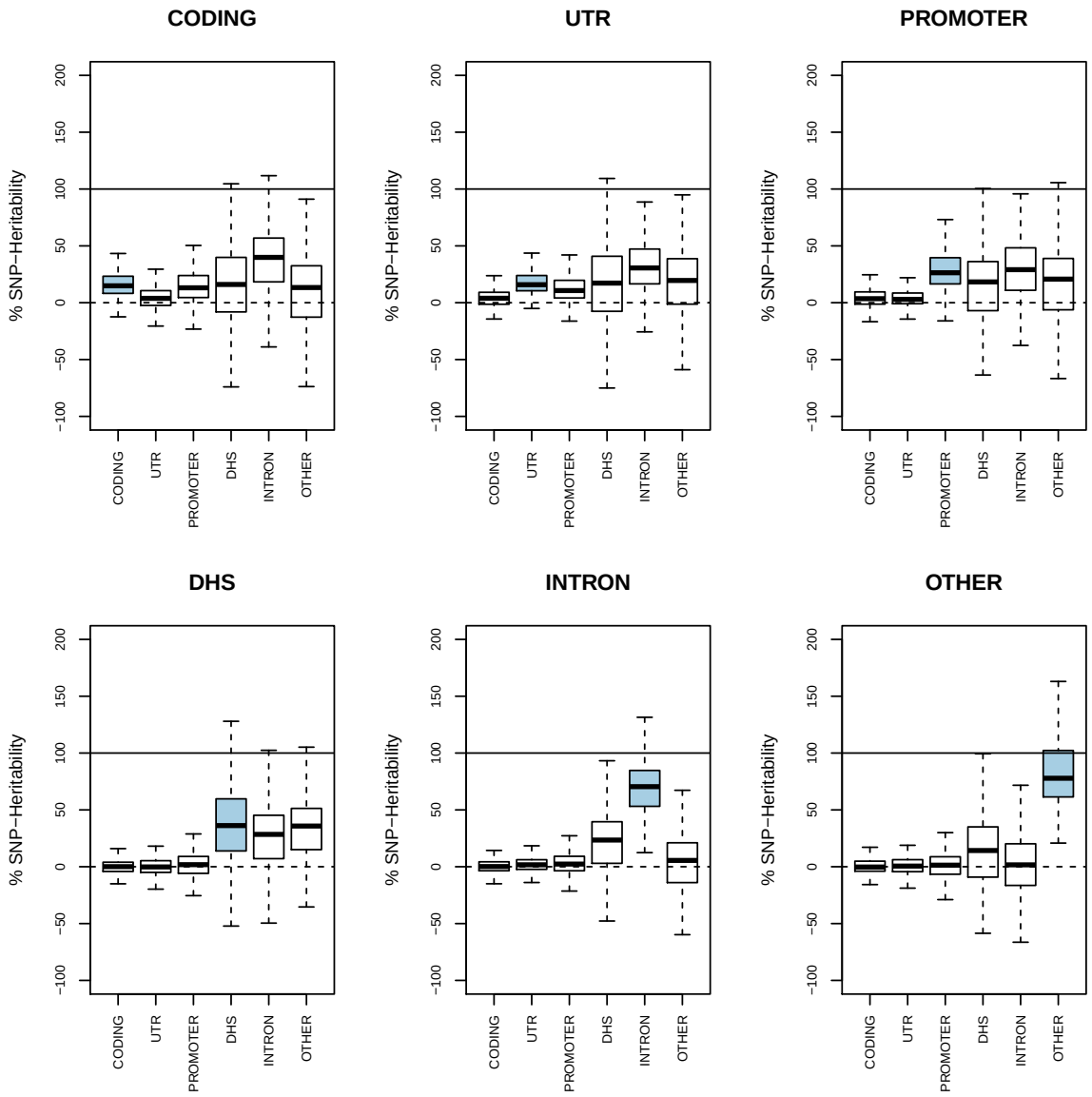

Figure S10. Partitioning of h_g^2 simulated in low-frequency imputed data. Estimate of h_g^2 from genotyped SNPs in each functional category for phenotypes simulated from imputed SNPs with MAF < 0.05. Each section of the figure describes average results from simulations where all h_g^2 was induced in the titular functional category (highlighted in blue). Estimate h_g^2 is spread across multiple functional categories due to incomplete tagging. Error-bars indicate standard error from 200 simulations.

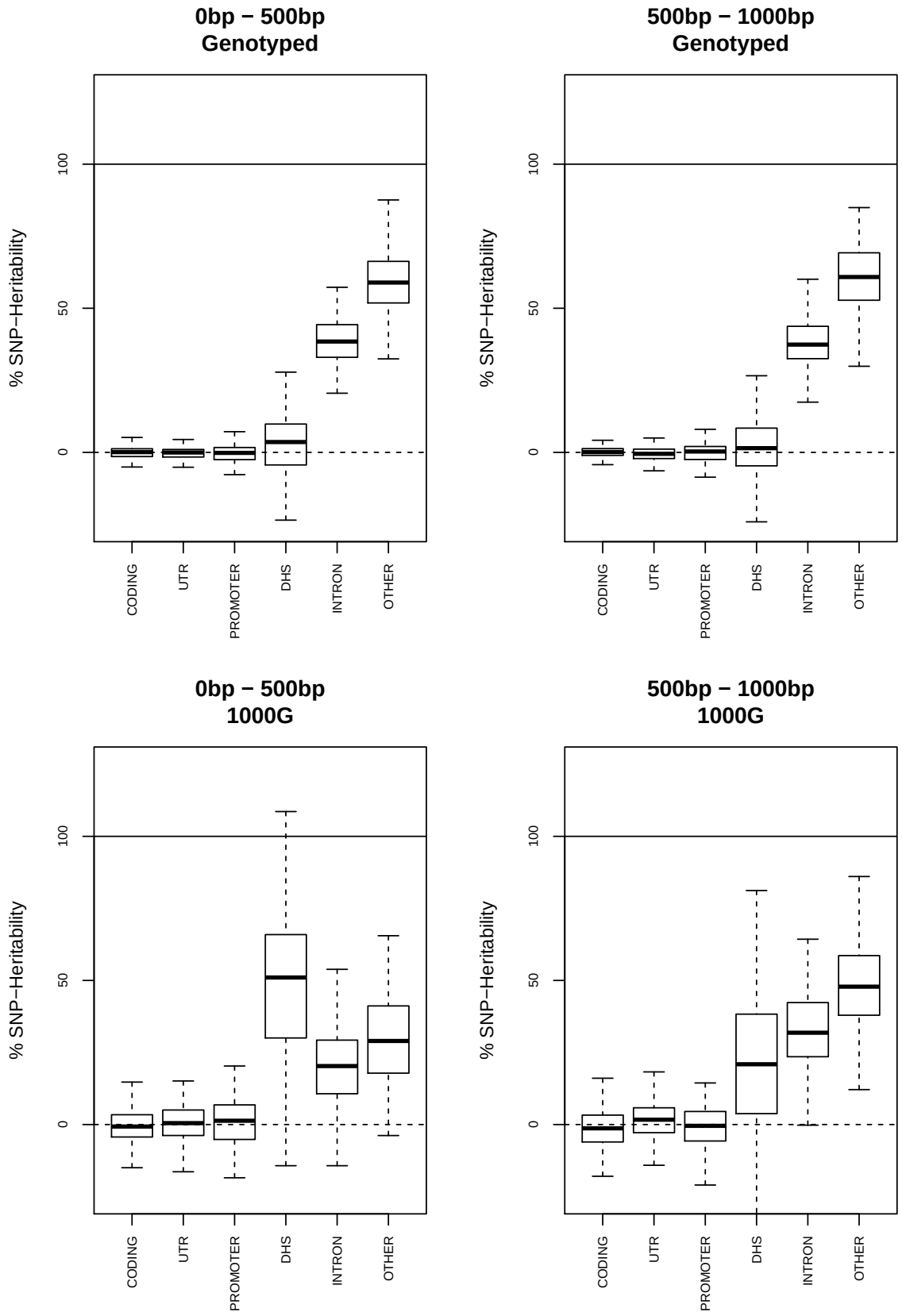

Figure S11. Partitioning of $h^2_{\mathbf{g}}$ from causal variants at the DHS boundary. Causal variants were sampled from non-DHS intronic and intergenic regions within 0-500bp (left, 29% of imputed SNPs) and 500-1,000bp (right, 15% of imputed SNPs) of any DHS region boundary. Box-plots shown % h^2_{g} estimates over 200 simulations with MAF-independent causal variants. Phenotypes and GRMs from genotyped SNPs (top) and from imputed SNPs (bottom).

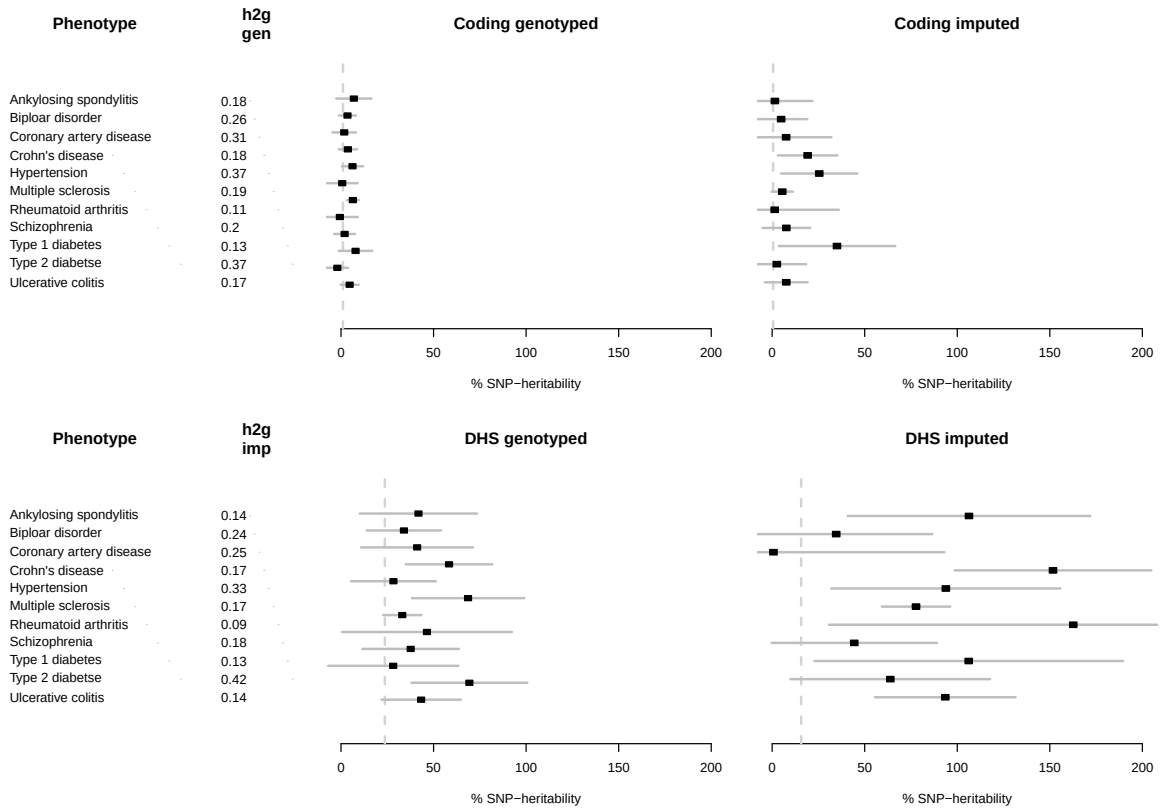

Figure S12. Individual trait analysis of coding and DHS variants. Forest plot of % h_g^2 inferred for each trait over coding SNPs (top) and DHS SNPs (bottom). Total h_g^2 shown for each trait and SNP platform in second column.

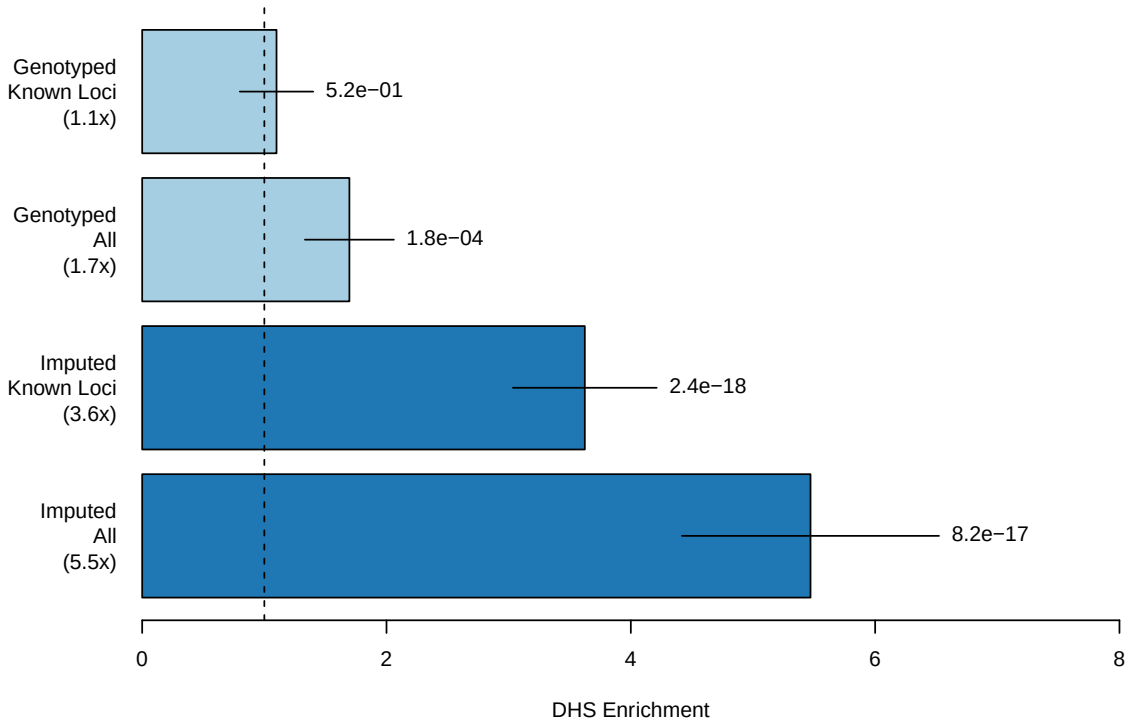

Figure S13. Functional enrichment of SNP-heritability in DHS regions. The ratio of observed % heritability over corresponding % of SNPs is reported as meta-analysis over all traits for four locus types. Light blue bars detail analysis of typed SNPs and dark blue bars detail analysis of typed and 1,000 Genomes imputed SNPs. "Known Loci" categories corespond to analysis restricted to 1MB regions around published genome-wide significant loci for the corresponding trait. We note that the choice of region size may impact the absolute enrichment, with larger regions expected to appear more like the genome-wide enrichment and yield a conservative estimate of the difference. This region size is expected to yield a representative estimate[1]. SP, HT, and BD had too few known loci or could not converge in the local analyses and were excluded from all computations, resulting in slightly different overall values from Figure 1. Error bars define 95% confidence interval after adjusting for shared controls.

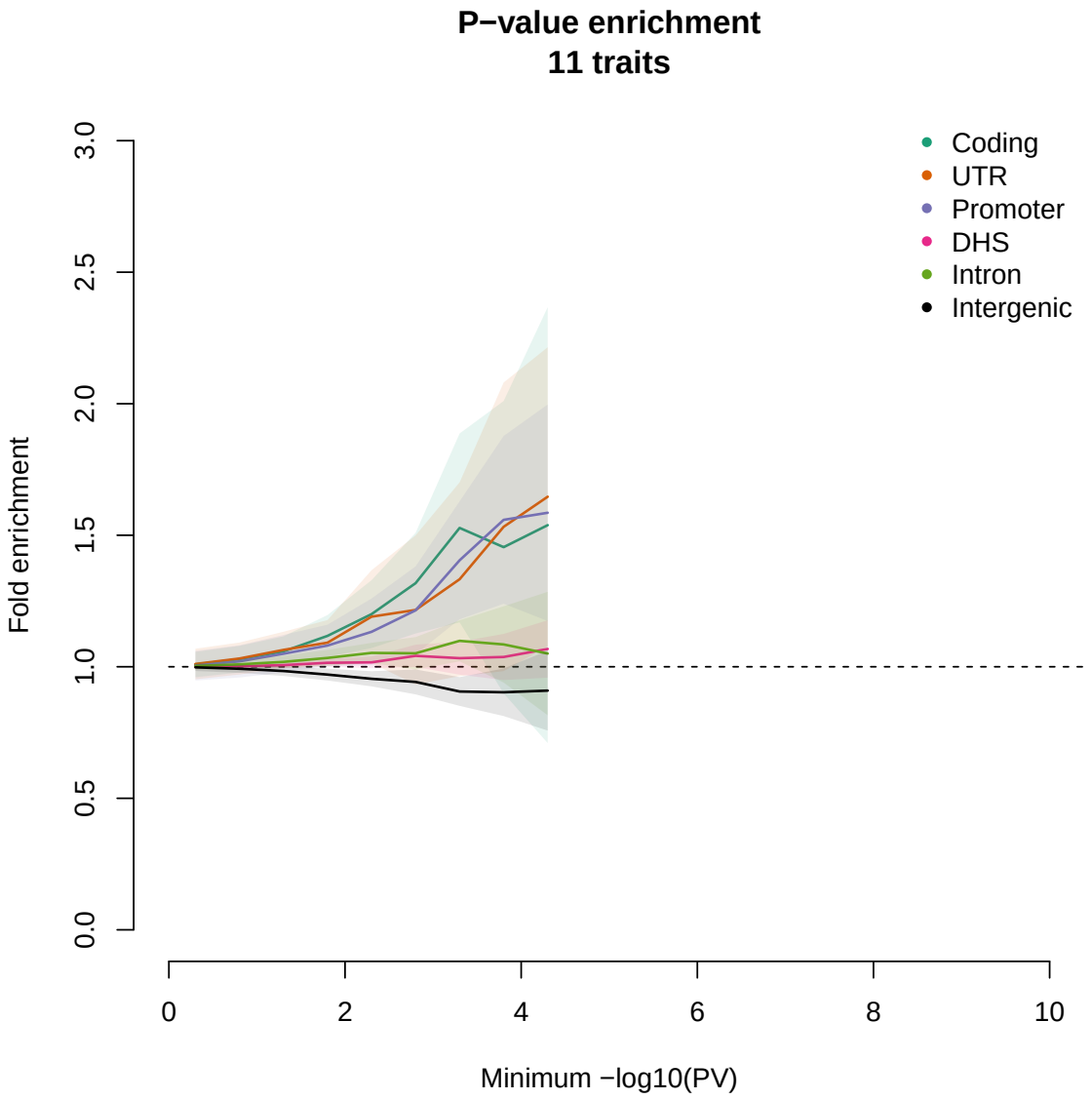

Figure S14. ***P*-value enrichment in 11 traits.** Fold-enrichment of P-values meeting a given significance threshold in each functional category. Enrichment plotted for all thresholds that contain at least 100 SNPs. Average over 11 traits shown in top-left for thresholds observed in all traits, with shaded region corresponding to standard error.

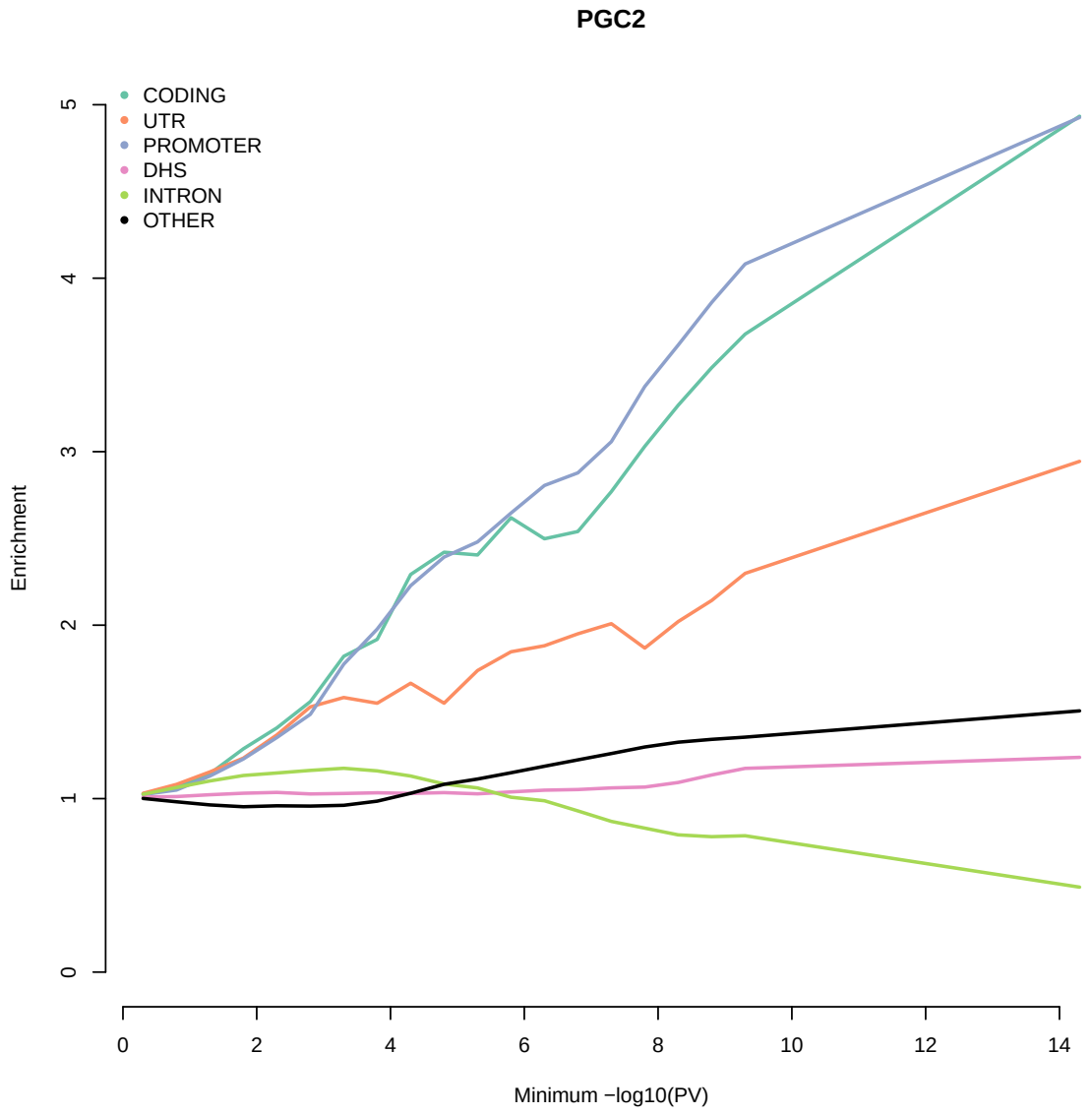

Figure S15. ***P*-value enrichment in PGC2.** Fold-enrichment of P-values meeting a given significance threshold in each functional category.

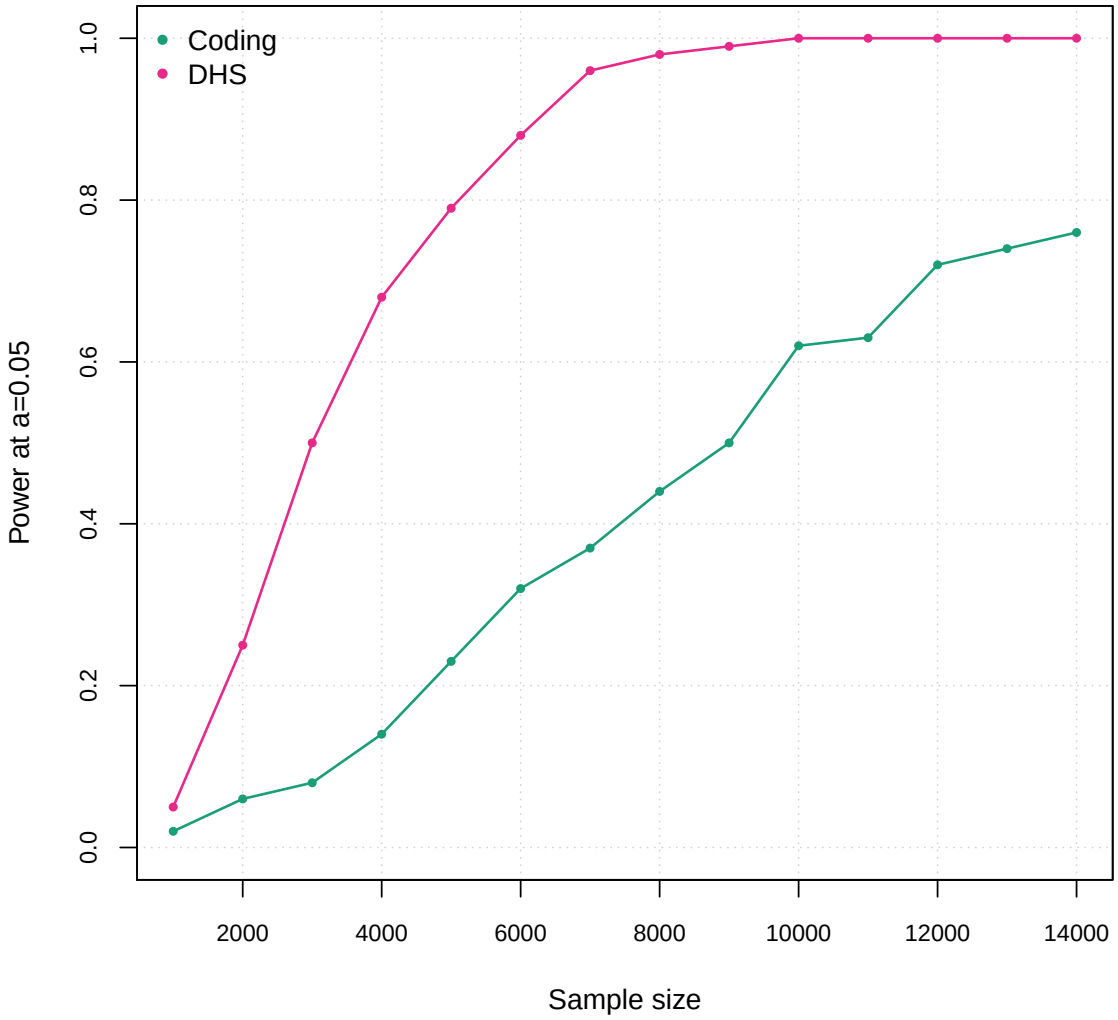

Figure S16. Power to detect significant $h^2_{\mathbf{g}}$ enrichment. Phenotypes were simulated with DHS and coding enrichment matching the observed meta-analysis values in a 33,000 sample cohort. Power was then inferred as the fraction of 100 simulations where enrichment was significant at $P < 0.05$ over increasing sample sizes.

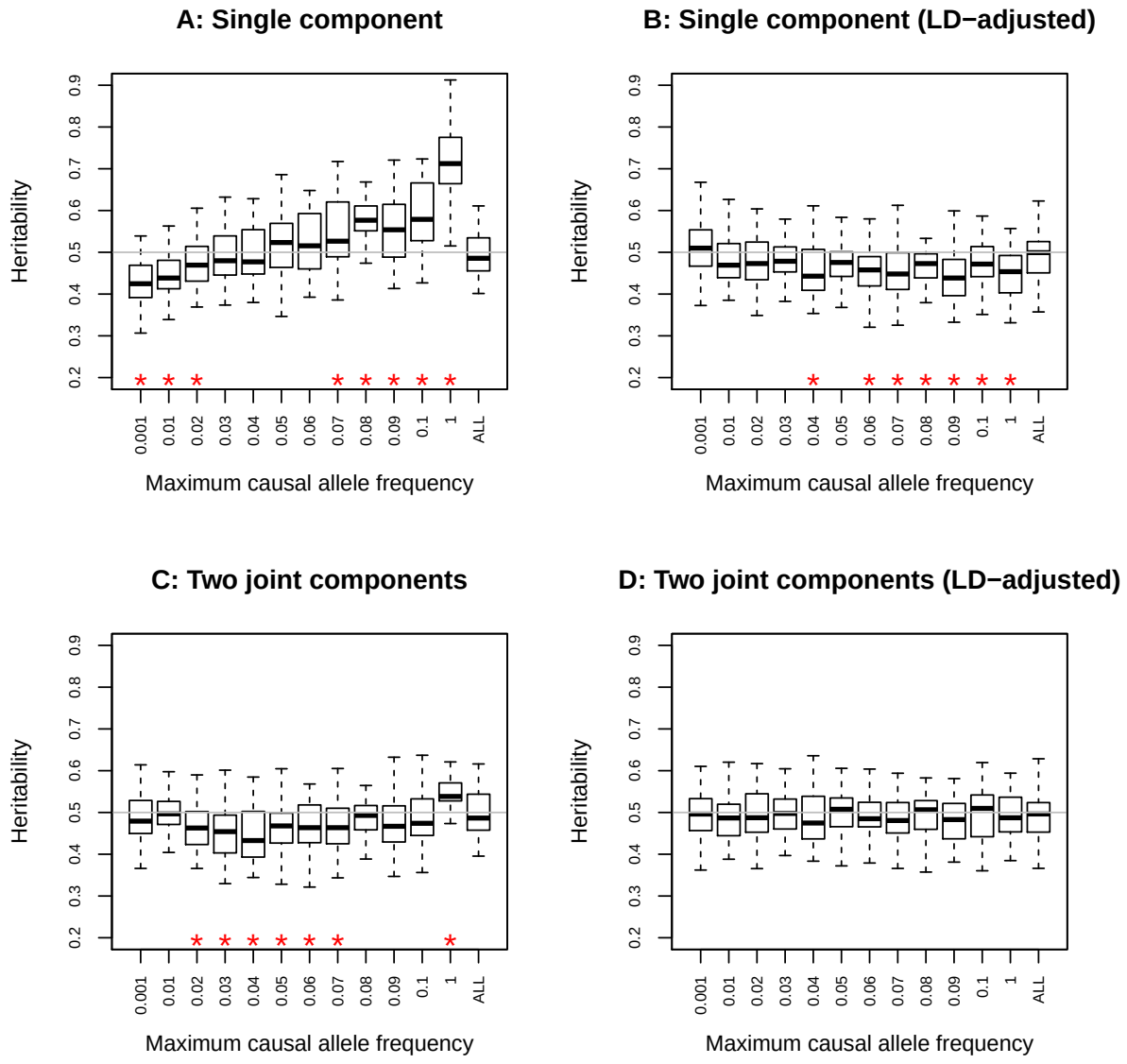

Figure S17. Heritability estimates in simulation with normalized allelic effect-sizes. Distribution of h^2_g inferred by four variance-component models is shown over a range of disease architectures. Additive phenotypes with $h^2 = 0.5$ were simulated from 1,000 randomly selected causal variants with maximum allele frequency from 0.01 to 0.1 (x-axis). Normalized SNP effect-sizes were drawn from the standard normal such that each SNP explains equal variance in expectation. Box-plots show inferred h^2_g over 40 random simulations. For the joint component model the sum of both inferred h^2_g values is reported. A red asterisk indicates significant difference from 0.5 by z-test after correcting for ten architectures tested. Under the un-adjusted single-component model we observe both kinds of bias depending on the causal allele frequency cutoff. When causal variants are primarily rare (MAF ≤ 0.02) the mean estimate is significantly deflated down to 0.45, whereas when causal variants are more common (MAF ≤ 0.1) the mean estimate is significantly deflated up to 0.59. LD adjustment [1] of the single component appears to fix the downwards bias, with mean estimate no lower than 0.49 (not significantly different from 0.50) but does not completely mitigate the upwards bias, with a mean estimate up to 0.57. Splitting the data into two components for rare and common SNPs entirely removes the upwards bias but introduces downwards bias in most instances where causal variants can be common. Combining the two strategies and using two internally LD-adjusted [1] components yields completely unbiased estimates with no disease architecture exhibiting h^2_g significantly different from 0.5.

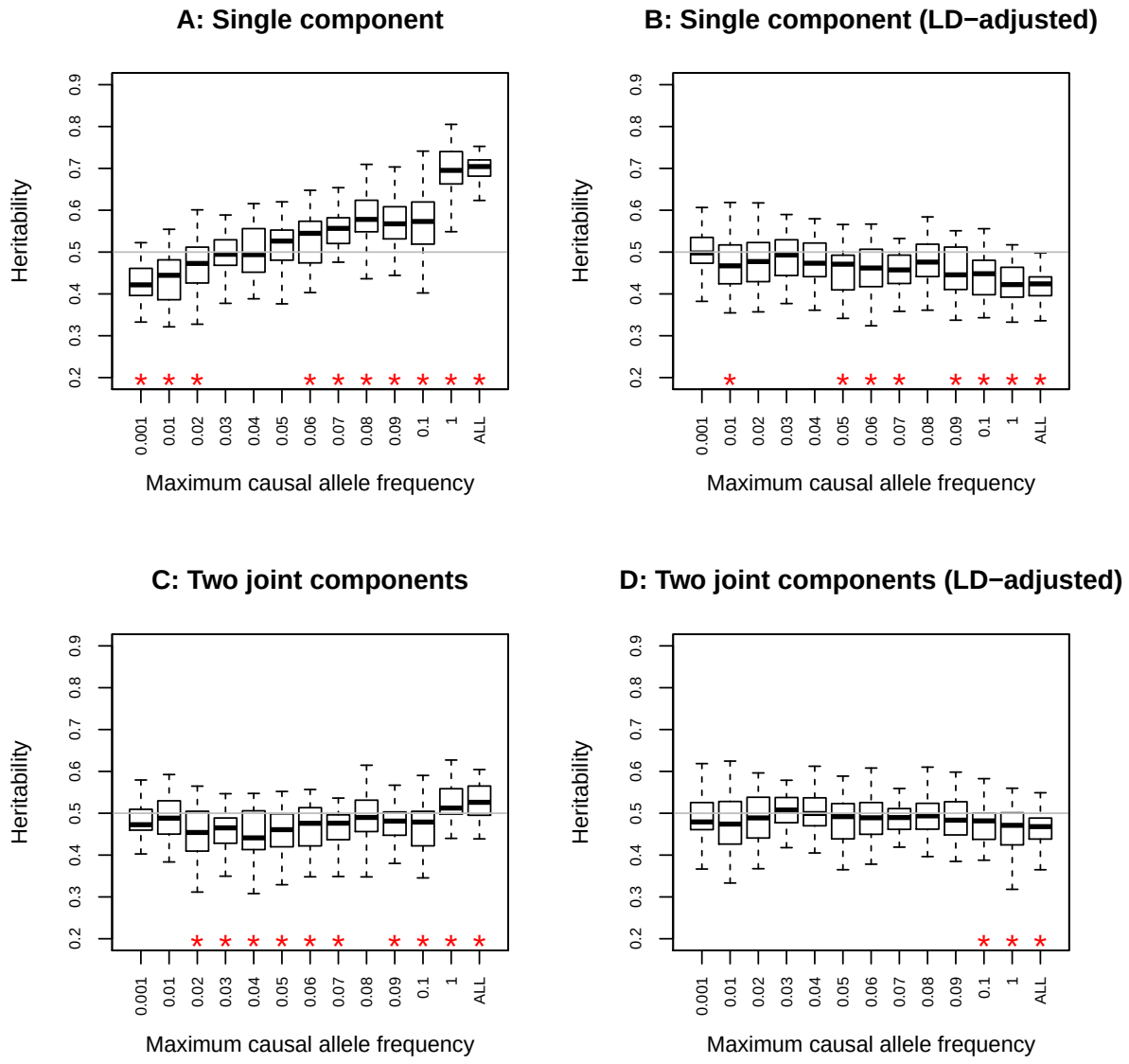

Figure S18. Heritability estimates in simulation with standard allelic effect-sizes. Distribution h^2_g inferred by four variance-component models is shown over a range of disease architectures. Additive phenotypes with $h^2 = 0.5$ were simulated from 1,000 randomly selected causal variants with maximum allele frequency from 0.01 to 0.1 (x-axis). Allelic effect-sizes were drawn from the standard normal such that common SNPs explain more variance in expectation. Box-plots show inferred h^2_g over 40 random simulations. For the joint component model the sum of both inferred h^2_g values is reported. A red asterisk indicates significant difference from 0.5 by z-test after correcting for ten architectures tested.

We considered weather the SNPs used to construct the GRM should be normalized by their observed variance or the expected variance $2p(1-p)$ based on the minor allele frequency p. We performed simulations for the two normalization schemes and two effect-size distributions. Under the infinitesimal model where every variant explains the same amount of phenotypic variance in expectation, we observed no differences between the normalizations for any class of SNPs. Under the neutral model where effect-size is proportional to the minor allele frequency, we observed a significant difference between the two normalizations when rare variants were included in the analysis, with the $2p(1-p)$ scaling resulting in a significant upwards bias. These findings indicate that rare variants have slight but consistent deviations from Hardy-Weinberg equilibrium that can affect the variance-component estimate under the $2p(1-p)$ normalization. To account for this, we use the observed variance to normalize markers in all analyses of rare variants.

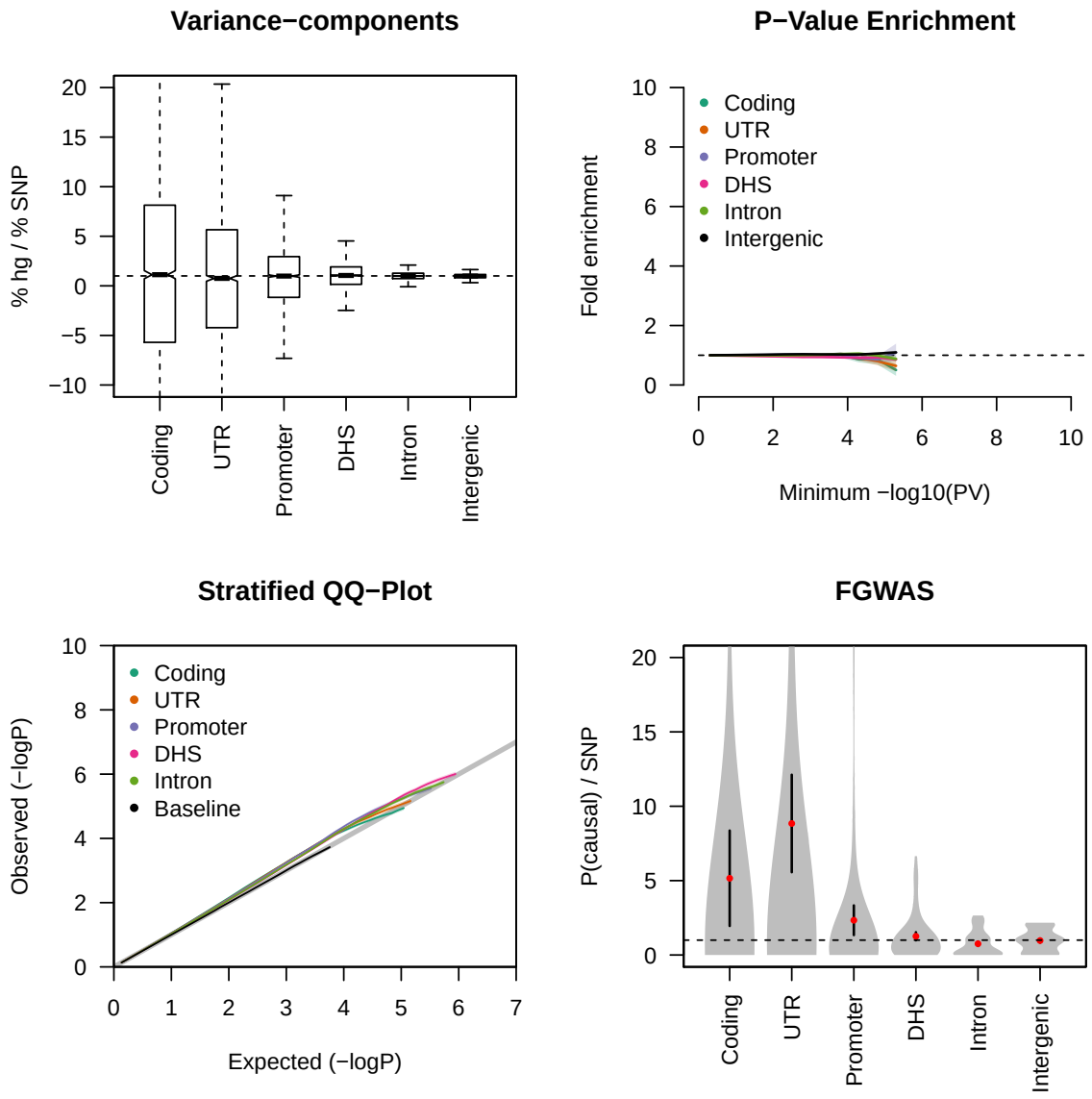

Figure S19. Estimates of functional enrichment under the null. We simulated a polygenic disease architecture with MAF-independent imputed causal SNPs uniformly drawn from all functional categories, corresponding to no enrichment. Simulated phenotypes were tested using the variance-component method (top left) from 3000 simulations; P-value enrichment (top right) from 100 simulations; stratified QQ-plot (bottom left) from 100 simulations; FGWAS (bottom right) from 100 converged simulations (out of $\sim$ 800 total). FGWAS plot contains mean (red point); 1.96$\times$ standard error (black line); and density function for the full distribution shown in gray. All methods showed no enrichment except FGWAS, which exhibited upward bias at smaller categories due to enrichment being restricted to the 0-1 scale.

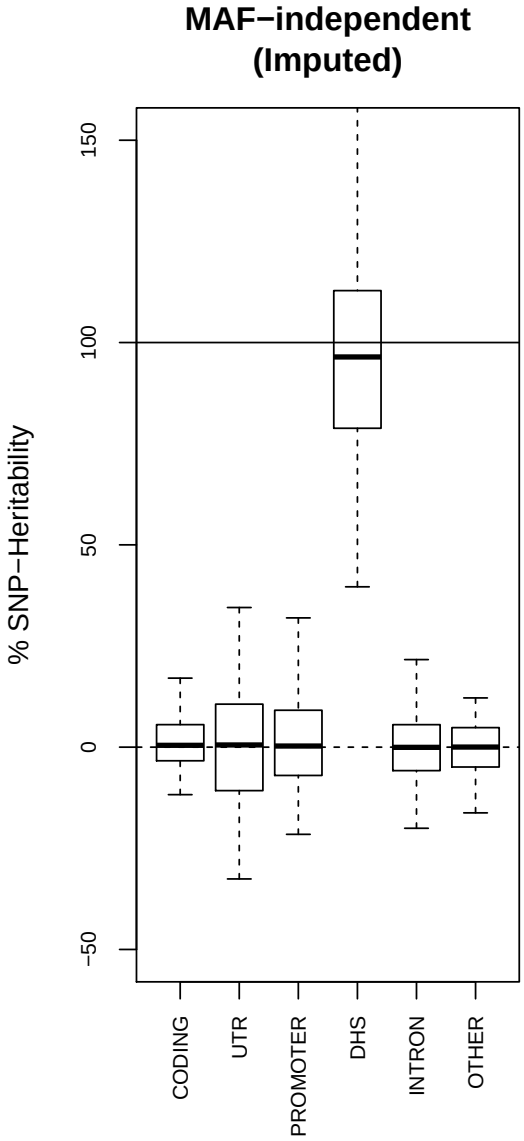

Figure S20. Partitioning of $h^2_{\mathbf{g}}$ with DHS centers enriched. Causal effect-sizes were sampled such that center of DHS (1% of genome) explains 25% of h^2 and remainder of DHS (15% of genome) explains 75% of h^2. Box-plots shown $\%h^2$ over 200 simulations with MAF-independent causal variants.

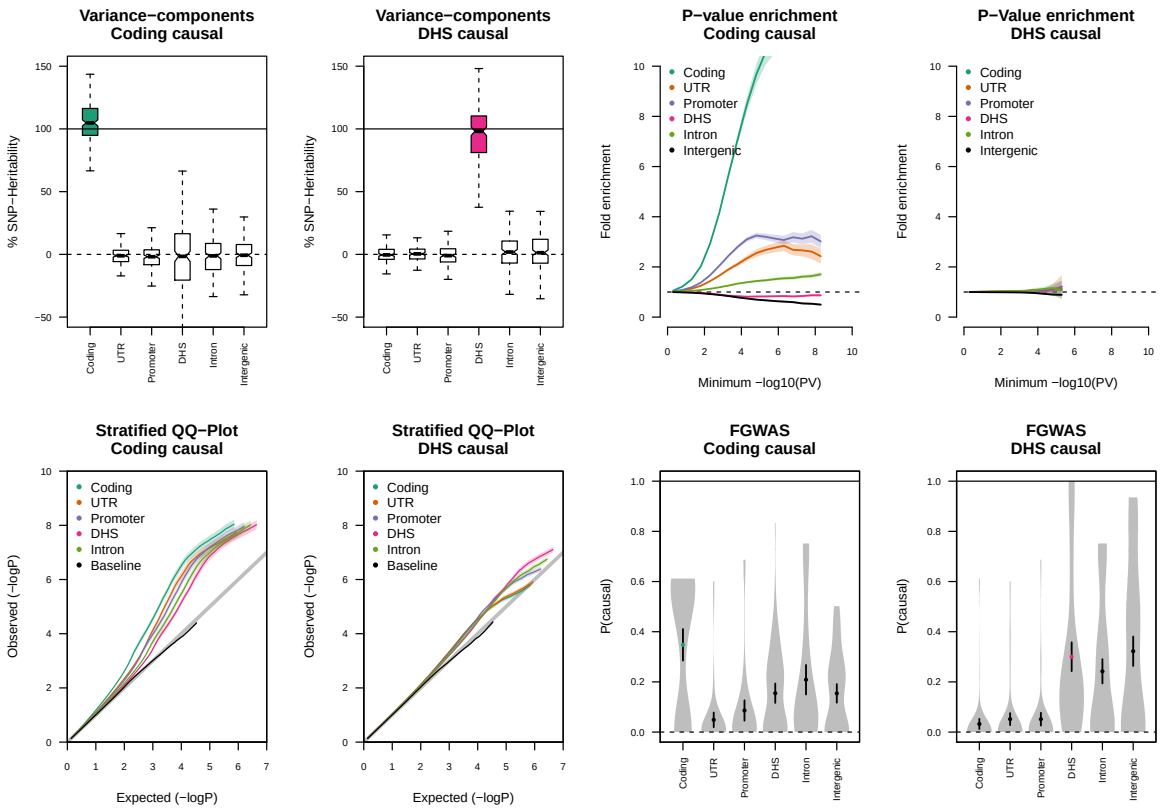

Figure S21. Estimates of functional enrichment from single causal category. We simulated a polygenic disease architecture with MAF-independent imputed causal SNPs drawn from a single functional category, corresponding to complete enrichment of the respective category. Simulated phenotypes were tested using the variance-component method (top left) from 1000 simulations; P-value enrichment (top right) from 100 simulations; stratified QQ-plot (bottom left) from 100 simulations; FGWAS (bottom right) from 100 converged simulations (out of $\sim$ 800 total). FGWAS plot contains mean (red point); 1.96$\times$ standard error (black line); and density function for the full distribution shown in gray. Stratified QQ-plot and P-value enrichment sub-plots show 1.96$\times$ standard error as shaded regions. In the DHS-causal scenario, GWAS-based methods underestimated the enrichment; while in the Coding-causal scenario, GWAS-based methods overestimated enrichment from other correlated categories. For each method, only the Coding causal and DHS causal scenarios are shown, additional simulations appear in Fig. S6, S7, S23, S24, S26, S27.

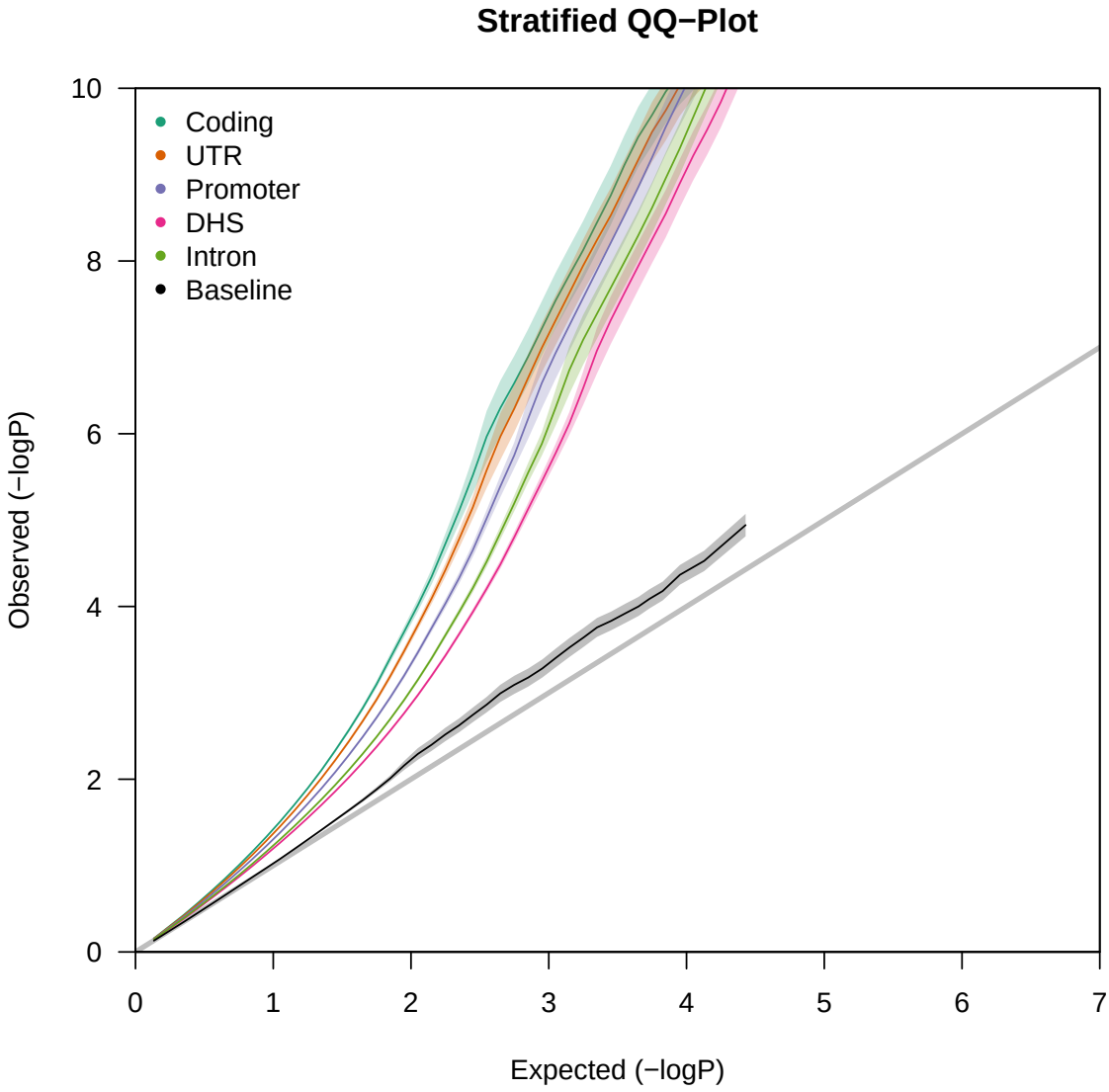

Figure S22. Stratified QQ-plot from realistic simulations. Stratified QQ-plots[2] display association statistics from variants in LD with each functional category in a probability plot, and assess significant enrichment of a given category visually or by a non-parametric test. The method accounts for LD by computing the sum of r^2 correlations between each GWAS SNP and all neighboring variants (within 1Mbp, including the SNP itself) belonging to a given functional category. A GWAS SNP is then considered part of a category if the corresponding score is ≥ 1 and QQ-lines are computed, separately, for each SNP in a category. We implemented this method as described in Schork et al.[2], using the European 1000 Genomes samples as the LD reference. As required, intergenic variants were defined as those having a score of zero to every other category, and we refer to them as "baseline" here to distinguish from the functionally intergenic category. Association statistics for each category were divided by the λ_{GC} observed in the baseline variants. Realistic traits were simulated in a 33,000 sample cohort with 8,300 causal SNPs where DHS and coding variants explaining 79% and 8% of h^2_g, respectively (no enrichment for other categories). Phenotypes and GWAS summary statistics were computed in a cohort of 32,000 samples. DHS appears to be the least enriched non-baseline category, while UTR, Promoter, and Intron appear falsely enriched due to LD to the truly causal Coding category. Shaded regions show standard error from 50 replicates. Under the null, the method correctly identified no enrichment for any disease architecture (Fig. S19). Under the causal category scenario, the stratified QQ-plots exhibited similar patterns of false-positive enrichment for correlated categories and false-negative estimates for DHS (Fig. S21, S23, S24). While the truly causal category generally had the highest deviation from the null in all instances except DHS-causal, it was not significantly distinguishable from the other truly null categories. Patterns were similar for the low-frequency architecture, with the DHS category further falsely depleted.

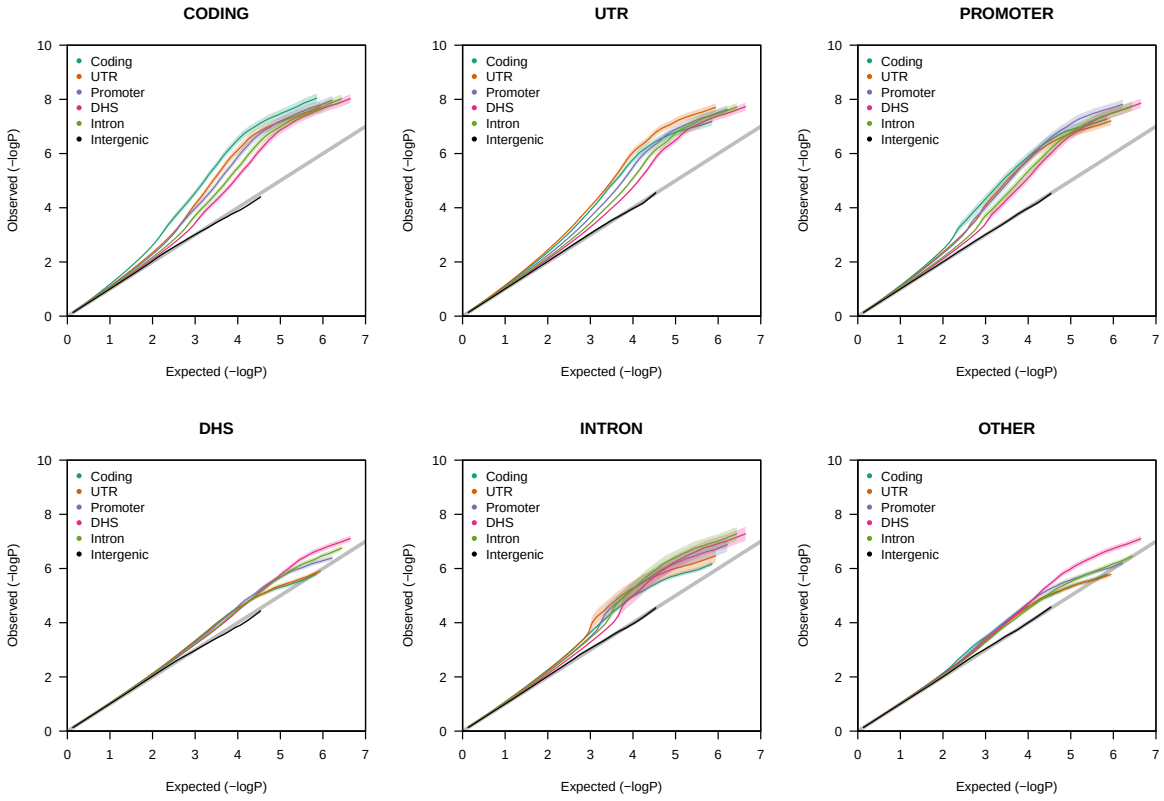

Figure S23. Stratified QQ-plots enrichment from simulated enrichment (MAF-independent). Each sub-figure shows stratified QQ-plot estimate when only the title category is causal. Non-intergenic categories appear falsely enriched in most instances. See Figure S22 for method and simulation details.

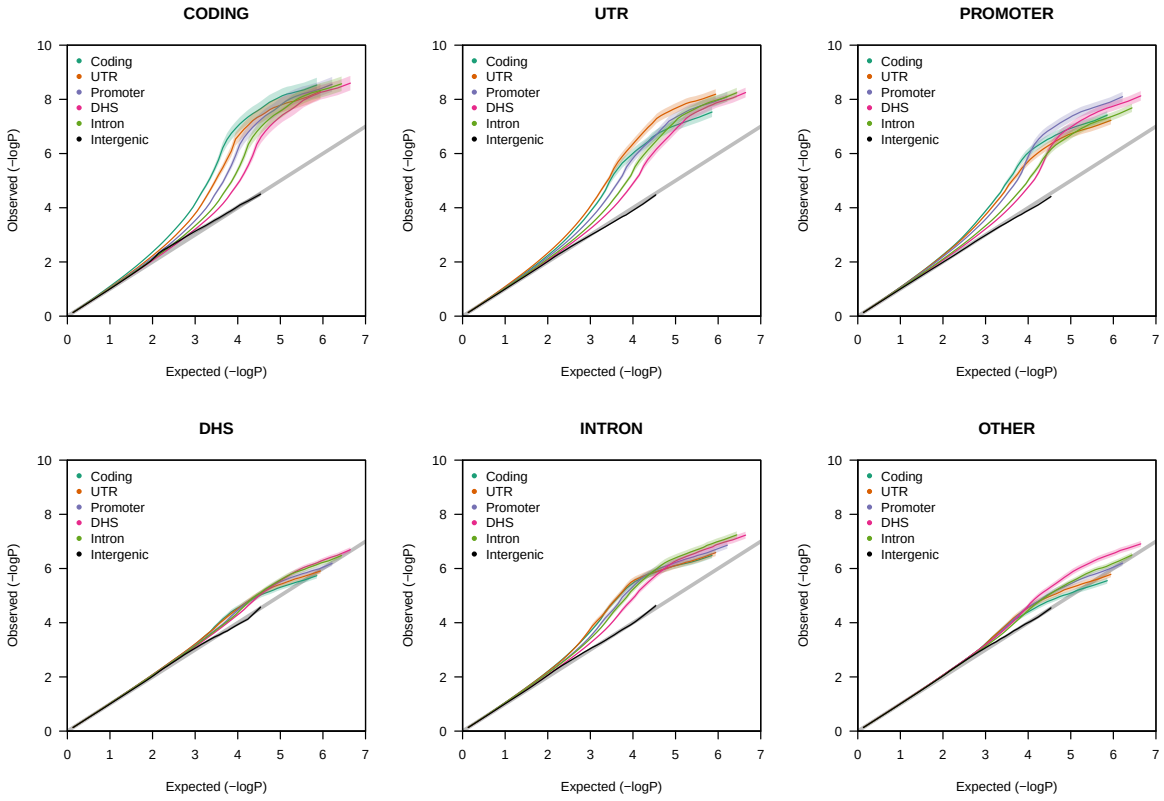

Figure S24. Stratified QQ-plots enrichment from simulated enrichment (low-frequency). Each sub-figure shows stratified QQ-plot estimate when only the title category is causal. Non-intergenic categories appear falsely enriched in most instances. See Figure S22 for method and simulation details.

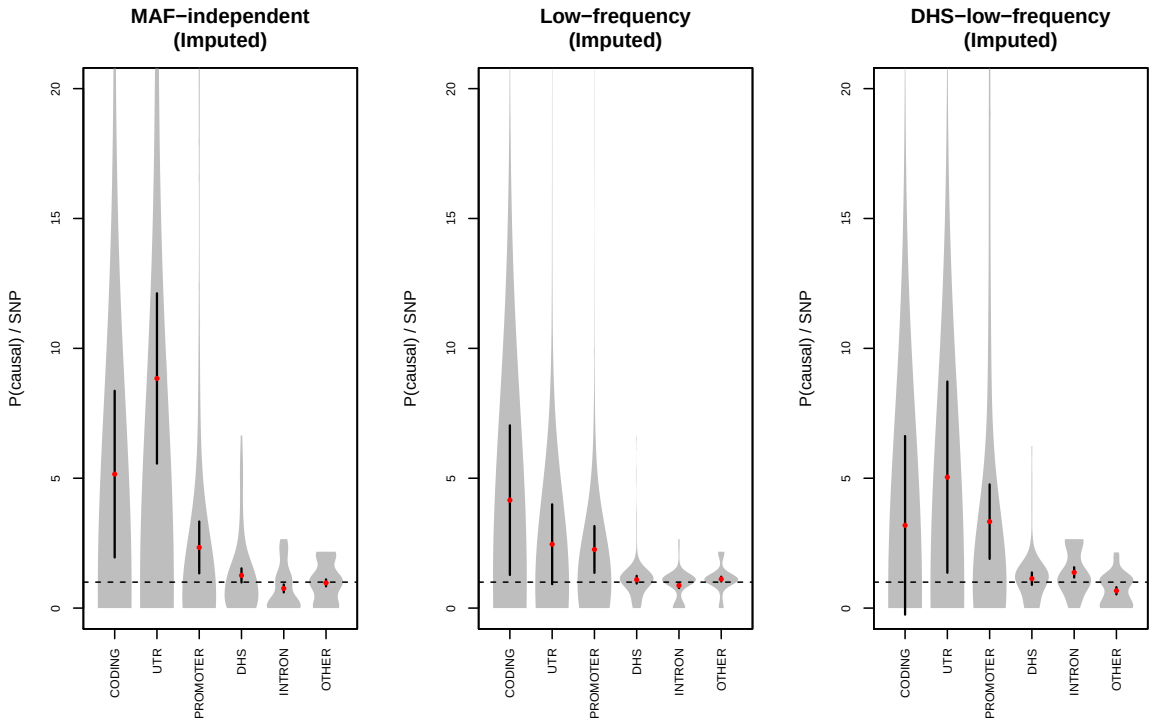

Figure S25. Enrichment from FGWAS under the null. Recently, Pickrell et al.[3] proposed a Bayesian hierarchical model that iteratively estimates category-specific enrichment priors and individual SNP association posteriors, implemented in the FGWAS software (see Web Resources). It's important to note that the main focus of this work was to identify functional enrichment at strongly associated variants, which is a fundamentally different question from enrichment in polygenic h^2_g. With this in mind, we provide comparisons to FGWAS here for completeness, and as a measure of the expected difference between polygenic architecture and the strong associations. We computed the estimates of P(causal), the probability that a SNP from the given annotation is causal, by running FGWAS in the "-print -p 0" mode to return individual SNP posteriors and then summing the posteriors over all SNPs in a given annotation divided by the sum of posteriors over all SNPs. We also report the odds-ratio estimate of enrichment inferred by FGWAS where appropriate. We ran FGWAS on our simulated phenotypes with window size set such that one causal variant is present in expectation (-k 500), to match the methodological assumptions. Due to computational constraints, FGWAS was only evaluated on chromosome 1 and any non-converging runs were excluded. In the null simulations, FGWAS was significantly upwardly biased for the smaller categories, perhaps due to the underlying metric having high variance and being restricted to the 0-1 scale or due to the large number of simulations not converging (Fig. S25A). The low-frequency and mixed architectures were generally similar to the MAF-independent architecture (Fig. S25B,C). For the causal simulations, the P(causal) at the true causal category was typically the highest but still underestimated by over 50%, with the larger non-causal categories also falsely identified as causal (Fig. S21, S26). This was most apparent when DHS is truly causal, with the intron and intergenic categories being indistinguishable from DHS. On the other hand, under the low-frequency architecture, estimates of P(causal) were not substantially different from the null; matching the overall category size regardless of true enrichment (Fig. S27). The FGWAS estimate has previously been shown to be unbiased when annotations are randomly sampled from the genome[4], and we suspect that the complex LD between contiguous categories results in the bias observed here. Each subplot shows estimates of enrichment from simulated phenotypes with no enrichment under different causal-variant architectures. Depending on the causality, smaller categories (Coding, UTR, Promoter) yield upward bias due to individual estimates being bounded to 0-1. The fraction of simulations that converged was 15%, 38%, and 16% respectively.

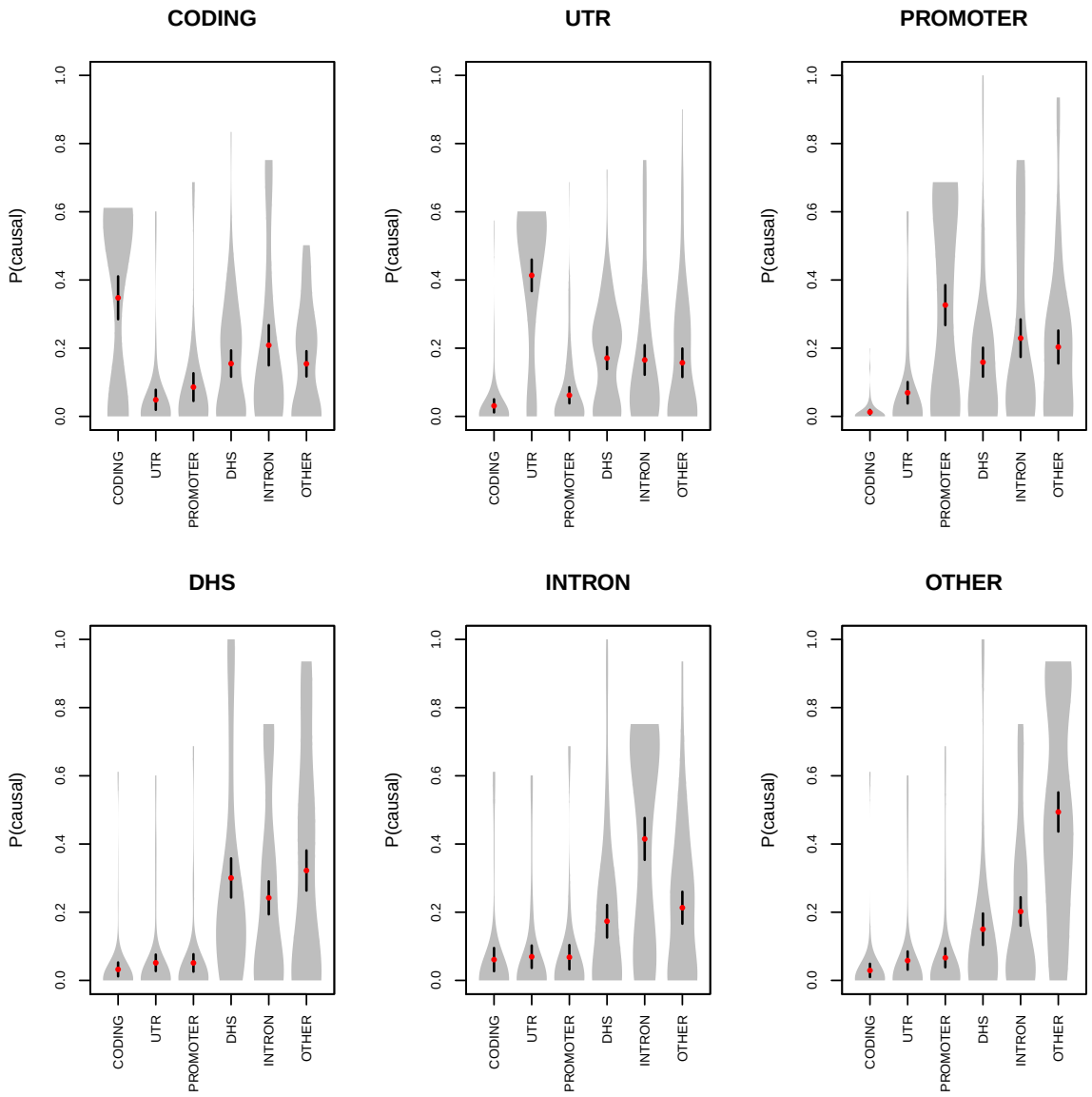

Figure S26. FGWAS estimates from simulated enrichment (MAF-independent). Each subfigure shows the FGWAS estimate whe only the title category is causal. Though the truly causal category is typically identified as most enriched, other categories (particularly the larger DHS/Intron/Intergenic) exhibit upward bias. The fraction of simulations that converged was 10% on average per category. See Figure S25 for method and simulation details.

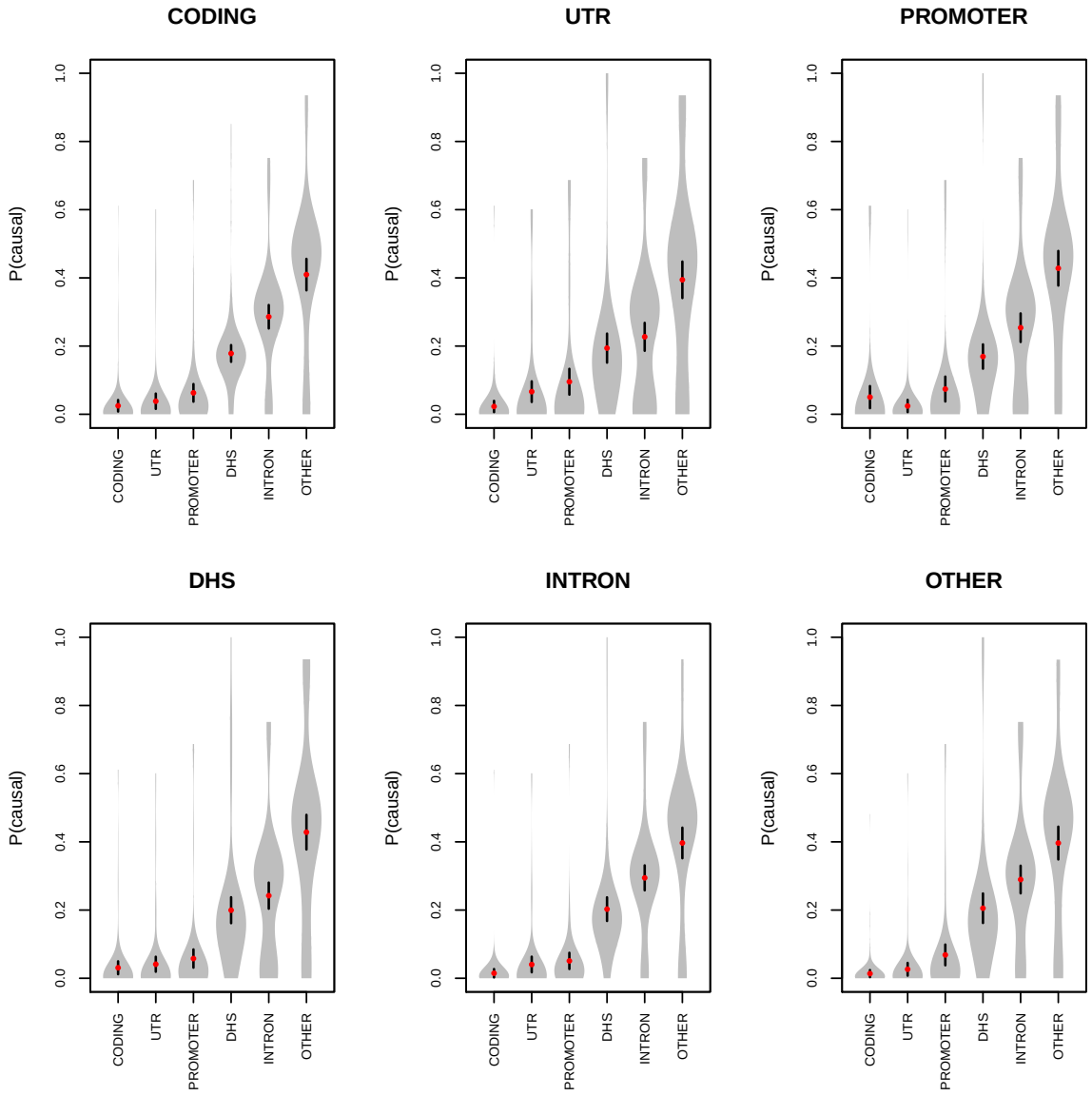

Figure S27. FGWAS estimates from simulated enrichment (low-frequency). Each sub-figure shows the FGWAS estimate whe only the title category is causal. Unlike the MAF-independent architecture, estimates of causality are not substantially different from the null category size. The fraction of simulations that converged was 30% on average per category. See Figure S25 for method and simulation details.

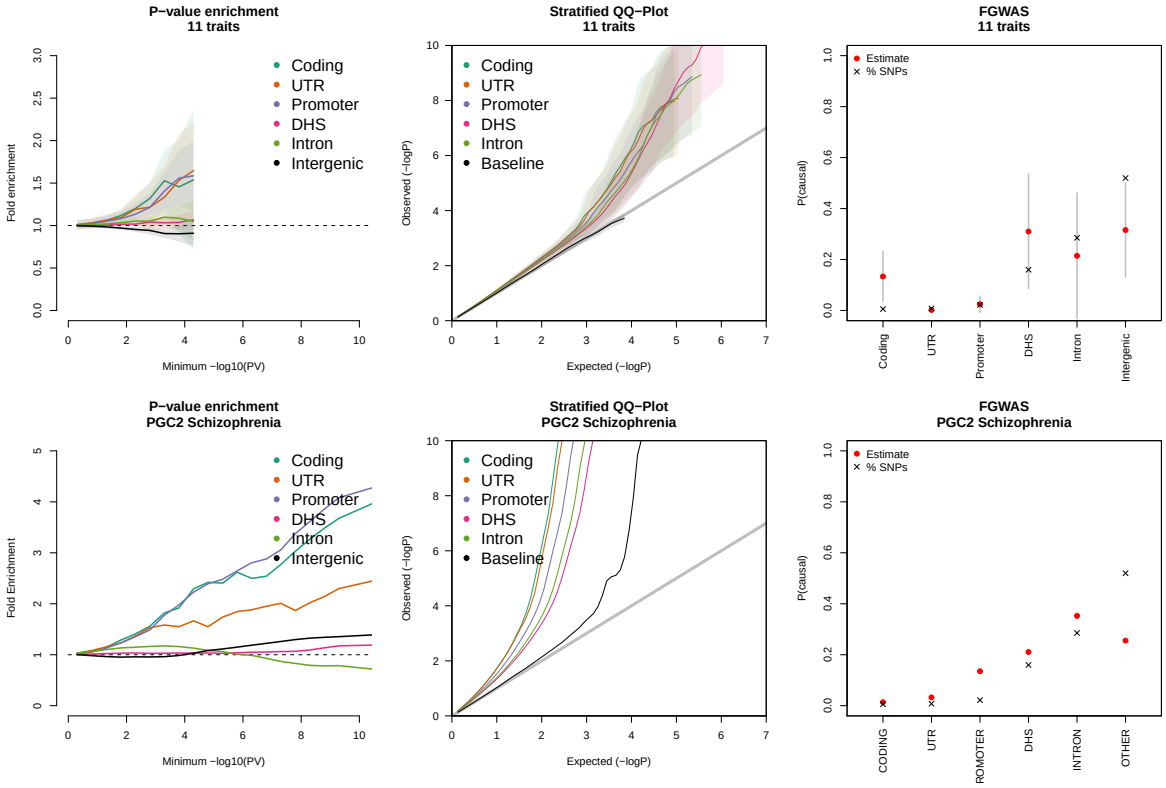

Figure S28. Enrichment from GWAS summary-statistics across 11 traits. Enrichment estimates from three GWAS-based methods are shown averaged over 11 traits (top) and for PGC2 Schizophrenia (bottom). Estimates for most enriched category were inconsistent across methods both in the WTCCC traits and in PGC. Shaded regions for P-value enrichment and QQ-plot, gray bars for FGWAS correspond to 1.96× standard error. FGWAS did not converge for BD, HT, and SP and they were excluded from the plot. We did not observe a clear consensus across the methods, with P-value enrichment showing promoter variants as significantly enriched; stratified QQ-plots showing significant enrichment in all non-intergenic categories; and FGWAS identifying coding variants as significantly enriched (Fig. 4A, S28). Likewise, in analyses of PGC2, none of the GWAS-based methods identified substantial enrichment at DHS variants (Fig. S28) nor did they agree on the most enriched category: promoter/coding for P-value enrichment; coding/UTR for stratified qq-plot; and promoter for FGWAS. This is consistent with our findings in realistic simulations, with stratified QQ-plots having similar results to P-value enrichment (Fig. 4B, S22, results from FGWAS were not shown due to lack of convergence).

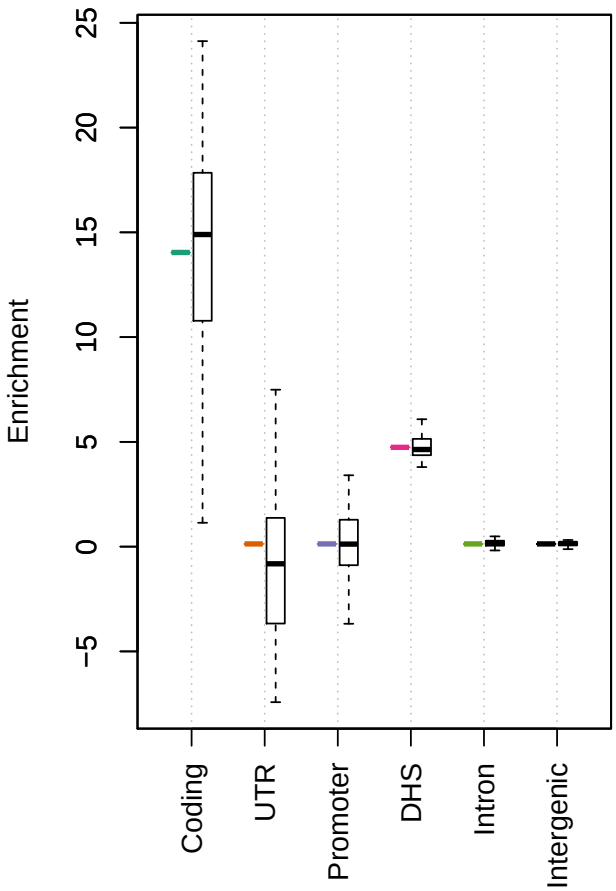

Figure S29. Partitioning of $h^2_{\mathbf{g}}$ with true DHS and coding enrichment. Inferred h^2_{g} enrichment from disease architecture mimicking observed DHS and coding enrichment in real data. Due to computational restrictions, enrichment was estimated from a random 15,000 samples of the 33,000 sample simulated GWAS cohort. Colored bars show the induced enrichment. Boxplots show the distribution of inferred enrichment over 50 trials.

Phenotype	Label	Prevalence	Cases	Controls	Genotyped SNPs	Imputed SNPs
WTCCC2:						
Schizophrenia	SP	0.010	2698	5458	394992	4345606
Ankylosing spondylitis	AS	0.003	1783	5239	408616	6162624
Multiple sclerosis	MS	0.001	9315	5211	396469	5795523
Ulcerative colitis	UC	0.001	2495	5428	447905	4620390
WTCCC:						
Bipolar disorder	BD	0.005	1550	2666	143054	4192374
Coronary artery disease	CAD	0.060	1746	2668	139567	4190156
Crohns disease	CD	0.001	1542	2662	146952	4199232
Hypertension	HT	0.050	1730	2669	139541	4185735
Rheumatoid arthritis	RA	0.005	1664	2664	143732	4190217
Type 1 diabetes	T1D	0.005	1746	2668	139206	4184291
Type 2 diabetes	T2D	0.080	1641	2671	142027	4195404
Other:						
Schizophrenia	PGC2	0.010	24926	33271	Varies	4-5 million
Schizophrenia	SWE ex-chip	0.010	5010	6197	238652	NA

Table S1. Datasets analyzed. Number of samples and markers for each dataset analyzed.

Cohort	Cases	Controls
boco	1754	2121
buls	192	595
denm	462	449
dubl	259	828
edin	363	281
ersw	260	311
lie5	486	383
mgs2	2583	2444
pewb	564	1779
ucla	688	598
clo3	2105	1975
cou3	530	678
egcu	234	1152
ersw	265	319
swe5	1764	2581
swe6	975	1145
umeb	341	577
umes	193	704
buls	195	608
butr	608	613
cims	67	65
clm2	3426	4085
lie2	133	269
msaf	325	139
pewb	574	1812
pews	150	236
aarh	876	871
boco	1773	2161
fii6	360	1082
gras	1067	1169
lacw	157	245
lie5	497	389
ucla	700	607

Table S2. PGC2 datasets analyzed.

Variant class	Homogenous	All
All coding	104,240	110,331
Singleton coding	19,860	19,329
Rare coding (MAF < 0.01, non-singleton)	64,040	70,569
Common coding (MAF ≥ 0.01)	20,340	20,433

Table S3. Summary of exome-chip data. Number of polymorphic variants by coding class and sub-cohort in the Swedish schizophrenia samples.

Category	Description	% physical	% 1000G	% array	% imputed
Coding (non-UTR)	Overlaps a coding exon.	1.1%	0.9%	0.9%	0.5%
UTR	Overlaps a 5' or 3' untranslated region.	1.0%	0.9%	1.0%	0.8%
Promoter	Within 2kbp of a transcription start site.	2.5%	2.6%	2.2%	2.2%
DHS	Overlaps DHS region observed in any cell-type.	14.6%	16.4%	23.3%	15.7%
Intron	Overlaps an intron.	29.1%	28.6%	26.8%	28.8%
Intergenic	All other intergenic variants.	51.8%	50.5%	44.8%	52.0%

Table S4. Coding and regulatory annotation categories. Description of functional categories and fraction occupied, respectively, by physical genome; all 1000 Genomes SNPs; average array SNPs; average imputed 1000G SNPs.

WTCCC1 Genotyped: Affymetrix				
Annotation	MAF	INFO	LD score	Cons
Coding	0.2330	NA	116.4	1.076
UTR	0.2388	NA	104.1	0.560
Promoter	0.2435	NA	118.6	0.231
DHS	0.2462	NA	92.6	0.346
Intron	0.2450	NA	111.0	0.177
Intergenic	0.2489	NA	116.9	0.135
WTCCC2 Imputed: Affymetrix				
Annotation	MAF	INFO	LD score	Cons
Coding	0.1700	0.9730	111.0	1.191
UTR	0.1773	0.9749	100.2	0.620
Promoter	0.1780	0.9739	114.5	0.266
DHS	0.1836	0.9775	89.0	0.388
Intron	0.1817	0.9776	108.2	0.194
Intergenic	0.1846	0.9773	111.6	0.148
WTCCC2 Imputed: Illumina				
Annotation	MAF	INFO	LD score	Cons
Coding	0.1672	0.9745	91.7	1.525
UTR	0.1735	0.9758	85.8	0.816
Promoter	0.1749	0.9751	97.1	0.358
DHS	0.1798	0.9778	79.2	0.498
Intron	0.1780	0.9791	94.7	0.254
Intergenic	0.1810	0.9780	101.3	0.188

Table S5. Functional category features. For each genotyping platform and functional category, the following features are reported: minor allele frequency (MAF), imputation quality (INFO), average number of LD partners (LD score), and GERP conservation score (Cons).

Annotation	Imputed % SNPs	Imputed % Effective SNPs	Genotyped % SNPs	Genotyped % Effective SNPs
Coding	0.5%	0.5%	1.0%	2.3%
UTR	0.8%	0.8%	1.0%	2.4%
Promoter	2.2%	2.0%	2.3%	4.8%
DHS	15.7%	18.9%	23.6%	33.6%
Intron	28.6%	29.8%	27.1%	23.2%
Other	52.2%	47.9%	45.0%	33.7%
DHS-Cell-Unique	4.0%	4.4%	NA	NA
DHS-DGF	7.8%	9.5%	NA	NA
DHS-Enhancer	3.2%	4.2%	NA	NA
DHS-Narrowed 1%	1.1%	1.3%	NA	NA
DHS-Narrowed 5%	5.2%	6.1%	NA	NA
DHS-Narrowed 10%	10.3%	12.1%	NA	NA

Table S6. Effective % of SNPs in analyzed categories. For each functional category analyzed, effective number of imputed SNPs was computed using LD in 1000G EUR samples[5,6]; defined as the number of SNPs divided by the average sum of r^2 between a SNP in the category and every other SNP in a 1Mbp window. Lower panel shows estimates from functional categories analyzed only in imputed data.

Table S7. Cell types analyzed. Excel spreadsheet detailing cell types and tissues used for cell-type specific DHS analysis.

				1x noise			2x noise
Category	% SNPs	% h^2_g (se)	SD	REML SE	% h^2_g (se)	SD	REML SE
CODING	0.6%	0.2% (0.5%)	5.5%	5.0%	0.4% (0.6%)	5.8%	5.2%
DHS	15.8%	16.0% (1.9%)	19.4%	17.9%	18.5% (1.7%)	17.2%	18.4%
PROMOTER	2.4%	1.4% (0.7%)	7.0%	6.5%	2.6% (0.7%)	6.9%	6.7%
UTR	0.9%	1.4% (0.5%)	4.8%	5.1%	1.5% (0.6%)	5.7%	5.3%
INTRON	29.1%	29.0% (1.1%)	10.8%	10.4%	28.7% (1.1%)	10.8%	10.7%
OTHER	51.3%	52.0% (1.1%)	11.4%	11.2%	48.3% (1.1%)	11.3%	11.4%

Table S8. Partitioned null h^2_g with simulated imputation noise. Null phenotypes were simulated from SNPs with realistic imputation noise (proportional to imputation INFO score) added and % h^2_g inferred using functional components without noise; corresponding to a scenario where genotypes are imputed with some inaccuracy. Under the assumption that INFO score is a reasonable proxy for imputation accuracy, substantial differences in imputation between categories would be expected to yield biased estimates. However, no significant deviations from the null were observed. As in previous simulations, a polygenic quantitative trait was constructed from 8,300 randomly selected causal variants for individuals in the WTCCC2:AS cohort. For each causal SNP s and corresponding INFO (imputation accuracy) score i_s, normally distributed noise was added to create a new SNP s' such that s had an R^2 of i_s with s'. A polygenic trait with $h^2_g = 0.50$ and no functional enrichment was then simulated from the noisy genotypes (identical to a model where phenotypes come from clean genotypes and the GRM is constructed from noisy ones). The "2x noise" column corresponds to a more extreme simulation where the added noise was double that observed in the real data (new $i_s = 1-2(1-i_s)$). The emperical standard deviation (SD) and the average analytical standard error (REML SE) is also shown for each scenario and do not deviate substantially. All estimates computed from 100 random simulations.

Category	1x noise % h^2_g (se)	2x noise % h^2_g (se)
CODING	99.5% (0.8%)	100.3% (0.9%)
DHS	95.7% (2.4%)	95.4% (2.4%)

Table S9. Partitioned causal h^2_g with simulated imputation noise. Simulations as described in Table S8 but with 100% h^2_g in the listed category.

Category	λ_{GC}	% h_g^2	se	P-value	adjusted se	adjusted P-value
Coding	1.26	7.5%	2.0%	4.74e-04	2.2%	1.83e-03
UTR	1.34	6.6%	2.0%	4.28e-03	2.4%	1.36e-02
Promoter	1.45	6.2%	2.6%	1.25e-01	3.1%	2.02e-01
DHS	1.32	79.5%	6.6%	3.64e-22	7.6%	3.74e-17
Intron	1.39	1.5%	3.9%	5.48e-12	4.7%	4.89e-09
Intergenic	1.70	-3.1%	4.0%	2.84e-42	5.3%	1.53e-25

Table S10. Meta-analysis adjusted for shared controls. We evaluated potential biases due to the use of shared controls by shifting the functional categories and performing the entire genotyped meta-analysis procedure to compute an empirical null distribution. Specifically, over 1,000 consecutive indices, we shifted all functional annotations ahead by 2MB (moving regions that crossed the chromosome boundary into the next chromosome) thereby preserving the total h_g^2, total sample relatedness, and relative dependence between categories but permuting any relationship to true function. For each shifted annotation, we re-computed GRMs from the genotyped data and estimated functional enrichment within each trait, as well as the meta-analysis value across all 11 traits, yielding $1,000 \times 6$ shifted meta-analysis estimates. We observed no enrichment or inflation of P-values within each study (Table S11), further supporting the robustness of the empirical standard error. We did observe inflation in the meta-analysis P-values ranging from λ_{GC} of 1.26 (coding) to 1.70 (intergenic). We adjusted the standard errors observed in real data by the corresponding $\sqrt{\lambda_{GC}}$, which yielded adjusted P-values that remained significant for all categories but UTR (Table S10). For each functional category, the empirical inflation of p-values due to shared controls (λ_{GC}) is reported. The raw meta-analysis estimate of h_g^2, standard error, and enrichment P-value is shown for imputed SNPs; followed by the corresponding λ_{GC} adjusted estimates.

Category	avg. enrichment	avg. Z-score
Coding	0.98	-0.04
UTR	1.01	-0.04
Promoter	1.04	-0.01
DHS	0.99	-0.02
Intron	1.00	0.00
Intergenic	1.00	0.00

Table S11. Estimates of enrichment from shifted regions. For each category, the average enrichment and Z-score observed in h^2_g estimates on real phenotypes and shifted functional annotations. Results averaged across 1,000 shifts and all traits.

	Genotyped			Imputed		
Annotation	% h_g^2 (s.e.)	Enrichment (s.e)	P-value	% h_g^2 (s.e.)	Enrichment (s.e.)	P-value
Coding	4% (1%)	4.12 (0.96)	1.1e-03	8% (2%)	13.84 (4.12)	1.8e-03
DHS	38% (4%)	1.63 (0.16)	1.0e-04	79% (8%)	5.07 (0.48)	3.7e-17
Promoter	5% (1%)	2.19 (0.62)	5.2e-02	6% (3%)	2.79 (1.41)	2.0e-01
UTR	4% (1%)	3.51 (0.95)	8.2e-03	7% (2%)	8.42 (3.01)	1.4e-02
Intron	23% (3%)	0.83 (0.11)	1.2e-01	2% (5%)	0.05 (0.16)	4.9e-9
Intergenic	25% (4%)	0.56 (0.08)	2.7e-08	-3% (5%)	-0.06 (0.10)	¡1e-20

Table S12. Components of heritability from regulatory elements in GWAS data (meta-analysis). Shared controls correction applied (see also Table S13, S10).

	Genotyped			Imputed		
Annotation	% h_g^2 (s.e.)	Enrichment (s.e)	P-value	% h_g^2 (s.e.)	Enrichment (s.e.)	P-value
Coding	4% (1%)	4.12 (0.85)	2.59e-04	8% (2%)	13.84 (3.67)	4.74e-04
DHS	38% (3%)	1.63 (0.14)	7.98e-06	79% (7%)	5.07 (0.42)	3.64e-22
Promoter	5% (1%)	2.19 (0.51)	1.94e-02	6% (3%)	2.79 (1.17)	1.25e-01
UTR	4% (1%)	3.51 (0.82)	2.21e-03	7% (2%)	8.42 (2.60)	4.28e-03
Intron	23% (2%)	0.83 (0.09)	6.40e-02	2% (4%)	0.05 (0.14)	5.48e-12
Intergenic	25% (3%)	0.56 (0.06)	4.11e-13	-3% (4%)	-0.06 (0.08)	2.84e-42

Table S13. Components of heritability from regulatory elements in GWAS data (meta-analysis). Meta-analysis estimates computed using inverse-variance weighting without shared controls correction.

Bipolar disorder (h^2 = 0.6-0.7)	Genotyped h_g^2 = 0.26 (0.032)				Imputed h_g^2 = 0.24 (0.035)			
Function	% h_g^2	(s.e.)	Enrichment	P-value	% h_g^2	(s.e.)	Enrichment	P-value
Coding	3.5%	(2.4%)	4.2	2.6e-01	4.9%	(7.2%)	9.0	5.5e-01
UTR	3.6%	(2.5%)	3.8	2.8e-01	11.6%	(7.6%)	15.3	1.5e-01
Promoter	-1.0%	(3.3%)	-0.5	3.4e-01	-11.0%	(9.4%)	-5.1	1.6e-01
DHS	34.0%	(10.3%)	1.4	3.1e-01	34.6%	(26.5%)	2.2	4.7e-01
Intron	22.9%	(8.0%)	0.9	6.2e-01	27.0%	(15.3%)	0.9	9.2e-01
Intergenic	37.0%	(8.9%)	0.8	3.3e-01	33.0%	(16.2%)	0.6	2.3e-01
Coronary artery disease (h^2 = 0.3-0.6)	Genotyped h_g^2 = 0.31 (0.057)				Imputed h_g^2 = 0.25 (0.062)			
Function	% h_g^2	(s.e.)	Enrichment	P-value	% h_g^2	(s.e.)	Enrichment	P-value
Coding	1.7%	(3.2%)	2.2	7.7e-01	7.5%	(12.5%)	14.0	5.7e-01
UTR	5.6%	(3.7%)	5.9	2.1e-01	10.5%	(13.0%)	13.8	4.6e-01
Promoter	4.5%	(5.1%)	2.0	6.6e-01	2.8%	(16.0%)	1.3	9.7e-01
DHS	41.1%	(15.4%)	1.8	2.5e-01	0.7%	(47.1%)	0.0	7.5e-01
Intron	24.5%	(12.0%)	0.9	8.4e-01	44.4%	(27.5%)	1.5	5.7e-01
Intergenic	22.6%	(13.8%)	0.5	9.3e-02	34.1%	(27.2%)	0.7	5.0e-01
Crohn's disease (h^2 = 0.6-0.8)	Genotyped h_g^2 = 0.18 (0.024)				Imputed h_g^2 = 0.17 (0.025)			
Function	% h_g^2	(s.e.)	Enrichment	P-value	% h_g^2	(s.e.)	Enrichment	P-value
Coding	3.7%	(2.5%)	4.6	2.5e-01	19.2%	(8.2%)	35.9	2.3e-02
UTR	-0.8%	(2.5%)	-0.8	4.8e-01	3.1%	(7.6%)	4.1	7.6e-01
Promoter	7.3%	(4.0%)	3.3	2.0e-01	-3.6%	(9.5%)	-1.7	5.4e-01
DHS	58.4%	(11.9%)	2.5	3.4e-03	151.7%	(27.1%)	9.7	5.2e-07
Intron	14.9%	(8.8%)	0.6	1.7e-01	-30.9%	(15.6%)	-1.1	1.3e-04
Intergenic	16.5%	(10.3%)	0.4	4.9e-03	-39.5%	(17.3%)	-0.8	1.1e-07
Hypertension (h^2 = 0.3-0.7)	Genotyped h_g^2 = 0.37 (0.053)				Imputed h_g^2 = 0.33 (0.059)			
Function	% h_g^2	(s.e.)	Enrichment	P-value	% h_g^2	(s.e.)	Enrichment	P-value
Coding	6.2%	(2.9%)	7.6	6.4e-02	25.4%	(10.5%)	47.2	1.8e-02
UTR	5.5%	(3.0%)	5.6	1.3e-01	12.7%	(9.5%)	16.7	2.1e-01
Promoter	4.9%	(4.2%)	2.2	5.1e-01	-3.0%	(11.7%)	-1.4	6.6e-01
DHS	28.3%	(11.7%)	1.2	6.8e-01	93.8%	(31.6%)	6.0	1.3e-02
Intron	19.4%	(9.4%)	0.7	4.2e-01	-32.4%	(18.6%)	-1.1	1.1e-03
Intergenic	35.6%	(10.5%)	0.8	3.4e-01	3.4%	(19.8%)	0.1	1.3e-02
Rheumatoid arthritis (h^2 = 0.6)	Genotyped h_g^2 = 0.11 (0.031)				Imputed h_g^2 = 0.09 (0.033)			
Function	% h_g^2	(s.e.)	Enrichment	P-value	% h_g^2	(s.e.)	Enrichment	P-value
Coding	-0.6%	(5.0%)	-0.8	7.7e-01	1.4%	(17.7%)	2.6	9.6e-01
UTR	7.1%	(5.7%)	7.3	2.9e-01	21.1%	(19.3%)	27.9	2.9e-01
Promoter	1.9%	(7.5%)	0.9	9.8e-01	28.3%	(24.6%)	13.2	2.9e-01
DHS	46.4%	(23.4%)	2.0	3.3e-01	162.7%	(67.4%)	10.4	2.9e-02
Intron	6.5%	(18.9%)	0.2	2.8e-01	-78.9%	(45.1%)	-2.8	1.7e-02
Intergenic	38.8%	(20.1%)	0.8	7.3e-01	-34.6%	(42.0%)	-0.7	3.8e-02
Type 1 diabetes (h^2 =0.9)	Genotyped h_g^2 = 0.13 (0.030)				Imputed h_g^2 = 0.13 (0.032)			

Function	% h^2_g	(s.e.)	Enrichment	P-value	% h^2_g	(s.e.)	Enrichment	P-value
Coding	7.9%	(4.6%)	9.5	1.3e-01	35.0%	(16.1%)	65.3	3.2e-02
UTR	5.1%	(4.4%)	5.2	3.5e-01	-1.8%	(12.9%)	-2.4	8.4e-01
Promoter	11.0%	(6.7%)	5.0	1.9e-01	28.8%	(18.3%)	13.5	1.4e-01
DHS	28.2%	(18.0%)	1.2	7.9e-01	106.2%	(42.5%)	6.8	3.3e-02
Intron	36.7%	(14.7%)	1.4	5.1e-01	-8.3%	(26.0%)	-0.3	1.5e-01
Intergenic	11.2%	(17.1%)	0.2	4.4e-02	-59.9%	(30.8%)	-1.1	2.7e-04
Type 2 diabetes (h^2 = 0.3-0.6)	Genotyped h^2_g = 0.37 (0.065)				Imputed h^2_g = 0.42 (0.070)			
Function	% h^2_g	(s.e.)	Enrichment	P-value	% h^2_g	(s.e.)	Enrichment	P-value
Coding	-2.0%	(3.0%)	-2.4	3.5e-01	2.5%	(8.1%)	4.7	8.1e-01
UTR	-0.7%	(3.2%)	-0.7	6.0e-01	8.7%	(8.5%)	11.4	3.5e-01
Promoter	-3.5%	(4.7%)	-1.6	2.3e-01	-3.3%	(10.3%)	-1.5	6.0e-01
DHS	69.3%	(16.0%)	3.0	4.0e-03	63.8%	(27.5%)	4.1	8.0e-02
Intron	26.1%	(11.4%)	1.0	9.4e-01	17.1%	(17.4%)	0.6	5.1e-01
Intergenic	10.7%	(13.6%)	0.2	1.0e-02	11.1%	(17.2%)	0.2	1.7e-02
Multiple sclerosis (h^2 = 0.3-0.8)	Genotyped h^2_g = 0.19 (0.009)				Imputed h^2_g = 0.17 (0.009)			
Function	% h^2_g	(s.e.)	Enrichment	P-value	% h^2_g	(s.e.)	Enrichment	P-value
Coding	6.4%	(1.7%)	3.6	6.2e-03	5.5%	(2.9%)	9.4	9.5e-02
UTR	3.9%	(1.4%)	3.2	6.3e-02	8.1%	(3.1%)	9.4	1.9e-02
Promoter	6.1%	(2.0%)	2.4	7.4e-02	11.7%	(4.0%)	5.0	1.8e-02
DHS	33.1%	(5.3%)	1.3	1.1e-01	77.7%	(9.4%)	4.9	5.5e-11
Intron	24.1%	(4.0%)	0.9	3.1e-01	1.5%	(5.7%)	0.1	9.1e-07
Intergenic	26.4%	(4.2%)	0.6	3.2e-04	-4.5%	(5.7%)	-0.1	1.0e-16
Ankylosing spondylitis ($h^2 > 0.90$)	Genotyped h^2_g = 0.18 (0.028)				Imputed h^2_g = 0.14 (0.027)			
Function	% h^2_g	(s.e.)	Enrichment	P-value	% h^2_g	(s.e.)	Enrichment	P-value
Coding	6.9%	(4.8%)	3.9	2.9e-01	1.5%	(10.4%)	2.6	9.3e-01
UTR	11.4%	(4.6%)	9.2	2.7e-02	20.9%	(12.0%)	24.5	9.6e-02
Promoter	5.2%	(5.7%)	2.0	6.5e-01	7.5%	(13.5%)	3.2	7.1e-01
DHS	41.8%	(16.1%)	1.7	2.8e-01	106.3%	(33.4%)	6.7	6.8e-03
Intron	14.9%	(11.7%)	0.5	2.6e-01	-23.6%	(20.6%)	-0.8	1.1e-02
Intergenic	19.8%	(12.7%)	0.5	8.6e-02	-12.6%	(20.2%)	-0.2	1.5e-03
Schizophrenia (h^2 = 0.7-0.8)	Genotyped h^2_g = 0.20 (0.025)				Imputed h^2_g = 0.18 (0.024)			
Function	% h^2_g	(s.e.)	Enrichment	P-value	% h^2_g	(s.e.)	Enrichment	P-value
Coding	1.9%	(2.9%)	2.6	6.8e-01	7.7%	(6.6%)	14.2	2.8e-01
UTR	2.5%	(3.1%)	2.6	6.2e-01	0.8%	(6.3%)	1.0	1.0e 00
Promoter	7.4%	(4.4%)	3.4	2.4e-01	-9.7%	(7.7%)	-4.2	1.2e-01
DHS	37.6%	(13.3%)	1.6	2.7e-01	44.4%	(22.8%)	2.8	2.1e-01
Intron	26.6%	(9.4%)	1.0	9.8e-01	37.3%	(14.0%)	1.3	5.3e-01
Intergenic	23.9%	(10.6%)	0.5	3.5e-02	19.6%	(14.1%)	0.4	2.2e-02
Ulcerative colitis (h^2 = 0.5)	Genotyped h^2_g = 0.17 (0.017)				Imputed h^2_g = 0.14 (0.016)			
Function	% h^2_g	(s.e.)	Enrichment	P-value	% h^2_g	(s.e.)	Enrichment	P-value

Coding	4.7%	(2.5%)	6.2	1.2e-01	7.6%	(5.9%)	14.5	2.3e-01
UTR	4.3%	(2.6%)	4.5	2.1e-01	-1.4%	(5.7%)	-1.7	7.0e-01
Promoter	8.7%	(3.7%)	4.0	7.3e-02	23.8%	(7.9%)	10.6	6.5e-03
DHS	43.3%	(10.9%)	1.9	6.5e-02	93.5%	(19.4%)	6.0	6.0e-05
Intron	21.2%	(7.6%)	0.8	4.5e-01	-5.6%	(11.5%)	-0.2	2.8e-03
Intergenic	17.9%	(8.7%)	0.4	1.1e-03	-18.0%	(12.4%)	-0.3	1.5e-08

Table S14. Components of heritability from regulatory elements in GWAS data. Family-based h^2 (from literature), total h^2_g, and function-specific h^2_g of liability is reported for eleven traits. Enrichment computed over the % of SNPs in each category and P-value computed from Z-score. For auto-immune traits (CD,RA,T1D,MS,AS,UC) the well-studied MHC locus was removed from analyses. By inverse-variance meta-analysis, the average total genotyped $h^2_g = 0.17$ (0.01) and imputed $h^2_g = 0.19$ (0.01) for a nominally significant difference of P=0.03. Using flat weights instead yielded % h^2_g DHS = 85% with standard deviation of 48% (corresponding to both trait and sampling variation) and root mean squared analytical standard error of 36% (corresponding to estimated sampling variation only), yielding a standard deviation of $\sqrt{0.48^2 - 0.36^2} = 32\%$ in the true unobserved values. All traits with % h^2_g estimates > 100% (CD, RA, T1D, AS) have compensatory components with negative estimates.

	Constrained			Standard		
Category	fraction h^2_g (se)	enrichment (se)	PV	fraction h^2_g (se)	enrichment (se)	PV
Coding	0.052 (0.019)	9.521 (3.418)	1.27e-02	0.075 (0.020)	13.838 (3.673)	4.74e-04
UTR	0.053 (0.019)	6.801 (2.443)	1.76e-02	0.066 (0.020)	8.417 (2.596)	4.28e-03
Promoter	0.069 (0.025)	3.126 (1.109)	5.52e-02	0.062 (0.026)	2.792 (1.168)	1.25e-01
DHS	0.710 (0.064)	4.532 (0.407)	3.82e-18	0.795 (0.066)	5.072 (0.421)	3.64e-22
Intron	0.061 (0.038)	0.211 (0.131)	1.74e-09	0.015 (0.039)	0.053 (0.137)	5.48e-12
Intergenic	0.046 (0.039)	0.088 (0.075)	9.68e-34	-0.031 (0.040)	-0.059 (0.078)	2.84e-42

Table S15. Constrained REML estimate of h^2_g. Comparison of constrained analysis (where components estimating h^2_g below zero are dropped from the analysis) and the standard un-constrained results. All values computed from meta-analysis over 11 traits. No shared-controls correction applied.

Enrichment	Coding	DHS	Promoter	UTR	Intron	Intergenic	Entropy (H)
Promoter (imputed)	0.00%	2.35%	0.14%	0.01%	7.94%	25.98%	0.65
Coding (imputed)	0.04%	2.28%	0.05%	0.01%	7.69%	25.17%	0.64
All categories (genotyped)	0.02%	6.02%	0.11%	0.03%	6.49%	13.03%	0.62
DHS (imputed)	0.00%	12.45%	0.01%	0.00%	2.02%	6.60%	0.52
All categories (imputed)	0.04%	12.45%	0.14%	0.05%	0.44%	-1.59%	0.30

Table S16. Theoretical entropy of functional partitions. Our estimates of the relative significance of different h^2_{g} enrichment scenarios were directly dependent on the standard error and overall sample size analyzed. Here, we consider an alternative figure of merit which relies only on the fraction of h^2_{g} in each category. We borrow from information theory the concept of entropy, which is a measure of uncertainty in the distribution of a random variable. Given $P(X_i)$, the probability mass function of a random variable, entropy can be quantified as $H = -\sum_{i=1}^{a} P(X_i)\log(P(X_i))$. Depending on the distribution and log-base, this is equivalent to the number of bits required to encode an observation, with higher entropy implying lower predictability. Applying this to functional categories, we define $P(X_i)$ as the joint probability that a SNP falls into the given category and is causal. Assuming that $\%h^2_{gi}$ corresponds to the probability of causality in category i, we compute $P(X_i) = \%\text{SNP}_i \times \%h^2_{gi}$. We then compute the entropy as outlined previously. Table S16 demonstrates the resulting entropy from multiple enrichment scenarios observed in the 11 traits, with entropy inversely correlated to the individual category significance. Highest entropy was computed for an enrichment scenario that only accounted for the (least significant) promoter category, and lowest entropy was observed for an enrichment scenario that accounted for all six categories. Interestingly, the six-category genotyped enrichment yielded higher entropy than a hypothetical DHS-only imputed enrichment. This formulation of "functional entropy" provides a standard metric for comparing real and hypothetical enrichment scenarios completely independent of sample size and data platform. Each row indicates a different enrichment scenarios observed in the 11 traits, with the rows listing individual annotations corresponding to an enrichment only at that category and no enrichmet in other categories. Each column then lists the probability of a SNP being causal (% h^2_{g} × % SNP for that category), as well as the resulting entropy computed as $H = -\sum p \times \log(p)$.

	A: 1000G imputed	B: Genome-wide significant		C: Known from NHGRI	
Category	% SNPs	% SNPs	(enrichment)	% SNPs	(enrichment[1])
Coding	0.5%	1.7%	(3.16)	8.5%	(16.98)
UTR	0.8%	1.7%	(2.20)	2.6%	(3.25)
Promoter	2.2%	2.3%	(1.04)	7.5%	(3.39)
DHS	15.7%	14.3%	(0.91)	27.4%	(1.74)
Intron	28.8%	30.7%	(1.07)	25.5%	(0.88)
Intergenic	52.0%	49.3%	(0.95)	28.6%	(0.55)

[1]Does not account for null distribution of NHGRI SNPs.

Table S17. Functional enrichment from GWAS hits. Fraction of SNPs partitioned into each category shown for (A) all 1000 Genomes imputed SNPs; (B) genome-wide significant imputed SNPs (single best association in 1MB locus); and (C) known associated SNPs from NHGRI catalog. Enrichments computed relative to 1000G imputed fractions, all values computed from union of 11 traits.

Annotation	% h_g^2 (s.e.)	Enrichment (s.e)	P-value
Coding	0.026 (0.014)	4.206 (2.192)	1.44e-01
UTR	0.075 (0.014)	8.934 (1.653)	1.59e-06
Promoter	0.040 (0.017)	1.814 (0.760)	2.84e-01
DHS	0.509 (0.047)	3.154 (0.291)	1.40e-13
Intron	0.193 (0.028)	0.651 (0.096)	2.82e-04
Intergenic	0.149 (0.029)	0.294 (0.057)	1.94e-35

Table S18. Components of heritability from regulatory elements in PGC2 schizophrenia.

Category	% category	% h_g^2	% DHS h_g^2 (se)	% DHS SNP	enrichment to DHS (se)	PV
DHS-Coding	27.4%	5% (1%)	5% (1%)	0.9%	5.30 (1.34)	1.35e-03
DHS-UTR	31.2%	5% (1%)	4% (1%)	1.4%	2.62 (0.93)	8.16e-02
DHS-Promoter	29.8%	9% (2%)	9% (2%)	3.9%	2.25 (0.47)	7.90e-03
DHS-Intron	NA	40% (3%)	35% (2%)	39.6%	0.87 (0.06)	3.74e-02
DHS-Intergenic	NA	51% (4%)	48% (2%)	54.2%	0.89 (0.04)	1.05e-02
non-DHS	NA	-16% (5%)	NA	NA	NA	NA

Table S19. Functional enrichment of main categories within DHS category. The extended DHS category was sub-partitioned into five annotations, and h_g^2 reported. % category reports the percent of main category covered by DHS. The remaining non-DHS category was significantly negative ($P = 0.002$), likely due to underestimating standard errors.

Tissue type	Cell type	Genotyped Autoimmune	Imputed Autoimmune	Genotyped Non-autoimmune
Blood	T Cell	5.8 (4.2e-06)	10.2 (1.3e-12)	2.1 (3.5e-01)
Blood	Leukemia Cells	3.5 (6.7e-06)	4.7 (5.3e-10)	1.0 (9.8e-01)
Blood	Lymphoblastoid Cell	3.3 (1.1e-05)	4.9 (5.4e-11)	1.0 (9.4e-01)
Fetal Kidney	Fetal Right Renal Pelvis	5.4 (1.4e-04)	8.2 (5.7e-08)	1.5 (7.4e-01)
Bone Marrow	Monocyte	4.1 (1.6e-04)	5.7 (2.2e-07)	1.3 (7.6e-01)
Blood	CD8 Primary Cell	3.0 (3.0e-04)	5.4 (1.8e-10)	1.0 (9.6e-01)
Fetal Thymus	Fetal Thymus Cell	2.6 (4.0e-04)	4.5 (3.2e-09)	0.8 (6.6e-01)
Blood	CD4 Primary Cell	2.3 (9.5e-04)	3.1 (6.5e-06)	0.9 (8.7e-01)
Blood	CD14 Primary Cell	2.7 (1.1e-03)	3.2 (2.8e-04)	1.3 (7.4e-01)
Liver	Hliver Cell	3.7 (1.4e-03)	6.4 (2.8e-08)	-0.3 (2.1e-01)
Fetal Kidney	Fetal Left Renal Cortex Cell	4.8 (2.0e-03)	8.3 (2.6e-07)	1.6 (7.1e-01)
Bone Marrow	Blast Cell	2.9 (2.1e-03)	5.1 (9.1e-09)	2.4 (9.0e-02)
Fetal Muscle	Fetal Back Muscle Cell	5.6 (6.5e-03)	9.5 (2.5e-08)	2.9 (2.3e-01)
Blood	CD34 Primary Cell	1.7 (9.9e-03)	1.7 (1.5e-02)	0.7 (3.9e-01)
Blood	CD34 Mobilized Primary Cell	0.1 (3.0e-02)	0.3 (1.2e-01)	1.2 (6.6e-01)
Bone Marrow	Erythroleukemic Cell	2.3 (3.8e-02)	3.2 (4.7e-04)	0.5 (5.4e-01)
Fetal Lung	Fetal Left Lung Cell	0.2 (3.9e-02)	-0.1 (1.6e-02)	1.1 (8.4e-01)
Blood	Lymphocyte	1.7 (4.0e-02)	2.4 (3.5e-04)	0.7 (4.7e-01)
Fetal Kidney	Fetal Left Renal Pelvis	2.3 (4.7e-02)	3.2 (3.7e-03)	0.9 (8.8e-01)
Fetal Large Intestine	Fetal Large Intestine Cell	1.6 (4.9e-02)	2.7 (2.2e-06)	0.9 (6.9e-01)

Table S20. Cell-type and phenotype specific DHS enrichment. Fold-enrichment of h_g^2 relative to SNPs reported for cell-types DHSs observed as significant in genotype data (without adjusting for 83 cell-types tested). Enrichment was measured in comparison to h_g^2 at DHS regions, accounting for the background DHS enrichment. Results shown separately from meta-analysis of 6 autoimmune traits and 5 non-autoimmune traits. No shared-controls adjustment applied.

Cell type	AS	CD	MS	RA	T1D	UC
Monocyte	9.3% (2.8)	9.2% (3.1)	14.8% (4.5)	20.7% (7.0)	11.6% (3.8)	13.8% (4.7)
Fetal Right Renal Pelvis	5.5% (2.7)	5.0% (3.3)	13.7% (7.0)	12.2% (7.8)	6.0% (3.9)	11.6% (7.6)
Lymphoblastoid Cell	35.1% (4.2)	15.2% (2.1)	35.8% (4.4)	28.3% (3.8)	12.5% (1.7)	29.9% (4.1)
CD8 Primary Cell	23.3% (3.1)	7.8% (1.2)	33.6% (4.5)	14.4% (2.2)	56.4% (8.7)	28.4% (4.4)
Fetal Thymus Cell	20.3% (2.1)	15.7% (1.8)	34.8% (3.6)	-2.3% (-0.3)	23.4% (2.8)	30.5% (3.5)
T Cell	19.2% (7.2)	11.3% (5.0)	15.9% (6.1)	7.9% (3.5)	NA	15.1% (7.0)
Leukemia Cells	10.5% (1.5)	19.3% (2.9)	29.8% (4.1)	19.9% (3.0)	30.2% (4.6)	29.5% (4.6)
Mean enrichment:	(3.4)	(2.8)	(4.9)	(3.9)	(4.2)	(5.1)

Table S21. Cell-type and phenotype specific DHS enrichment by trait. For cell-types reported as significant in Table 1, % h^2_g and fold-enrichment relative to DHS is shown for each autoimmune trait, estimated from genotyped data.

		Joint with main categories		
% genome	% h_g^2 univar	% h_g^2 (se)	enrichment (se)	P-value
1%	0.610	0.198 (0.040)	18.094 (3.655)	2.91e-06
5%	0.853	0.415 (0.070)	7.971 (1.335)	1.76e-07
10%	0.948	0.704 (0.073)	6.884 (0.717)	2.35e-16
16% (all DHS)	0.985	0.795 (0.066)	5.072 (0.421)	3.64e-22

Table S22. h_g^2 from narrowed DHS regions. DHS regions were narrowed (to the center of the region) to achieve set % of genome, and h_g^2 estimates are reported from a single DHS component (univar) as well as jointly with the five other main components. For comparison, a randomly sampled 16% of SNPs yielded an average % h_g^2 univar of 0.86.

Variant class (and SNPs in LD)	All h^2_g (se)	Hom. h^2_g (se)	Hom. h^2_{gLD} (se)
All	0.307 (0.027)	0.366 (0.038)	0.370 (0.040)
GWAS chip	0.273 (0.020)	0.314 (0.028)	0.317 (0.042)
Exome chip	0.116 (0.022)	0.157 (0.032)	0.158 (0.034)
Variant class (exclusive)	All h^2_g (se) P-value	Hom. h^2_g (se) P-value	Hom. h^2_{gLD} (se) P-value
GWAS chip	0.242 (0.020)	0.282 (0.029)	0.291 (0.028)
Exome chip	0.065 (0.021) 2.0×10^{-06}	0.084 (0.031) 2.0×10^{-03}	0.079 (0.034) 1.2×10^{-02}
Exome chip (rare)	0.014 (0.019) 2.1×10^{-01}	0.040 (0.028) 7.7×10^{-02}	0.037 (0.029) 1.0×10^{-01}
Exome chip (common)	0.051 (0.011) 5.2×10^{-07}	0.044 (0.015) 1.3×10^{-03}	0.042 (0.017) 7.7×10^{-03}

Table S23. Components of heritability of Schizophrenia from exome chip. Estimates of h^2_g are reported from variance components in the homogenous Swedish sub-population as well as all samples. Top panel shows estimates (without accounting for shared variance due to LD between classes) in All samples, homogenous Swedish sub-population, and LD-adjusted[1] estimates (h^2_{gLD}) from the homogenous Swedish sub-population. Bottom panel shows corresponding joint estimates accounting for shared variance due to LD. In bottom panel, P-values from a likelihood ratio test on the corresponding component are shown below each row.

A: GWAS chip + Exome chip					
Annotation	h^2_g (se)	% h^2_g (se)	% Non-coding SNPs	Enrichment (se)	P-value
Coding (common)	0.049 (0.015)	NA	NA	NA	NA
Coding (rare)	0.037 (0.028)	NA	NA	NA	NA
UTR	0.003 (0.007)	1.1% (2.4%)	1.9%	0.59 (1.24)	7.4e-01
Promoter	0.006 (0.008)	2.2% (3.0%)	3.0%	0.73 (1.00)	7.8e-01
DHS	0.114 (0.023)	41.2% (8.0%)	25.7%	1.60 (0.31)	5.3e-02
Intron	0.083 (0.019)	30.2% (6.4%)	26.2%	1.15 (0.25)	5.3e-01
Intergenic	0.070 (0.023)	25.3% (7.4%)	43.2%	0.59 (0.17)	1.6e-02
B: GWAS chip					
Annotation	h^2_g (se)	% h^2_g (se)	% SNPs	Enrichment (se)	P-value
Coding	0.014 (0.007)	4.3% (2.3%)	2.0%	2.15 (1.13)	3.1e-01
UTR	0.005 (0.007)	1.6% (2.1%)	1.9%	0.84 (1.12)	8.9e-01
Promoter	0.009 (0.008)	2.8% (2.6%)	2.9%	0.95 (0.90)	9.5e-01
DHS	0.118 (0.023)	37.8% (7.1%)	25.2%	1.50 (0.28)	7.3e-02
Intron	0.092 (0.019)	29.5% (5.8%)	25.7%	1.15 (0.22)	5.1e-01
Intergenic	0.075 (0.023)	24.0% (6.6%)	42.4%	0.57 (0.16)	5.7e-03
C: 1000G imputed + Exome chip					
Annotation	h^2_g (se)	% h^2_g (se)	% Non-coding SNPs	Enrichment (se)	P-value
Coding (common)	0.050 (0.016)	NA	NA	NA	NA
Coding (rare)	0.035 (0.028)	NA	NA	NA	NA
UTR	0.030 (0.016)	11.0% (5.9%)	0.8%	13.26 (7.08)	8.3e-02
Promoter	-0.017 (0.019)	-6.1% (7.0%)	2.3%	-2.63 (2.99)	2.3e-01
DHS	0.144 (0.059)	53.0% (20.4%)	16.9%	3.13 (1.21)	7.7e-02
Intron	0.044 (0.032)	16.0% (11.8%)	28.7%	0.56 (0.41)	2.8e-01
Intergenic	0.071 (0.035)	26.1% (12.6%)	51.2%	0.51 (0.25)	4.7e-02
D: 1000G imputed					
Annotation	h^2_g (se)	% h^2_g (se)	% SNPs	Enrichment (se)	P-value
Coding	0.056 (0.014)	17.6% (4.8%)	0.4%	45.89 (12.43)	3.1e-04
UTR	0.023 (0.015)	7.2% (4.7%)	0.8%	8.66 (5.68)	1.8e-01
Promoter	-0.024 (0.018)	-7.7% (5.7%)	2.3%	-3.31 (2.47)	8.2e-02
DHS	0.169 (0.055)	53.6% (16.4%)	16.9%	3.18 (0.97)	2.5e-02
Intron	0.025 (0.030)	7.8% (9.5%)	28.6%	0.27 (0.33)	2.8e-02
Intergenic	0.068 (0.032)	21.6% (10.1%)	51.0%	0.42 (0.20)	3.7e-03

Table S24. Components of heritability from regulatory elements in SWE-SCZ schizophrenia. Estimates are reported from the homogenous Swedish sub-population.

Coding $f_{\max}$	Homogenous	All
Singleton	0.000 (0.007)	0.000 (0.004)
0.001	0.000 (0.009)	0.000 (0.006)
0.005	0.000 (0.010)	0.004 (0.007)
0.010	0.000 (0.011)	0.006 (0.008)
0.050	0.025 (0.013)	0.031 (0.009)

Table S25. Collapsed-variant $\hat{h}^2_g$ of Schizophrenia from exome chip. For a given cohort, the variance of the heritability estimate tends to grow with the number of markers analyzed. Borrowing from gene-based burden association tests[7,8], we considered a strategy for reducing the variance of this estimate by collapsing rare variants in a gene into a single polymorphic site when computing the GRM. Over the full data-set, this procedure collapses the 60,000 effective SNPs into approximately 16,000 genes that contain polymorphic SNPs. This technique also has the benefit of incorporating singleton variants that violate the traditional variance-components model normality assumptions. However, as with burden-tests, the model assumes that all SNPs have identical normalized effect-sizes and will exhibit downwards bias when this assumption is violated. Formally, the method recodes each gene as a multi-allelic "pseudo-SNP" where samples that carry a minor allele below frequency threshold $f_{\max}$ are considered carriers of the pseudo-SNP allele equal to the number of such variants they carry. The pseudo-SNPs are then normalized to have mean=0 and variance=1 and a new GRM is computed over the normalized pseudo-SNPs as in the standard model. The corresponding measure of $h^2_{\text{g,collapsed}}$ is estimated from this collapsed variance-component, jointly with a single non-coding component, which fully accounts for the minimal tagging of h^2_{g} from non-coding regions by collapsed variants (Table S41). Our simulations show that disease architectures with $> 50\%$ non-causal (or non-deleterious) variants capture substantially less heritability as to make this approach underpowered compared to the standard model considering all SNPs (Table S42, S43). This table reports estimates of heritability from gene-based collapsed variants computed in two sub-groups of Swedish samples with increasing allele frequency thresholds. Analytical standard error reported in parenthesis.

Phenotype	Study	# SNPs	# Samples
RA	Stahl et al. 2010	2,556,272	25,708
T2D	Morris et al. 2012	2,473,442	149,821
CAD	Schunkert et al. 2011	2,420,361	22,233
SCZ	PGC2 2014	9,444,246	150,064

Table S26. GWAS summary statistics used for fine-mapping.

	# SNPs in critical set			P(causal in critical set)		
Sample size	No prior[1]	True prior[2]	Wrong prior[3]	No prior[1]	True prior[2]	Wrong prior[3]
4414	1132	309	310	0.95	0.94	0.84
8828	1024	279	280	0.97	0.96	0.87
13242	903	246	245	0.98	0.97	0.88
17656	747	202	201	0.98	0.97	0.89
22070	603	161	163	0.98	0.96	0.88
26484	485	128	125	0.97	0.96	0.88
30898	375	97	95	0.96	0.96	0.87
35312	276	70	72	0.96	0.94	0.86
39726	199	50	52	0.96	0.94	0.85
44140	145	38	39	0.95	0.93	0.83

For (Coding, UTR, Promoter, DHS, Intron, Intergenic) respectively the following models and priors were used:
[1]Trait = (13.8 , 8.4 , 2.8 , 5.1 , 0.05 , 0.001); prior = (1.0 , 1.0 , 1.0 , 1.0 , 1.0 , 1.0)
[2]Trait = prior = (13.8 , 8.4 , 2.8 , 5.1 , 0.05 , 0.001)
[3]Trait = (6.6 , 3.3 , 0.5 , 4.3 , 0.3 , 0.1); prior = (13.8 , 8.4 , 2.8 , 5.1 , 0.05 , 0.001)

Table S27. Simulated fine-mapping analyses and calibration. Loci harboring a single typed causal variant were simulated from imputed SNPs and evaluated for fine-mapping over increasing sample sizes. The 95% critical set of causal variants was then computed with and without SNP priors, with set size and fraction of instances where the causal variant is in the critical set reported. "No prior" corresponds to a generative model where enrichment matches mean estimate from imputed data in main text and no prior is used for fine-mapping. "True prior" corresponds to the same generative model and the true enrichment is used as prior for fine-mapping. "Wrong prior" corresponds to the same fine-mapping priors but true enrichment set to the boundary of the confidence interval reported in main text. Each value is the mean from 2,000 simulations.

Component	Univariate R^2	Step-wise R^2	Step-wise PV	Multivariate PV
DHS	0.055	0.055	4.24e-104	7.49e-12
Intron	0.034	0.056	1.60e-03	2.83e-04
Intergenic	0.031	0.059	8.50e-07	6.22e-07
UTR	0.021	0.062	1.05e-07	7.34e-07
Promoter	0.016	0.062	3.46e-01	2.42e-01
Coding	0.009	0.062	2.24e-01	2.24e-01

We computed the expected GBLUP prediction accuracy using the previously derived[9,10] relationship that M effective SNPs, N training samples, and h^2_g are expected to yield prediction $r^2 = (h^2_g h^2_g)/(h^2_g + M/N)$. We did not account for ascertainment because prediction was assessed by cross-validation. For the PGC analysis, the observed-scale $h^2_g = 0.49$, $N = 10000$ and we assumed $M = 60000$, which is expected to yield genome-wide $r^2 = 0.037$. Assuming independent variance components, we similarly estimated expected r^2 of the functionally stratified predictor by evaluating (jointly estimated) component-specific h^2_g directly in the data, estimating M from the fraction of SNPs in each component, and summing all of the functional expected r^2 to compute the genome-wide prediction. For the PGC analysis, this yielded an expected genome-wide $r^2 = 0.077$, or a 2.08× increase over the standard predictor.

Table S28. BLUP prediction accuracy in PGC. The h^2_g for a set of SNPs is an upper-bound on the prediction accuracy of a polygenic score constructed from those SNPs in unrelated samples[9–11]. To evaluate the impact of functional partitioning on risk prediction, we compared GBLUP[12,13] prediction accuracy using six jointly estimated functional components vs. a single genome-wide component in the phase 1 subset of the PGC schizophrenia data (11,000 samples, see Materials and Methods). BLUP coefficients were computed in GCTA[14] (see Web Resources) using the imputed data in a model with a single genome-wide component and a separate model with the six functional category components and converted into SNP effects. Risk scores were then computed from the SNPs and effects in each component. We assessed prediction accuracy using 10-fold cross-validation, where component-specific heritability and BLUP values were only estimated in the ∼10,000 training samples. To account for population structure we included 10 principal components as fixed-effects in training the BLUP. We also included the same number of PCs when evaluating the predicted phenotype in a logistic regression with the true phenotype, reporting the Nagelkerke pseudo-R^2 of each model minus that of the principal components. Results are reported in Table S28. In this table, prediction R^2 and significance is reported for GBLUPs estimated from six functional categories jointly in 10-fold cross-validation. Univariate R^2 column reports the accuracy of a 1-dof predictor from each of the component individually. Step-wise R^2 column reports the accuracy of a multiple-dof prediction with each component added as an additional predictor in turn. Step-wise PV column reports P-value from the newly added predictor. Multivariate PV reports P-value from each predictor in the final 6-dof prediction model. In all instances, principal components were included as additional fixed-effects and subtracted from prediction R^2. Of the six jointly estimated components, DHS yielded the highest individual R^2 (0.055) and coding yielded the lowest (0.009). A single degree of freedom GBLUP prediction from the sum of all six components yielded a highly significant R^2 of 0.061 ($P < 10^{-20}$). However, GBLUP prediction using a single component was only slightly less accurate, with $R^2 = 0.058$ ($P = 2.6 \times 10^{-7}$ for difference). On the observed-scale OLS R^2, this corresponds to a genome-wide $r^2 = 0.043$ and a stratified $r^2 = 0.046$. Though highly statistically significant, the observed-scale increase of 1.07× is substantially lower than the 2.08× that would be expected in the case of independent markers (see Foonote). This indicates that the assumption of component independence is strongly violated and significant enrichments in component h^2_g do not necessarily translate into increased prediction accuracy.

Genotyped:						
	causal MAF< 0.50		causal MAF< 0.05		causal $\text{MAF}_{\text{DHS}} < 0.05$	
Category	empirical sd	REML se	empirical sd	REML se	empirical sd	REML se
Coding	0.009	0.008	0.011	0.010	0.011	0.010
UTR	0.009	0.009	0.011	0.011	0.011	0.011
Promoter	0.015	0.015	0.017	0.016	0.016	0.016
DHS	0.059	0.053	0.050	0.051	0.051	0.051
Intron	0.047	0.047	0.041	0.042	0.042	0.042
Intergenic	0.058	0.058	0.050	0.050	0.051	0.050
Imputed:						
	causal MAF< 0.50		causal MAF< 0.05		causal $\text{MAF}_{\text{DHS}} < 0.05$	
Category	empirical sd	REML se	empirical sd	REML se	empirical sd	REML se
Coding	0.033	0.032	0.033	0.032	0.032	0.032
UTR	0.033	0.033	0.033	0.033	0.032	0.033
Promoter	0.042	0.041	0.042	0.041	0.044	0.041
DHS	0.124	0.124	0.125	0.124	0.126	0.124
Intron	0.069	0.068	0.069	0.068	0.067	0.069
Intergenic	0.077	0.075	0.073	0.075	0.077	0.075

Table S29. Empirical and analytical standard error of partitioned $h^2_{\mathbf{g}}$. Over 1,000 simulations for each of three disease architectures, the emperical standard deviation and average REML analytical standard error is reported for each functional category.

Phenotype	Coding	DHS	Promoter	UTR	Intron	Intergenic
SP (REML)	0.020 (0.029)	0.376 (0.133)	0.074 (0.044)	0.025 (0.031)	0.266 (0.094)	0.239 (0.106)
SP (jknife)	0.024 (0.027)	0.368 (0.168)	0.085 (0.047)	0.022 (0.019)	0.258 (0.143)	0.244 (0.088)
AS (REML)	0.069 (0.048)	0.418 (0.161)	0.052 (0.057)	0.114 (0.046)	0.149 (0.117)	0.198 (0.127)
AS (jknife)	0.086 (0.048)	0.419 (0.188)	0.040 (0.057)	0.102 (0.052)	0.115 (0.149)	0.238 (0.112)
MS (REML)	0.064 (0.017)	0.331 (0.053)	0.061 (0.020)	0.039 (0.014)	0.242 (0.040)	0.264 (0.042)
MS (jknife)	0.073 (0.023)	0.339 (0.076)	0.050 (0.021)	0.046 (0.012)	0.235 (0.034)	0.258 (0.068)
UC (REML)	0.047 (0.025)	0.433 (0.109)	0.087 (0.037)	0.043 (0.026)	0.212 (0.076)	0.179 (0.086)
UC (jknife)	0.045 (0.027)	0.425 (0.105)	0.085 (0.039)	0.045 (0.028)	0.223 (0.069)	0.177 (0.077)
BD (REML)	0.035 (0.024)	0.340 (0.103)	-0.010 (0.033)	0.036 (0.025)	0.229 (0.080)	0.370 (0.089)
BD (jknife)	0.030 (0.027)	0.321 (0.129)	-0.021 (0.035)	0.047 (0.022)	0.245 (0.088)	0.377 (0.093)
CAD (REML)	0.017 (0.032)	0.411 (0.154)	0.045 (0.052)	0.056 (0.037)	0.245 (0.120)	0.226 (0.138)
CAD (jknife)	0.018 (0.025)	0.432 (0.137)	0.048 (0.054)	0.058 (0.039)	0.225 (0.105)	0.220 (0.134)
CD (REML)	0.037 (0.025)	0.584 (0.119)	0.073 (0.040)	-0.008 (0.025)	0.149 (0.088)	0.165 (0.103)
CD (jknife)	0.036 (0.025)	0.619 (0.113)	0.071 (0.050)	0.005 (0.029)	0.134 (0.107)	0.133 (0.117)
HT (REML)	0.062 (0.029)	0.283 (0.117)	0.049 (0.042)	0.055 (0.030)	0.194 (0.094)	0.356 (0.105)
HT (jknife)	0.062 (0.027)	0.319 (0.110)	0.058 (0.052)	0.057 (0.026)	0.210 (0.083)	0.293 (0.116)
RA (REML)	-0.006 (0.049)	0.464 (0.234)	0.019 (0.076)	0.071 (0.057)	0.065 (0.189)	0.388 (0.201)
RA (jknife)	-0.017 (0.037)	0.444 (0.325)	-0.007 (0.085)	0.069 (0.063)	0.063 (0.193)	0.443 (0.241)
T1D (REML)	0.079 (0.046)	0.282 (0.180)	0.110 (0.067)	0.051 (0.044)	0.367 (0.147)	0.112 (0.171)
T1D (jknife)	0.077 (0.050)	0.301 (0.158)	0.114 (0.076)	0.061 (0.054)	0.361 (0.103)	0.088 (0.161)
T2D (REML)	-0.020 (0.030)	0.694 (0.160)	-0.035 (0.047)	-0.007 (0.032)	0.261 (0.114)	0.107 (0.136)
T2D (jknife)	-0.020 (0.041)	0.769 (0.146)	-0.030 (0.048)	-0.019 (0.040)	0.208 (0.160)	0.096 (0.180)
meta (REML)	0.040 (0.008)	0.384 (0.033)	0.050 (0.012)	0.035 (0.008)	0.226 (0.025)	0.250 (0.028)
	2.59e-04	7.98e-06	1.94e-02	2.21e-03	6.40e-02	4.11e-13
meta (jknife)	0.039 (0.009)	0.418 (0.038)	0.043 (0.013)	0.040 (0.008)	0.228 (0.024)	0.238 (0.032)
	1.10e-03	1.50e-06	1.07e-01	6.45e-05	6.72e-02	1.78e-11

Table S30. Comparison of analytical and jack-knife % $h^2_{\mathbf{g}}$ from genotyped SNPs. For each trait and functional category, the % h^2_{g} and standard error (in parentheses) is shown from a the standard REML and a weighted block-jackknife dropping each chromosome in turn. Results from meta-analysis for each method shown at the bottom, with P-values for enrichment below each entry.

Phenotype	Coding	DHS	Promoter	UTR	Intron	Intergenic
BD (REML)	0.049 (0.072)	0.346 (0.265)	-0.110 (0.093)	0.116 (0.076)	0.270 (0.153)	0.330 (0.162)
BD (jknife)	0.029 (0.079)	0.244 (0.294)	-0.110 (0.119)	0.155 (0.086)	0.338 (0.173)	0.345 (0.172)
CAD (REML)	0.075 (0.125)	0.007 (0.468)	0.028 (0.160)	0.105 (0.130)	0.444 (0.275)	0.341 (0.272)
CAD (jknife)	0.052 (0.113)	0.058 (0.598)	0.024 (0.163)	0.125 (0.103)	0.449 (0.352)	0.301 (0.307)
CD (REML)	0.192 (0.082)	1.517 (0.271)	-0.036 (0.095)	0.031 (0.076)	-0.309 (0.156)	-0.395 (0.173)
CD (jknife)	0.201 (0.090)	1.506 (0.367)	-0.042 (0.106)	0.044 (0.088)	-0.289 (0.196)	-0.417 (0.176)
HT (REML)	0.255 (0.105)	0.938 (0.316)	-0.030 (0.118)	0.127 (0.095)	-0.324 (0.186)	0.034 (0.198)
HT (jknife)	0.253 (0.096)	0.902 (0.431)	-0.020 (0.181)	0.122 (0.086)	-0.266 (0.169)	0.008 (0.261)
RA (REML)	0.014 (0.176)	1.627 (0.674)	0.283 (0.246)	0.212 (0.193)	-0.789 (0.451)	-0.346 (0.420)
RA (jknife)	0.026 (0.213)	1.592 (0.952)	0.285 (0.218)	0.250 (0.204)	-0.826 (0.347)	-0.340 (0.460)
T1D (REML)	0.350 (0.161)	1.062 (0.425)	0.288 (0.183)	-0.018 (0.129)	-0.083 (0.260)	-0.599 (0.308)
T1D (jknife)	0.370 (0.165)	0.992 (0.509)	0.290 (0.194)	0.004 (0.121)	-0.126 (0.314)	-0.528 (0.287)
T2D (REML)	0.025 (0.081)	0.638 (0.275)	-0.033 (0.102)	0.087 (0.085)	0.171 (0.174)	0.111 (0.172)
T2D (jknife)	0.022 (0.063)	0.668 (0.164)	-0.048 (0.064)	0.084 (0.094)	0.165 (0.186)	0.110 (0.095)
SP (REML)	0.077 (0.066)	0.443 (0.228)	-0.097 (0.077)	0.008 (0.063)	0.373 (0.140)	0.196 (0.141)
SP (jknife)	0.077 (0.067)	0.401 (0.186)	-0.063 (0.090)	0.007 (0.052)	0.373 (0.130)	0.206 (0.134)
MS (REML)	0.055 (0.029)	0.777 (0.094)	0.117 (0.040)	0.080 (0.031)	0.015 (0.057)	-0.045 (0.057)
MS (jknife)	0.058 (0.026)	0.782 (0.144)	0.115 (0.052)	0.089 (0.039)	0.002 (0.080)	-0.048 (0.071)
AS (REML)	0.015 (0.104)	1.063 (0.334)	0.075 (0.135)	0.209 (0.120)	-0.236 (0.206)	-0.126 (0.202)
AS (jknife)	0.019 (0.107)	1.065 (0.437)	0.073 (0.118)	0.193 (0.178)	-0.232 (0.265)	-0.120 (0.235)
UC (REML)	0.076 (0.059)	0.935 (0.194)	0.238 (0.079)	-0.014 (0.057)	-0.056 (0.115)	-0.180 (0.124)
UC (jknife)	0.079 (0.058)	0.897 (0.235)	0.250 (0.100)	-0.043 (0.052)	-0.023 (0.123)	-0.161 (0.157)
meta (REML)	0.075 (0.020)	0.795 (0.066)	0.062 (0.026)	0.066 (0.020)	0.015 (0.039)	-0.031 (0.040)
	4.74e-04	3.64e-22	1.25e-01	4.28e-03	5.48e-12	2.84e-42
meta (jknife)	0.073 (0.019)	0.710 (0.077)	0.047 (0.029)	0.057 (0.022)	0.028 (0.047)	-0.002 (0.043)
	3.35e-04	5.45e-13	3.98e-01	2.25e-02	4.17e-08	3.18e-33

Table S31. Comparison of analytical and jack-knife % $h^2_{\mathbf{g}}$ from imputed SNPs. For each trait and functional category, the % h^2_{g} and standard error (in parentheses) is shown from a the standard REML and a weighted block-jackknife dropping each chromosome in turn. Results from meta-analysis for each method shown at the bottom, with P-values for enrichment below each entry.

		No fixed-effects		PCs as fixed-effects	
Phenotype	Prevalence	REML (se)	Regression (se)	REML (se)	Regression (se)
BD	0.005	0.31 (0.033)	0.40 (0.034)	0.24 (0.035)	0.24 (0.035)
CAD	0.060	0.27 (0.061)	0.28 (0.059)	0.25 (0.062)	0.22 (0.059)
CD	0.001	0.18 (0.025)	0.22 (0.025)	0.17 (0.025)	0.20 (0.025)
HT	0.050	0.58 (0.097)	0.59 (0.093)	0.55 (0.098)	0.50 (0.093)
RA	0.005	0.10 (0.033)	0.11 (0.032)	0.09 (0.033)	0.08 (0.032)
T1D	0.005	0.14 (0.032)	0.15 (0.031)	0.13 (0.032)	0.13 (0.032)
T2D	0.080	0.50 (0.068)	0.62 (0.067)	0.42 (0.070)	0.43 (0.067)
SP	0.010	0.75 (0.013)	10.00 (0.021)	0.18 (0.024)	0.25 (0.055)
MS	0.001	0.29 (0.007)	2.91 (0.008)	0.17 (0.009)	0.21 (0.013)
AS	0.003	0.15 (0.027)	0.16 (0.026)	0.14 (0.027)	0.14 (0.026)
UC	0.001	0.15 (0.016)	0.15 (0.015)	0.14 (0.016)	0.14 (0.015)

Table S32. Total liability-scale $h^2_{\mathbf{g}}$ from four inference methods. For each trait, the total estimate of h^2_{g} is shown from the standard REML method and Haseman-Elston regression with and without included fixed-effects. Estimates were transformed to liability-scale using the given prevalence.

Phenotype	Prevalence	No fixed-effects		PCs as fixed-effects	
		REML	Regression	REML	Regression
BD	0.005	0.48 (0.20)	0.63 (0.14)	0.35 (0.27)	0.43 (0.24)
CAD	0.060	-0.08 (0.44)	-0.05 (0.39)	0.01 (0.47)	-0.10 (0.49)
CD	0.001	1.46 (0.26)	1.49 (0.21)	1.52 (0.27)	1.58 (0.24)
HT	0.050	0.91 (0.29)	1.06 (0.26)	0.94 (0.32)	1.12 (0.31)
RA	0.005	1.37 (0.57)	1.38 (0.52)	1.63 (0.67)	1.76 (0.75)
T1D	0.005	1.21 (0.40)	1.35 (0.36)	1.06 (0.43)	1.27 (0.43)
T2D	0.080	0.70 (0.24)	0.70 (0.18)	0.64 (0.28)	0.52 (0.26)
SP	0.010	0.56 (0.06)	0.75 (0.00)	0.44 (0.23)	0.09 (0.39)
MS	0.001	0.72 (0.06)	0.79 (0.00)	0.78 (0.09)	0.91 (0.11)
AS	0.003	1.09 (0.31)	1.09 (0.28)	1.06 (0.33)	1.07 (0.33)
UC	0.001	0.91 (0.18)	1.00 (0.16)	0.94 (0.19)	1.03 (0.18)

Table S33. Fraction of DHS $h^2_{\mathbf{g}}$ from four inference methods. For each trait, the DHS estimate of % h^2_{g} is shown from the standard REML method and Haseman-Elston regression with and without included fixed-effects.

Annotation	Regression % h_g^2	REML % h_g^2 (s.e.)
Coding	10%	8% (2%)
DHS	90%	79% (7%)
Promoter	5%	6% (3%)
UTR	8%	7% (2%)
Intron	-4%	2% (4%)
Intergenic	-9%	-3% (4%)

Table S34. Regression and variance-component estimates of functional enrichment. The meta-analyzed estimate of % h_g^2 is shown for analyses using regression and variance-components (REML). No shared-control adjustment was performed.

Category	% SNPs	polygenic % h^2_g (se 100 trials)	DHS high-effect % h^2_g (se 400 trials)
CODING	0.8%	0.7% (0.9%)	1.6% (0.4%)
UTR	1.1%	1.2% (0.9%)	0.7% (0.5%)
PROMOTER	2.8%	2.7% (1.1%)	3.5% (0.6%)
DHS	16.7%	15.6% (2.7%)	17.1% (1.8%)
INTRON	31.1%	30.2% (2.0%)	29.5% (1.1%)
OTHER	47.5%	49.6% (1.7%)	47.7% (1.1%)

Table S35. Partitioned h^2_g with simulated case-control ascertainment. We simulated case-control ascertainment under two disease architectures and estimated % h^2_g to asess ascertainment induced biases. Phenotypes were simulated on imputed chr1 SNPs (10% of genome) of the 33,000 sample combined WTCCC2 cohort, using 830 causal variants with no functional enrichment and $h^2_g = 0.50$. "Polygenic" columns present results from simulation with randomly selected causal variants. "DHS high-effect" columns present results from simulation with only 16 causal DHS variants (each explaining 1% of the h^2_g), and 814 randomly selected non-DHS causal variants. Neither disease architecture lead to significant deviations from null enrichment. Ascertainment was induced by setting the top 1% of phenotypes to be cases (327 samples) and randomly selecting 654 non-cases to be controls, yielding a trait with 1% prevalance and 1:2 case:control ascertainment. Category-specific GRMs were then constructed for each ascertained cohort and h^2_g was evaluated on the liability scale. Restricting to chromosome 1 resulted in an M/N equal to that of a ∼10,000 sample cohort (where M is the effective number of SNPs, and N is the number of samples).

		polygenic
Category	empirical sd	REML se
CODING	8.0%	7.1%
UTR	8.0%	8.4%
PROMOTER	10.5%	10.0%
DHS	25.1%	27.3%
INTRON	18.4%	16.3%
OTHER	15.7%	17.6%

Table S36. Empirical and analytical standard error of partitioned $h^2_{\mathbf{g}}$ with case-control ascertainment. The emperical standard deviation and REML analytical standard-error shown for estimates of % h^2_{g} for a simulated 1:2 case:control ascertained trait with prevalance of 1% (see Table S35 for simulation details). Under this quasi-polygenic architecture with 830 causal variants, the analytical SE is 0.2% higher on average. Estimates shown over 100 random simulations.

Joint GRM:	h^2_g (se)
known,non-coding + non-coding	0.018 (0.004)
known,non-coding + non-coding	0.287 (0.028)
known,coding + known,non-coding + non-coding	0.006 (0.004)
known,coding + known,non-coding + non-coding	0.018 (0.004)
known,coding + known,non-coding + non-coding	0.286 (0.028)

Table S37. Components of heritability for known Schizophrenia loci. h^2_g for multiple joint estimates at known schizophrenia loci are reported for the underlined component in the homogenous Swedish cohort.

Joint GRM	$\hat{h}^2_g$ (se)	$\hat{h}^2_{\text{gLD}}$ (se)
Common coding	10.7% (0.7%)	11.8% (0.9%)
Rare coding (non-singleton)	1.7% (1.6%)	0.7% (2.0%)
Joint GRM + non-coding	$\hat{h}^2_g$ (se)	$\hat{h}^2_{\text{gLD}}$ (se)
Common coding	-1.2% (0.7%)	-1.1% (0.9%)
Rare coding (non-singleton)	-0.6% (1.7%)	-2.3% (2.1%)

Table S38. Fraction of simulated common non-coding heritability inferred by coding variants. Another potential source of confounding when estimating exome h^2_{g} is heritability from nearby non-coding variants that is tagged by exonic variants due to LD. Because our interest is in identifying the purely exonic contribution to phenotype, we consider the heritability from these non-coding variants to "contaminate" our estimates. Using the GWAS chip data from this cohort allows us to quantify the amount of contamination expected due to common non-coding SNPs. We simulated a standard polygenic phenotype with $h^2 = 0.50$ coming exclusively from 5,000 randomly selected GWAS chip non-coding SNPs and then inferred h^2_{g} using variance-components constructed from coding SNPs. No coding SNPs were used to generate the phenotypes, and if no contamination was present we expect the inferred h^2_{g} to equal zero.Bottom panel shows results when a third variance-component corresponding to non-coding variants is estimated jointly in the model. Values reported represent the fraction of simulated heritability inferred averaged over 50 trials (with standard error in parenthesis). We found that all coding variants together accounted for an average of 17.4% of the non-coding heritability (Table S38), significantly different from zero. This further broke down to slight but non-significant contamination of 2.7% at rare coding variants (MAF < 0.01) and a highly significant average of 11.8% from common coding variants (MAF ≥ 0.01), consistent with common variants being generally better tags of nearby common variation. Given the small physical size of the exome, contamination of 11.8% of the non-coding heritability could substantially bias the estimates from coding variants when estimated directly from exome chip data. To account for this contamination, we model an additional component consisting of the non-coding GWAS variants. When we conditioned in this way and estimate using a three variance-component model, we see statistically zero heritability attributed to the rare and common coding components. Because we only have genome-wide GWAS chip data available, which does not include rare variants and these variants are notoriously difficult to impute, the non-coding component is unlikely to account for contamination from rare non-coding variants. However, these variants would need to be physically close and in similar frequency to be strongly tagged by the rare coding variants we examined.

	Causal variants	
GRM genotypes	Rare coding	Common coding
non-coding	0.051 (0.012)	0.426 (0.006)
rare coding	0.509 (0.011)	0.043 (0.015)
common coding	0.024 (0.003)	0.514 (0.008)
	Causal Variants	
Joint GRM genotypes	Rare coding	Common coding
rare coding + non-coding	0.486 (0.003)	0.002 (0.001)
common coding + non-coding	0.025 (0.002)	0.485 (0.003)
rare coding + common coding	0.486 (0.004)	0.001 (0.001)
rare coding + common coding	0.000 (0.001)	0.482 (0.004)

Table S39. $\hat{h}_g^2$ of phenotypes simulated from coding variants. We set out to estimate the fraction of exome h^2 that is tagged by non-coding SNPs from the GWAS chip and 1,000 Genomes imputation. We simulate two groups of standard additive phenotypes from the rare and common exome variants, respectively, and infer $h^2_{\text{g,non-coding}}$ of these phenotypes from the non-coding SNPs. $\hat{h}_g^2$ inferred from different classes of GRMs is shown, with standard error over 10 trials in parenthesis. Lower panel shows results from multiple GRMs fit jointly, with bolded GRM corresponding to the reported variance-component estimate. The ratio of $\hat{h}^2_{\text{g,non-coding}}$ to simulated $h^2_{\text{g,exome}}$ gives us an estimate of the fraction of exome heritability tagged by non-coding variants. In 10 simulations from chromosome 22 with $h^2_{\text{g,exome}} = 0.5$ the average ratio is 0.85 for common coding variants and 0.11 for rare coding variants (Table S39). However, the tagging between components is fully accounted for by a joint, three component model (Table S40).

Simulated h^2_{g}			Jointly inferred $\hat{h}^2_g$ (se)		
rare coding	common coding	all non-coding	rare coding	common coding	all non-coding
0.25	0.25	0.25	0.247 (0.003)	0.262 (0.002)	0.256 (0.003)

Table S40. Joint $h^2_{\mathbf{g}}$ from simulated phenotype in Swedish schizophrenia cohort.

	$f_{\max}$				
GRM	Singleton	0.001	0.005	0.010	0.050
Collapsed	-0.009 (0.002)	-0.002 (0.002)	-0.002 (0.003)	-0.000 (0.003)	0.009 (0.004)
Collapsed + non-coding	-0.007 (0.002)	-0.004 (0.002)	-0.004 (0.003)	-0.004 (0.003)	0.001 (0.003)

Table S41. Collapsed $\hat{h}_g^2$ of phenotypes simulated from non-coding variants. An infinitesimal trait with $h_g^2 = 0.50$ was simulated from non-coding variants and $\hat{h}_g^2$ was inferred from coding variants collapsed below designated minor allele frequency $f_{\max}$. Mean and standard error are reported over 50 random trials. See Table S25 for method details.

		Fraction causal			
$f_{\max}$	Effect distribution	100%	50%	10%	1%
0.001	Uniform	0.49 (0.002)	0.33 (0.003)	0.21 (0.002)	0.17 (0.005)
0.001	Allelic	0.39 (0.003)	0.28 (0.003)	0.20 (0.003)	0.17 (0.007)
0.001	Normalized	0.33 (0.002)	0.22 (0.003)	0.16 (0.002)	0.16 (0.006)
0.005	Uniform	0.47 (0.002)	0.33 (0.005)	0.22 (0.002)	0.19 (0.006)
0.005	Allelic	0.37 (0.003)	0.28 (0.004)	0.21 (0.004)	0.18 (0.007)
0.005	Normalized	0.28 (0.003)	0.19 (0.004)	0.14 (0.002)	0.14 (0.006)
0.010	Uniform	0.47 (0.003)	0.34 (0.005)	0.24 (0.002)	0.20 (0.006)
0.010	Allelic	0.38 (0.002)	0.29 (0.006)	0.22 (0.003)	0.20 (0.007)
0.010	Normalized	0.24 (0.004)	0.17 (0.003)	0.13 (0.002)	0.15 (0.007)
0.050	Uniform	0.42 (0.003)	0.35 (0.006)	0.27 (0.003)	0.23 (0.008)
0.050	Allelic	0.35 (0.003)	0.30 (0.006)	0.28 (0.003)	0.23 (0.010)
0.050	Normalized	0.22 (0.003)	0.16 (0.005)	0.12 (0.002)	0.14 (0.006)

Table S42. Collapsed $\hat{h}_g^2$ of phenotypes simulated from rare coding variants. A quasi-infinitesimal trait was simulated from specified exome-wide causal fraction of coding variants and varying $f_{\max}$ and total $h_g^2 = 0.5$. Effect-sizes were sampled from a standard normal distribution on the normalized-variant scale or the allelic-variant scale, and forced to be uni-directional within each gene. The collapsed $\hat{h}_g^2$ was then estimated from coding variants at the given $f_{\max}$. No more than half of the true h_{g}^2 can be recovered from collapsing under any disease architecture. See Table S25 for method details.

		Fraction causal			
$f_{\max}$	Effect distribution	100%	50%	10%	1%
0.001	Uniform	1.51	0.80	0.40	0.31
0.001	Allelic	1.00	0.63	0.38	0.31
0.001	Normalized	0.77	0.44	0.27	0.26
0.005	Uniform	1.54	0.92	0.49	0.42
0.005	Allelic	1.12	0.70	0.46	0.39
0.005	Normalized	0.72	0.41	0.24	0.26
0.010	Uniform	1.58	0.97	0.56	0.45
0.010	Allelic	1.14	0.76	0.51	0.46
0.010	Normalized	0.57	0.34	0.22	0.29
0.050	Uniform	1.31	0.96	0.72	0.59
0.050	Allelic	0.97	0.80	0.73	0.62
0.050	Normalized	0.48	0.31	0.20	0.26

Table S43. Power of collapsed vs. non-collapsed $\hat{h}_g^2$ for rare coding variants. The ratio of LRT statistics from collapsed / non-collapsed SNPs is reported for simulations with rare coding variants. Values < 1 indicate greater power for direct (rather than collapsed) estimates. See Table S25 for method details.

References

[1] Gusev, A., Bhatia, G., Zaitlen, N., Vilhjalmsson, B., Diogo, D., Stahl, E., Gregersen, P., Worthington, J., Klareskog, L., Raychaudhuri, S., et al. (2013). Quantifying missing heritability at known gwas loci. PLoS Genet *9*, e1003993.

[2] Schork, A. J., Thompson, W. K., Pham, P., Torkamani, A., Roddey, J. C., Sullivan, P. F., Kelsoe, J. R., O'Donovan, M. C., Furberg, H., Schork, N. J., et al. (2013). All snps are not created equal: Genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated snps. PLoS Genet *9*, e1003449.

[3] Pickrell, J. K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. The American Journal of Human Genetics *94*, 559–573.

[4] Pickrell, J. K. (2014). Are parameter estimates from fgwas unbiased?

[5] Bulik-Sullivan, B., Loh, P.-R., Finucane, H., Ripke, S., Yang, J., Psychiatric Genomics Consortium, S. W. G., Patterson, N., Daly, M. J., Price, A. L., and Neale, B. M. (2014). Ld score regression distinguishes confounding from polygenicity in genome-wide association studies. bioRxiv.

[6] Yang, J., Zaitlen, N., Goddard, M., Visscher, P., and Price, A. (2014). Advantages and pitfalls in the application of mixed-model association methods. Nat Genet *46*, 100–106.

[7] Li, B. and Leal, S. M. (2008). Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. Am J Hum Genet *83*, 311–321.

[8] Price, A. L., Kryukov, G. V., de Bakker, P. I. W., Purcell, S. M., Staples, J., Wei, L.-J., and Sunyaev, S. R. (2010). Pooled association tests for rare variants in exon-resequencing studies. Am J Hum Genet *86*, 832–838.

[9] Daetwyler, H. D., Villanueva, B., and Woolliams, J. A. (2008). Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLoS One *3*, e3395.

[10] Lee, S. H. and Wray, N. R. (2013). Novel genetic analysis for case-control genome-wide association studies: Quantification of power and genomic prediction accuracy. PloS one *8*, e71494.

[11] Wray, N. R., Yang, J., Hayes, B. J., Price, A. L., Goddard, M. E., and Visscher, P. M. (2013). Pitfalls of predicting complex traits from snps. Nat Rev Genet *14*, 507–515.

[12] Habier, D., Fernando, R. L., and Garrick, D. J. (2013). Genomic blup decoded: a look into the black box of genomic prediction. Genetics *194*, 597–607.

[13] Speed, D. and Balding, D. J. (2014). Multiblup: improved snp-based prediction for complex traits. Genome Res.

[14] Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). Gcta: a tool for genome-wide complex trait analysis. Am J Hum Genet *88*, 76–82.