# SNP-to-gene linking strategies reveal contributions of enhancer-related and candidate master-regulator genes to autoimmune disease

**Kushal K. Dey**[1,8,*], **Steven Gazal**[1], **Bryce van de Geijn**[1,3], **Samuel Sungil Kim**[1,4], **Joseph Nasser**[7], **Jesse M. Engreitz**[5,6,7], **Alkes L. Price**[1,2,7]

[1]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA

[2]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA

[3]Genentech, South San Francisco, CA 94080, USA

[4]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[5]Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA

[6]BASE Initiative, Betty Irene Moore Children's Heart Center, Lucile Packard Children's Hospital, Stanford University School of Medicine, Stanford, CA 94304, USA

[7]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

[8]Lead contact

## SUMMARY

We assess contributions to autoimmune disease of genes whose regulation is driven by enhancer regions (enhancer-related) and genes that regulate other genes in *trans* (candidate master-regulator). We link these genes to SNPs using several SNP-to-gene (S2G) strategies and apply heritability analyses to draw three conclusions about 11 autoimmune/blood-related diseases/traits. First, several characterizations of enhancer-related genes using functional genomics data are informative for autoimmune disease heritability after conditioning on a broad set of regulatory annotations. Second, candidate master-regulator genes defined using *trans*-eQTL in blood are also conditionally informative for autoimmune disease heritability. Third, integrating enhancer-related and master-regulator gene sets with protein-protein interaction (PPI) network information magnified their disease signal. The resulting PPI-enhancer gene score produced >2-fold stronger

heritability signal and >2-fold stronger enrichment for drug targets, compared with the recently proposed enhancer domain score. In each case, functionally informed S2G strategies produced 4.1- to 13-fold stronger disease signals than conventional window-based strategies.

## Graphical abstract
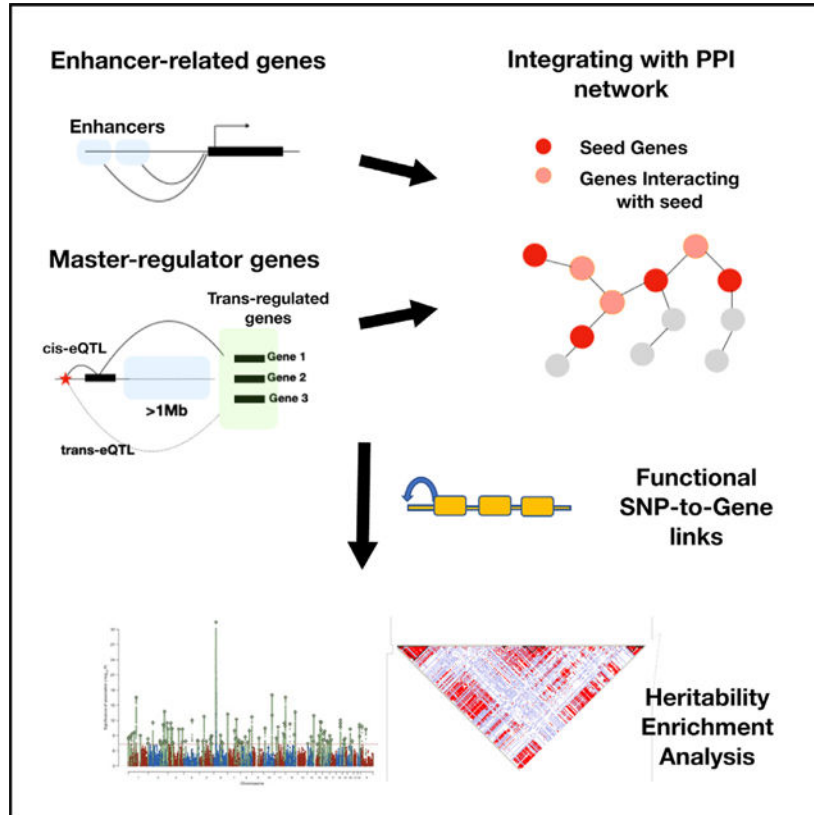


## In brief

Disease risk variants associated with complex traits and diseases predominantly lie in non-coding regulatory regions of the genes, motivating the need to assess the relative importance of genes for disease through the lens of gene regulation. Here, Dey et al. assess contributions to autoimmune disease of enhancer-related genes and candidate master-regulator genes in blood using SNP-to-gene linking strategies.

## INTRODUCTION

Disease risk variants associated with complex traits and diseases predominantly lie in non-coding regulatory regions of the genes, motivating the need to assess the relative importance of genes for disease through the lens of gene regulation.[1–6] Several recent studies have performed disease-specific gene-level prioritization by integrating genome-wide association study (GWAS) summary statistics data with functional genomics data, including gene expression and gene networks.[7–14] Here, we investigate the contribution to autoimmune disease of gene sets reflecting two specific aspects of gene regulation in blood—genes

with strong evidence of enhancer-related regulation and candidate master-regulator genes that may potentially regulate many other genes. Previous studies suggested that both of these characterizations are important for understanding human disease.[9,15–24] For example, several common non-coding variants associated with Hirschsprung disease have been identified in intronic enhancer elements of *RET* gene and have been shown to synergistically regulate its expression,[25,26] and *NLRC5* acts as a master-regulator of MHC class genes in immune response.[27] Our two main goals are to characterize which types of genes are important for autoimmune disease and to construct SNP annotations derived from those genes that are conditionally informative for disease heritability, conditional on all other annotations.

A major challenge in gene-level analyses of disease is to link genes to SNPs that may regulate them, a prerequisite to integrative analyses of GWAS summary statistics. Previous studies have often employed window-based strategies such as ±100 kb,[8,9,11] linking each gene to all SNPs within 100 kb; however, this approach lacks specificity. Here, we incorporated functionally informed SNP-to-gene (S2G) linking strategies that capture both distal and proximal components of gene regulation. We evaluated the resulting SNP annotations by applying stratified linkage disequilibrium (LD) score regression[28] (S-LDSC) conditional on a broad set of coding, conserved, regulatory, and LD-related annotations from the baseline-LD model,[29,30] meta-analyzing the results across 11 autoimmune diseases and blood cell traits. We focused on autoimmune diseases and blood cell traits because the functional data underlying the gene scores and S2G strategies that we analyze is primarily measured in blood. We also assessed gene-level enrichment for disease-related gene sets, including approved drug targets for autoimmune disease.[10]

# RESULTS

## Overview of methods

We define an annotation as an assignment of a numeric value to each SNP with minor allele count 5 in a 1000 Genomes Project European reference panel,[31] as in our previous work;[28] we primarily focus on annotations with values between 0 and 1. We define a gene score as an assignment of a numeric value between 0 and 1 to each gene; gene scores predict the relevance of each gene to disease. We primarily focus on binary gene sets defined by the top 10% of genes; we made this choice to be consistent with Finucane et al.,[9] and to ensure that all resulting SNP annotations (gene scores × S2G strategies; see below) were of reasonable size (0.2% of SNPs or larger). We consider 11 gene scores prioritizing enhancer-related genes, candidate master-regulator genes, and genes with high network connectivity to enhancer-related or candidate master-regulator genes (Table 1 and Figure S1); these gene scores were only mildly correlated (average r = 0.08, Figure S2). We considered enhancer-related and candidate master-regulator genes because previous studies suggested that both of these characterizations are important for understanding human disease.[9,15–24]

We define an S2G linking strategy as an assignment of 0, 1, or more linked genes to each SNP. We consider ten S2G strategies capturing both distal and proximal gene regulation (see STAR Methods, Figure 1A, and Table 2); these S2G strategies aim to link SNPs to genes that they regulate. For each gene score X and S2G strategy Y, we define a corresponding

combined annotation X × Y by assigning to each SNP the maximum gene score among genes linked to that SNP (or 0 for SNPs with no linked genes); this generalizes the standard approach of constructing annotations from gene scores using window-based strategies.[8,9] For example, enhancer domain score (EDS)-binary × Activity-by-Contact (ABC) annotates SNPs linked by ABC enhancer-gene links[32,39] to any gene from the EDS-binary gene set, whereas EDS-binary × 100 kb annotates all SNPs within 100 kb of any gene from the EDS-binary gene set. For each S2G strategy, we also define a corresponding binary S2G annotation defined by SNPs linked to the set of all genes. We have publicly released all gene scores, S2G links, and annotations analyzed in this study (see URLs).

We assessed the informativeness of the resulting annotations for disease heritability by applying S-LDSC[28] to 11 independent blood-related traits (six autoimmune diseases and five blood cell traits; average $N_{case}$ = 13K for autoimmune diseases and N = 443K for blood cell traits, Table S1) and meta-analyzing S-LDSC results across traits; we also assessed results meta-analyzed across autoimmune diseases or blood cell traits only, as well as results for individual diseases/traits. We conditioned on 86 coding, conserved, regulatory, and LD-related annotations from the baseline-LD model (v2.1)[29,30] (see URLs). S-LDSC uses two metrics to evaluate informativeness for disease heritability: enrichment and standardized effect size (τ*). Enrichment is defined as the proportion of heritability explained by SNPs in an annotation divided by the proportion of SNPs in the annotation,[28] and generalizes to annotations with values between 0 and 1.[34] Standardized effect size (τ*) is defined as the proportionate change in per-SNP heritability associated with a 1 SD increase in the value of the annotation, conditional on other annotations included in the model.[29]

As a preliminary assessment of the potential of the ten S2G strategies, we considered the ten S2G annotations defined by SNPs linked to the set of all genes. The S2G annotations were only weakly positively correlated (average r = 0.09; Figure S3). We analyzed the ten S2G annotations via a marginal analysis, running S-LDSC[28] conditional on the baseline-LD model and meta-analyzing the results across the 11 blood-related traits. In the marginal analysis, all ten S2G annotations were significantly enriched for disease heritability, with larger enrichments for smaller annotations (Figure 2A and Table S2); values of standardized enrichment (defined as enrichment scaled by the SD of the annotation[11]) were more similar across annotations (Figure S4 and Table S3). Seven S2G annotations attained conditionally significant τ* values after Bonferroni correction (p < 0.05/10) (Figure 2B and Table S2). In the joint analysis, three of these seven S2G annotations were jointly significant: transcription start site (TSS) (joint τ* = 0.97), Roadmap (joint τ* = 0.84), and ABC (joint τ* = 0.44) (Figure 2B and Table S4). This suggests that these three S2G annotations are highly informative for disease. Subsequent analyses were conditioned on the baseline-LD+ model defined by 86 baseline-LD model annotations plus all S2G annotations (except Coding, TSS, and Promoter, which were already part of the baseline-LD model), to ensure that conditionally significant τ* values for (gene scores × S2G strategies) annotations are specific to the gene scores and cannot be explained by (all genes × S2G strategies) annotations. Accordingly, we confirmed that (random genes × S2G strategies) annotations did not produce conditionally significant τ* values for any S2G strategy (Table S5).

We validated the gene scores implicated in our study by investigating whether they were enriched in five "gold-standard" disease-related gene sets: 195 approved drug target genes for autoimmune diseases;[10,45] 550 Mendelian genes related to immune dysregulation,[46] 390 Mendelian genes related to blood disorders,[47] 146 "Bone Marrow/Immune" genes defined by the Developmental Disorders Database/Genotype-Phenotype Database (DDD/G2P),[48] and 2,200 (top 10%) high-pLI genes[49] (Figure 3C and Table S6). (We note that the high-pLI genes should not be viewed as a strict gold standard, as not all of these genes are disease-related, but ≈30% of these genes have an established human disease phenotype.[49])

Subsequent subsections are organized in the following order: description of gene scores for that subsection; marginal analyses using S-LDSC; joint analyses using S-LDSC; and validation of the gene scores implicated in our study using gold-standard disease-related gene sets.

**Enhancer-related genes are conditionally informative for autoimmune disease heritability**

We assessed the disease informativeness of seven gene scores prioritizing enhancer-related genes in blood. We defined these gene scores based on distal enhancer-gene connections, tissue-specific expression, or tissue-specific expression quantitative trait loci (eQTL), all of which can characterize enhancer-related regulation (Figure 1B, Table 1, and STAR Methods). Some of these gene scores were derived from the same functional data that we used to define S2G strategies (e.g., ABC[32,39] and assay for transposase-accessible chromatin using sequencing [ATAC-seq];[33] see URLs). We included two published gene scores: (binarized) blood-specific EDS[24] and specifically expressed genes in Genotype-Tissue Expression (GTEx) whole blood[9] (SEG-GTEx). We use the term "enhancer-related" to broadly describe gene scores with high predicted functionality under a diverse set of metrics, notwithstanding the fact that all genes require the activation of enhancers and their promoters. Four of our enhancer-related gene scores (ABC-G, ATAC-distal, EDS-binary, and promoter-capture Hi-C [PC-HiC]) were explicitly defined based on distal enhancer-gene connections. Using the established EDS-binary (derived from the published EDS[24]) as a point of reference, we determined that the other three gene scores (ABC-G, ATAC-distal, and PC-HiC) had an average excess overlap of 1.73 with the EDS-binary score (p values per gene score: $2 \times 10^{-8}$ to $6 \times 10^{-6}$; Table S7), confirming that they prioritize enhancer-related genes. Three of our enhancer-related scores (eQTL-CTS, Expecto-MVP (magnitude of variation potential), and SEG-GTEx) were not explicitly defined based on distal enhancer-gene connections. We determined that these three gene scores also had an average excess overlap of 1.5× with the EDS-binary score (p values per gene score: $4 \times 10^{-7}$ to $1 \times 10^{-4}$; Table S7), confirming that they prioritize enhancer-related genes; notably, the excess overlap of 1.5× was almost as large as the excess of overlap of 1.7× for gene scores defined based on distal enhancer-gene connections.

We combined the seven enhancer-related gene scores with the ten S2G strategies (Table 2) to define 70 annotations. In our marginal analysis using S-LDSC conditional on the baseline-LD+ model (meta-analyzing S-LDSC results across 11 autoimmune diseases and blood cell traits), all 70 enhancer-related annotations were significantly enriched for disease heritability, with larger enrichments for smaller annotations (Figure S5 and Table S8); values

of standardized enrichment were more similar across annotations (Figure S6 and Table S9). Thirty-seven of the 70 enhancer-related annotations attained conditionally significant $\tau^*$ values after Bonferroni correction ($p < 0.05/110$) (Figure 3A and Table S8). We observed the strongest conditional signal for ATAC-distal × ABC ($\tau^* = 1.0 \pm 0.2$). ATAC-distal is defined by the proportion of mouse gene expression variability across blood cell types that is explained by distal ATAC-seq peaks in mouse;[33] the mouse genes are mapped to orthologous human genes. Four of the seven gene scores (ABC-G, ATAC-distal, EDS-binary, and SEG-GTEx) produced strong conditional signals across many S2G strategies; however, none of them attained Bonferroni-significant $\tau^*$ for all ten S2G strategies (Figure 3A). Among the S2G strategies, average conditional signals were strongest for the ABC strategy (average $\tau^* = 0.59$) and TSS strategy (average $\tau^* = 0.52$), which greatly outperformed the window-based S2G strategies (average $\tau^* = 0.04–0.07$), emphasizing the high added value of S2G strategies incorporating functional data (especially the ABC and TSS strategies).

We compared meta-analyses of S-LDSC results across six autoimmune diseases versus five blood cell traits (Figures 3B and S7; Tables S1, S10, and S11). Results were broadly concordant ($r = 0.57$ between $\tau^*$ estimates), with slightly stronger signals for autoimmune diseases (slope = 1.3). We also compared meta-analyses of results across two granulocyte-related blood cell traits (white blood cell count and eosinophil count) versus three red blood cell or platelet-related blood cell traits (red blood cell count, red blood cell distribution width, and platelet count) (Figure S8; Tables S12 and S13). Results were broadly concordant ($r = 0.65$, slope = 1.1). We also examined S-LDSC results for individual disease/traits and applied a test for heterogeneity[50] (Figures S9 and S10; Tables S14 and S15). Results were generally underpowered (false discovery rate [FDR] < 5% for 16 of 770 annotation-trait pairs), with limited evidence of heterogeneity across diseases/traits (FDR < 5% for 11 of 70 annotations).

We jointly analyzed the 37 enhancer-related annotations that were Bonferroni-significant in our marginal analysis (Figure 3A and Table S8) by performing forward stepwise elimination to iteratively remove annotations that had conditionally non-significant $\tau^*$ values after Bonferroni correction. Of these, six annotations were jointly significant in the resulting enhancer-related joint model (Figure S11 and Table S16), corresponding to four enhancer-related gene scores: ABC-G, ATAC-distal, EDS-binary, and SEG-GTEx.

We assessed the enrichment of the seven enhancer-related gene scores (Table 1) in five gold-standard disease-related gene sets: drug target genes,[10,45] Mendelian genes (Freund),[46] Mendelian genes (Vuckovic),[47] immune genes,[48] and high-pLI genes[49] (Figure 3C and Table S6). Six of the seven gene scores were significantly enriched (after Bonferroni correction; $p < 0.05/55$) in the drug target genes, all seven were significantly enriched in both Mendelian gene sets, three of seven were significantly enriched in the immune genes, and five of seven were significantly enriched in the high-pLI genes. The largest enrichment was observed for SEG-GTEx genes in the drug target genes (2.4×, SE 0.1) and Mendelian genes (Freund) (2.4×, SE 0.1). These findings validate the high importance to disease of enhancer-related genes.

We performed five secondary analyses. First, for each of the six annotations from the enhancer-related joint model (Figure S11), we assessed their functional enrichment for fine-mapped SNPs for blood-related traits from two previous studies.[51,52] We observed large and significant enrichments for all six annotations (Table S17), consistent with the S-LDSC results. Second, for each of the seven enhancer-related gene scores, we performed pathway enrichment analyses to assess their enrichment in pathways from the ConsensusPathDB database;[53] all seven gene scores were significantly enriched in immune-related and signaling pathways (Table S18). Third, we explored other approaches to combining information across genes that are linked to a SNP using S2G strategies by using either the mean across genes or the sum across genes of the gene scores linked to a SNP, instead of the maximum across genes. We determined that results for either the mean or the sum were very similar to the results for the maximum, with no significant difference in standardized effect sizes of the resulting SNP annotations (Tables S8, S19, and S20). Fourth, we repeated our analyses of the five enhancer-related gene scores for which the top 10% (of genes) threshold was applied, using top 5% or top 20% thresholds instead (Tables S21 and S22). We observed very similar results, with largely non-significant differences in standardized effect sizes. Fifth, we confirmed that our forward stepwise elimination procedure produced identical results when applied to all 70 enhancer-related annotations, instead of just the 37 enhancer-related annotations that were Bonferroni-significant in our marginal analysis.

We conclude that four of the seven characterizations of enhancer-related genes are conditionally informative for autoimmune diseases and blood-related traits when using functionally informed S2G strategies.

## Genes with high network connectivity to enhancer-related genes are even more informative

We assessed the disease informativeness of a gene score prioritizing genes with high connectivity to enhancer-related genes in a protein-protein interaction (PPI) network (PPI-enhancer).We hypothesized that (1) genes that are connected to enhancer-related genes in biological networks are likely to be important, and (2) combining potentially noisy metrics defining enhancer-related genes would increase the statistical signal. We used the STRING PPI network[38] to quantify the network connectivity of each gene with respect to each of the fourjointly informative enhancer-related gene scores from Figure S11 (ABC-G, ATAC-distal, EDS-binary, and SEG-GTEx) (Figure 1D). Network connectivity scores were computed using a random walk with restart algorithm[10,54] (see STAR Methods). We defined the PPI-enhancer gene score based on genes in the top 10% of average network connectivity across the four enhancer-related genescores (Table 1). The PPI-enhancer gene score was only moderately positively correlated with the four underlying enhancer-related gene scores (average r = 0.28; Figure S2).

We combined the PPI-enhancer gene score with the ten S2G strategies (Table 2) to define ten annotations. In our marginal analysis using S-LDSC (meta-analyzing S-LDSC results across 11 autoimmune diseases and blood cell traits), all ten PPI-enhancer annotations were significantly enriched for disease heritability, with larger enrichments for smaller

annotations (Figure S5 and Table S23); values of standardized enrichment were more similar across annotations (Figure S6 and Table S24). All ten PPI-enhancer annotations attained conditionally significant $\tau^*$ values after Bonferroni correction (p < 0.05/110) (Figure 3A and Table S23). Notably, the maximum $\tau^*$ (2.0 [SE 0.3] for PPI-enhancer × ABC) was >2-fold larger than the maximum $\tau^*$ for the recently proposed EDS[24] (0.91 [SE 0.21] for EDS-binary × ABC). All ten PPI-enhancer annotations remained significant when conditioned on the enhancer-related joint model from Figure S11 (Table S25). In a comparison of meta-analyses of S-LDSC results across five blood cell traits versus six autoimmune diseases, results were broadly concordant (r = 0.93 between $\tau^*$ estimates) but with much stronger signals for autoimmune diseases (slope = 2.2) (Figures 3B and S7; Tables S10 and S11). In a comparison of meta-analyses across two granulocyte-related blood cell traits versus three red blood cell or platelet-related blood cell traits, results were broadly concordant (r = 0.83), but with much stronger signals for granulocyte-related blood cell traits (slope = 2.1), providing a further validation that the PPI-enhancer gene score is related to immune response (Figure S8; Tables S12 and S13). In analyses of individual traits, 62 of 110 PPI-enhancer annotation-trait pairs were significant (FDR < 5%) (Figures S9 and S10; Table S14), eight of them with evidence of heterogeneity across diseases/traits (FDR < 5% for eight of ten PPI-enhancer annotations) (Table S15).

We jointly analyzed the six enhancer-related annotations from the enhancer-related joint model (Figure S11) and the ten marginally significant PPI-enhancer annotations conditional on the enhancer-related joint model in Table S25. Of these, three enhancer-related and four PPI-enhancer annotations were jointly significant in the resulting PPI-enhancer-related joint model (Figure 3D and Table S26). The joint signal was strongest for PPI-enhancer × ABC ($\tau^*$ = 1.2 ± 0.21), highlighting the informativeness of the ABC S2G strategy. Three of the seven annotations attained $\tau^*$ > 0.5; annotations with $\tau^*$ > 0.5 are unusual and considered to be important.[55]

We assessed the enrichment of the PPI-enhancer gene score in the five gold-standard disease-related gene sets: drug target genes,[10,45] Mendelian genes (Freund),[46] Mendelian genes (Vuckovic),[47] immune genes,[48] and high-pLI genes[49] (Figure 3C and Table S6). The PPI-enhancer gene score showed significant enrichment in all five gene sets, with higher magnitude of enrichment compared with any of the seven enhancer-related gene scores. In particular, the PPI-enhancer gene score was 5.3× (SE 0.1) enriched in drug target genes and 4.6× (SE 0.1) enriched in Mendelian genes (Freund), a 2-fold stronger enrichment in each case than the EDS-binary gene score[24] (2.1× [SE 0.1] and 2.3× [SE 0.1]).

We performed three secondary analyses. First, for each of the four jointly significant PPI-enhancer annotations from Figure 3D, we assessed their functional enrichment for fine-mapped SNPs for blood-related traits from two previous studies.[51,52] We observed large and significant enrichments for all four annotations (Table S17), consistent with the S-LDSC results (and with the similar analysis of enhancer-related annotations described above). Second, we performed a pathway enrichment analysis to assess the enrichment of the PPI-enhancer gene score in pathways from the ConsensusPathDB database;[53] this gene score was enriched in immune-related pathways (Table S18). Third, we confirmed that our forward stepwise elimination procedure produced identical results when applied

to all 80 enhancer-related and PPI-enhancer annotations, instead of just the six enhancer-related annotations from the enhancer-related joint model (Figure S11) and the ten PPI-enhancer annotations. Additional analyses assessing the relative importance of PPI-network information versus combining different enhancer-related gene scores are described in Methods S1.

We conclude that genes with high network connectivity to enhancer-related genes are conditionally informative for autoimmune diseases and blood-related traits when using functionally informed S2G strategies.

## Candidate master-regulator genes are conditionally informative for autoimmune disease heritability

We assessed the disease informativeness of two gene scores prioritizing candidate master-regulator genes in blood. We defined these gene scores using whole-blood eQTL data from the eQTLGen consortium[56] (*Trans*-master) and a published list of known transcription factors in humans (TF)[37] (Figure 1C, Table 1, and STAR Methods). We note that TF genes do not necessarily act as master regulators and only a small number of transcription factors regulate many downstream genes, but TF genes can still be viewed as candidate master regulators. Using 97 known master-regulator genes from 18 master-regulator families[57–61] as a point of reference, we determined that *Trans*-master and TF genes had 3.5× and 5.4× excess overlaps with the 97 candidate master-regulator genes (p values: $5.6 \times 10^{-72}$ and $2.2 \times 10^{-160}$; Tables S27 and S28), confirming that they prioritize candidate master-regulator genes.

In detail, *Trans*-master is a binary gene score defined by genes that significantly regulate three or more other genes in *trans* via SNPs that are significant *cis*-eQTLs of the focal gene (10% of genes); the median value of the number of genes *trans*-regulated by a *Trans*-master gene is 14. Notably, *trans*-eQTL data from the eQTLGen consortium[56] was only available for 10,317 previously disease-associated SNPs. It is possible that genes with significant *cis*-eQTL that are disease-associated SNPs may be enriched for disease heritability irrespective of *trans* signals. To account for this gene-level bias, we conditioned all analyses of *Trans*-master annotations on both (1) ten annotations based on a gene score defined by genes with at least one disease-associated *cis*-eQTL, combined with each of the ten S2G strategies, and (2) ten annotations based on a gene score defined by genes with at least three unlinked disease-associated *cis*-eQTL, combined with each of the ten S2G strategies; we chose the number 3 to maximize the correlation between this gene score and the *Trans*-master gene score (r = 0.32). Thus, our primary analyses were conditioned on 93 baseline-LD+ and 20 additional annotations (113 baseline-LD+ *cis* model annotations); additional secondary analyses are described below. We did not consider a SNP annotation defined by *trans*-eQTLs because the *trans*-eQTLs in eQTLGen data were restricted to disease-associated SNPs, which would bias our results.

We combined the *Trans*-master gene score with the ten S2G strategies (Table 2) to define ten annotations. In our marginal analysis using S-LDSC conditional on the baseline-LD+ *cis* model, all ten *Trans*-master annotations were strongly and significantly enriched for disease heritability, with larger enrichments for smaller annotations (Figure S5 and Table

S29); values of standardized enrichment were more similar across annotations (Figure S6 and Table S30). All ten *Trans*-master annotations attained conditionally significant $\tau^*$ values after Bonferroni correction (p < 0.05/110) (Figure 4A and Table S29). We observed the strongest conditional signals for *Trans*-master × TSS ($\tau^* = 1.6$ versus $\tau^* = 0.37$–0.39 for candidate master-regulator × window-based S2G strategies). We observed similar (slightly more significant) results when conditioning on baseline-LD+ annotations only (Table S31).

As noted above, *trans*-eQTL data from the eQTLGen consortium[56] were only available for 10,317 previously disease-associated SNPs, and we thus defined and conditioned on baseline-LD+ *cis* model annotations to account for gene-level bias. We verified that conditioning on annotations derived from gene scores defined by other minimum numbers of *cis*-eQTL and/or unlinked *cis*-eQTL produced similar results (Tables S32–S36). To verify that our results were not impacted by SNP-level bias, we adjusted each of the ten *Trans*-master annotations by removing all disease-associated *trans*-eQTL SNPs in the eQTLGen data from the annotation, as well as any linked SNPs (STAR Methods). We verified that these adjusted annotations produced similar results (Table S37). TF is a binary gene score defined by a published list of 1,639 known transcription factors in humans.[37] We combined TF with the ten S2G strategies (Table 2) to define ten annotations. In our marginal analysis conditional on the baseline-LD+ *cis* model, all ten TF annotations were significantly enriched for heritability but with smaller enrichments than the *Trans*-master annotations (Table S29; see Table S30 for standardized enrichments). Nine TF annotations attained significant $\tau^*$ values after Bonferroni correction (Figure 4A and Table S29) (the same nine annotations were also significant conditional on the baseline-LD+ model; Table S31). Across all S2G strategies, $\tau^*$ values of *Trans*-master annotations were larger than those of TF annotations (Table S29).

We compared meta-analyses of S-LDSC results across six autoimmune diseases versus five blood cell traits (Figures 4B and S7; Tables S1, S38, and S39). Results were broadly concordant (r = 0.56 between $\tau^*$ estimates), with slightly stronger signals for blood cell traits (slope = 0.57). We also compared meta-analyses of results across two granulocyte-related blood cell traits versus three red blood cell or platelet-related blood cell traits (Figure S8; Tables S40 and S41). Results were broadly concordant (r = 0.94, slope = 1.12). We also examined S-LDSC results for individual disease/traits and applied a test for heterogeneity[50] (Figures S12 and S13; Tables S14 and S15). We observed several annotation-trait pairs with disease signal (FDR < 5% for 96 of 220 annotation-trait pairs), with evidence of heterogeneity across diseases/traits (FDR < 5% for 10 of 20 annotations).

We jointly analyzed the ten *Trans*-master and nine TF annotations that were Bonferroni-significant in our marginal analysis (Figure 4A and Table S29) by performing forward stepwise elimination to iteratively remove annotations that had conditionally non-significant $\tau^*$ values after Bonferroni correction. Of these, three *Trans*-master annotations and two TF annotations were jointly significant in the resulting candidate master-regulator joint model (Figure S14 and Table S42). The joint signal was strongest for *Trans*-master × Roadmap ($\tau^* = 0.81$, SE = 0.13), emphasizing the high added value of the Roadmap S2G strategy.

We assessed the enrichment of the *Trans*-master and TF gene scores in the five gold-standard disease-related gene sets: drug target genes,[10,45] Mendelian genes (Freund),[46] Mendelian genes (Vuckovic),[47] immune genes,[48] and high-pLI genes[49] (Figure 4C and Table S6). The *Trans*-master gene score showed higher enrichment in all five gene sets compared with the TF gene score. The enrichments for candidate master-regulator genes were lower (1.4×, SE 0.07) for drug target genes in comparison with some enhancer-related genes and the PPI-enhancer gene score (Figure 3C); this can be attributed to the fact that candidate master-regulator genes may tend to disrupt genes across several pathways, rendering them unsuitable as drug targets.

We performed seven secondary analyses. First, for comparison purposes, we defined a binary gene score (*Trans*-regulated) based on genes with at least one significant *trans*-eQTL. We combined *Trans*-regulated genes with the ten S2G strategies to define ten annotations. In our marginal analysis using S-LDSC conditional on the baseline-LD+ *cis* model, none of the *Trans*-regulated annotations attained conditionally significant $\tau^*$ values after Bonferroni correction ($p < 0.05/110$) (Table S43). (In contrast, three of the annotations were significant when conditioning only on the baseline-LD+ model; Table S44.) Second, a potential complexity is that *trans*-eQTL in whole blood may be inherently enriched for blood cell trait-associated SNPs (since SNPs that regulate the abundance of a specific blood cell type would result in *trans*-eQTL effects on genes that are specifically expressed in that cell type[56]), potentially limiting the generalizability of our results to non-blood-cell traits. To ensure that our results were robust to this complexity, we verified that analyses restricted to the five autoimmune diseases (Table S1) produced similar results (Table S45). Third, for each of the five annotations from the candidate master-regulator joint model (Figure S14), we assessed their functional enrichment for fine-mapped SNPs for blood-related traits from two previous studies.[51,52] We observed large and significant enrichments for all five annotations (Table S17), consistent with the S-LDSC results (and with similar analyses described above). Fourth, we performed pathway enrichment analyses to assess the enrichment of the *Trans*-master and TF gene scores in pathways from the ConsensusPathDB database.[53] The *Trans*-master gene score was significantly enriched in immune-related pathways (Table S18). Fifth, we explored other approaches to combining information across genes that are linked to a SNP using S2G strategies, by using either the mean across genes or the sum across genes of the gene scores linked to a SNP, instead of the maximum across genes. We determined that results for either the mean or the sum were very similar to the results for the maximum, with no significant difference in standardized effect sizes of the resulting SNP annotations (Tables S29, S19, and S20). Sixth, we repeated our analyses of the *Trans*-master gene score, defined in our primary analyses based on 2,215 genes that *trans*-regulate ≥3 genes, using either 3,717 genes that *trans*-regulate ≥1 gene (most of which *trans*-regulate multiple genes) or 1,170 genes that *trans*-regulate ≥10 genes (Table S46). We observed very similar results, with largely non-significant differences in standardized effect sizes. Seventh, we confirmed that our forward stepwise elimination procedure produced identical results when applied to all 20 candidate master-regulator annotations, instead of just the 19 candidate master-regulator annotations that were Bonferroni-significant in our marginal analysis.

We conclude that candidate master-regulator genes are conditionally informative for autoimmune diseases and blood-related traits when using functionally informed S2G strategies.

## Genes with high network connectivity to candidate master-regulator genes are even more informative

We assessed the disease informativeness of a gene score prioritizing genes with high connectivity to candidate master-regulator genes in the STRING PPI network[38] (PPI-master, analogous to PPI-enhancer; see STAR Methods and Table 1). The PPI-master gene score was positively correlated with the two underlying candidate master-regulator gene scores (average r = 0.43) and modestly correlated with PPI-enhancer (r = 0.22) (Figure S2). In addition, it had an excess overlap of 7.2× with the 97 known master-regulator genes[57–61] (p = 2 × 10$^{-214}$; Tables S27 and S28).

We combined the PPI-master gene score with the ten S2G strategies (Table 2) to define ten annotations. In our marginal analysis using S-LDSC conditional on the baseline-LD+ *cis* model, all ten PPI-master annotations were significantly enriched for disease heritability, with larger enrichments for smaller annotations (Figure 4A and Table S47); values of standardized enrichment were more similar across annotations (Figure S6 and Table S48). All ten PPI-master annotations attained conditionally significant τ* values after Bonferroni correction (p < 0.05/110) (Figure 4B and Table S47) (as expected, results were similar when conditioning only on the baseline-LD+ model; Table S49). We observed the strongest conditional signals for PPI-master combined with TSS (τ* = 1.7, SE 0.16), Coding (τ* = 1.7, SE 0.14), and ABC (τ* = 1.6, SE 0.17) S2G strategies, again emphasizing the high added value of S2G strategies incorporating functional data (Table S47). Nine of the ten PPI-master annotations remained significant when conditioning on the candidate master-regulator joint model from Figure S14 (Table S50). In a comparison of meta-analyses of S-LDSC results across five blood cell traits versus six autoimmune diseases, results were broadly concordant (r = 0.81 between τ* estimates, slope = 0.93) (Figures 4B and S7; Tables S38 and S39). In a comparison of meta-analyses across two granulocyte-related blood cell traits versus three red blood cell or platelet-related blood cell traits, results were broadly concordant but with slightly stronger signals for granulocyte-related traits (r = 0.92, slope = 1.3), providing further validation that the PPI-master gene score is related to immune response (Figure S8; Tables S40 and S41). In the analyses of individual traits, 101 of 110 PPI-enhancer annotation-trait pairs were significant (FDR < 5%) (Figures S12 and S13; and Table S14), with evidence of heterogeneity across diseases/traits (FDR < 5% for six of ten PPI-master annotations) (Table S15).

We jointly analyzed the five candidate master-regulator annotations from the candidate master-regulator joint model (Figure S14 and Table S42) and the nine PPI-master annotations significant conditional on the candidate master-regulator joint model in Table S50. Of these, two *Trans*-master and three PPI-master annotations were jointly significant in the resulting PPI-master-regulator joint model (Figure 4D and Table S51). The joint signal was strongest for PPI-master × Roadmap (τ* = 0.94 ± 0.14), and four of the five annotations attained τ* > 0.5.

We assessed the enrichment of the PPI-master gene score in the five gold-standard disease-related gene sets: drug target genes,[10,45] Mendelian genes (Freund),[46] Mendelian genes (Vuckovic),[47] immune genes,[48] and high-pLI genes[49] (Figure 4C and Table S6). The PPI-master gene score showed significant enrichment in all five gene sets, with higher magnitude of enrichment compared with either of the candidate master-regulator gene scores. In particular, the PPI-master gene score was 2.7× (SE 0.1) enriched in drug target genes and 3.4× (SE 0.1) enriched in Mendelian genes (Freund).

We performed three secondary analyses. First, for each of the three jointly significant PPI-master annotations from Figure 4D, we assessed their functional enrichment for fine-mapped SNPs for blood-related traits from two previous studies.[51,52] We observed large and significant enrichments for all three annotations (Table S17), consistent with the S-LDSC results (and with similar analyses described above). Second, we performed a pathway enrichment analysis to assess the enrichment of the PPI-master gene score in pathways from the ConsensusPathDB database[53] and report the top enriched pathways (Table S18). Third, we confirmed that our forward stepwise elimination procedure produced identical results when applied to all 30 candidate master-regulator and PPI-master annotations, instead of just the five candidate master-regulator annotations from the candidate master-regulator joint model (Figure S14) and the nine PPI-master annotations that were Bonferroni-significant in our marginal analysis.

We conclude that genes with high network connectivity to candidate master-regulator genes are conditionally informative for autoimmune diseases and blood-related traits when using functionally informed S2G strategies.

## Combined joint model

We constructed a combined joint model containing annotations from the above analyses that were jointly significant, contributing information conditional on all other annotations. We merged the baseline-LD+ *cis* model with annotations from the PPI-enhancer (Figure 3D) and PPI-master (Figure 4D) joint models, and performed forward stepwise elimination to iteratively remove annotations that had conditionally non-significant $\tau^*$ values after Bonferroni correction (p < 0.05/110). The combined joint model contained eight new annotations, including two enhancer-related, two PPI-enhancer, two *Trans*-master, and two PPI-master annotations (Figure 5 and Table S52). The joint signals were strongest for PPI-enhancer × ABC ($\tau^* = 0.99$, SE 0.23) and PPI-master × Roadmap ($\tau^* = 0.91$, SE 0.12), highlighting the importance of two distal S2G strategies, ABC and Roadmap; five of the eight new annotations attained $\tau^* > 0.5$. Analyses confirmed that the combined joint model outperforms other heritability models (Methods S1).

We investigated the biology of individual loci by examining 1,198 SNPs that were previously confidently fine-mapped (posterior inclusion probability [PIP] > 0.90) for one autoimmune disease and five blood cell traits from the UK Biobank. Focusing on the four highly enriched regulatory annotations from Figure 5 (enrichment ≥ 18; 1.5% of SNPs in total), 194 of the 1,198 SNPs belonged to one or more of these four annotations. A list of these 194 SNPs is provided in Table S53. We highlight three notable examples. First, rs231779, a fine-mapped SNP (PIP = 0.91) for "All Auto Immune Traits" (Table S1), was

linked by the ABC S2G strategy to *CTLA4*, a high-scoring gene for the PPI-enhancer gene score (ranked 109) (Figure S17A). *CTLA4* acts as an immune checkpoint for activation of T cells and is a key target gene for cancer immunotherapy.[62–64] Second, rs6908626, a fine-mapped SNP (PIP = 0.99) for "All Auto Immune Traits" (Table S1), was linked by the Roadmap S2G strategy to *BACH2*, a high-scoring gene for the *Trans*-master gene score (ranked 311) (Figure S17B). *BACH2* is a known master-regulator transcription factor that functions in innate and adaptive lineages to control immune responses,[65,66] has been shown to control autoimmunity in mice knockout studies,[67] and has been implicated in several autoimmune and allergic diseases including lupus, type 1 diabetes, and asthma.[68–70] Third, *rs113473633*, a fine-mapped SNP (PIP = 0.99 and PIP = 0.99) for white blood cell (WBC) count and eosinophil count (Table S1), was linked by the Roadmap S2G strategy to *NFKB1*, a high-scoring gene for the *Trans*-master and PPI-master gene scores (ranked 409 and 111) (Figure S17C for WBC count, Figure S17D for eosinophil count). *NFKB1* is a major transcription factor involved in immune response[71] and is critical for development and proliferation of lymphocytes,[72,73] and has previously been implicated in blood cell traits.[47] In each of these examples, we nominate both the causal gene and the SNP-gene link.

We performed five secondary analyses. First, we investigated whether the eight annotations of the combined joint model still contributed unique information after including the pLI gene score,[49] which has previously been shown to be conditionally informative for disease heritability.[42,74,75] We confirmed that all eight annotations from Figure 5 remained jointly significant (Figure S18 and Table S54). Second, we constructed a less restrictive combined joint model by conditioning on the baseline-LD+ model instead of the baseline-LD+ *cis* model. The less restrictive combined joint model included one additional annotation, SEG-GTEx × Coding (Table S55). This implies that the combined joint model is largely invariant to conditioning on the baseline-LD+ or baseline-LD+ *cis* model. Third, we analyzed binarized versions of all 11 gene scores (Table 1) using MAGMA,[76] an alternative gene set analysis method. Nine of the 11 gene scores produced significant signals (Table S56), 11 marginally significant gene scores (Figures 3 and 4), and five gene scores included in the combined joint model of Figure 5 in the S-LDSC analysis. However, MAGMA does not allow for conditioning on the baseline-LD model, does not allow for joint analysis of multiple gene scores to assess joint significance, and does not allow for incorporation of S2G strategies. Fourth, we confirmed that our forward stepwise elimination procedure produced identical results when applied to all 110 enhancer-related, candidate master-regulator, PPI-enhancer, and PPI-master annotations, instead of just the 12 annotations from the PPI-enhancer (Figure 3D) and PPI-master (Figure 4D) joint models. Fifth, we assessed the model fit of the final joint model by correlating the residuals from stratified LD score regression with the independent variables in the regression (annotation-specific LD scores) for each of the 11 blood-related traits (Figure S19). We observed an average squared correlation of 0.02 across annotation-specific LD scores and traits, suggesting good model fit.

We conclude that enhancer-related genes and candidate master-regulator genes, as well as genes with high network connectivity to those genes, are jointly informative for autoimmune diseases and blood-related traits when using functionally informed S2G strategies.

# DISCUSSION

## Biological significance

We have assessed the contribution to autoimmune disease of enhancer-related genes and candidate master-regulator genes, incorporating PPI-network information and ten functionally informed S2G strategies. Our results provide information about which genes impact disease risk, distinguishing specific types of genes that play a greater role in genetic risk of disease (and have not previously been implicated in playing a greater role in genetic risk of disease). In some ways, our results distinguishing genes that are important for disease provide a quantitative improvement over previous work (e.g., versus EDS-binary, a previously proposed enhancer-related gene score[24]). However, in other ways our results provide qualitatively new findings (e.g., candidate master-regulator genes and genes that interact with enhancer-related genes or master-regulator genes without being directly implicated). Our characterization of genes that are important for disease is validated by their enrichment in gold-standard gene sets, including autoimmune disease drug targets and Mendelian genes related to immune dysregulation; these enrichments were higher than for previously published characterizations. Notably, 22 out of 196 drug target genes were uniquely implicated by PPI-enhancer gene score as compared with other enhancer-regulated gene scores (based on top 10% genes) (Table S57). These include three genes, *CCL2*, *IFNA1*, and *IKBKB*[77–79] known to be particularly important for autoimmune disease; for IKBKB, we further note that the SNP rs4737010 (chromosome 8) is a fine-mapped SNP for lymphocyte count that is implicated by our PPI-enhancer × ABC annotation (the annotation that is most conditionally informative for disease in our combined joint model; Figure 5 and Table S53). Similarly, 34 of 196 drug target genes were uniquely implicated by PPI-master gene score as compared with other candidate master-regulator gene scores (Table S57). Furthermore, although gold-standard gene sets may be viewed as positive controls, our results are expected to also implicate true disease genes that are not previously known. Genes uniquely implicated by PPI-enhancer that may be important for autoimmune disease include *CD70*, a known target for cancer immunotherapy,[80,81] and the *STAT* family genes (*STAT4*, *STAT5A*, and *STAT6*), which serve to organize the epigenetic landscape of immune cells[82]—both of which were not implicated by known gold-standard gene sets. Our results provide a route to performing functional follow-up experiments to elucidate and validate specific biological mechanisms (see below).

## Downstream implications

Our work has several downstream implications. First, the PPI-enhancer gene score, which attained a particularly strong enrichment for approved autoimmune disease drug targets, will aid prioritization of drug targets that share similar characteristics with previously discovered drugs, analogous to pLI[49] and LOEUF.[83,84] Second, it is not practical to perform functional experiments on every SNP or genomic locus in the genome; using our results, specific gene-linked regulatory regions implicated by our results can be targeted for functional follow-up experiments (e.g., CRISPR base editing targeted at GWAS fine-mapped autoimmune disease SNPs linked to genes implicated by our gene scores) to elucidate and validate specific biological mechanisms. Third, our results implicate the ABC and Roadmap S2G linking strategies as highly informative distal S2G strategies, and TSS

as a highly informative proximal S2G strategy, when linking SNPs to genes in analyses prioritizing genes or pathways; these S2G strategies should be used instead of or in combination with standard gene window-based S2G strategies. Fourth, our framework for disease heritability analysis incorporating regulatory S2G strategies (instead of conventional window-based approaches) is broadly applicable to other gene sets, e.g., characterizing cell types and cellular processes, as in our more recent work.[85] Fifth, at the level of genes, our findings have immediate potential for improving gene-level probabilistic fine-mapping of transcriptome-wide association studies[86] and gene-based association statistics,[87] using the gene scores as gene-level features to inform gene-level priors based on functional similarity of genes. Sixth, at the level of SNPs, our findings have immediate potential for improving functionally informed fine-mapping[52,88–90] (including experimental follow-up[91]), polygenic localization,[52] and polygenic risk prediction;[92,93] specifically, SNP annotations derived from SNPs linked to high-scoring genes can be used to inform SNP-level priors used in these applications.

## Limitations of the study

Our work has several limitations, representing important directions for future research. First, we caution the readers that the terms "enhancer-related genes" and "candidate master-regulator" genes are inherently broad, and individual gene scores and annotations should be interpreted based on their specific meanings. Second, our results do not provide an understanding of specific biological mechanisms at individual disease loci, necessitating functional follow-up. Third, our findings distinguish specific types of genes that play a greater role in genetic risk of disease but do not localize disease risk to a small number of genes, motivating more precise gene-level characterizations. Fourth, we restricted our analyses to enhancer-related and candidate master-regulator genes in blood, focusing on autoimmune diseases and blood-related traits; this choice was primarily motivated by the better representation of blood cell types in functional genomics assays and *trans*-eQTL studies. However, it will be invaluable to extend our analyses to other tissues and traits as more functional data become available. Fifth, the *trans*-eQTL data from eQTLGen consortium[56] is restricted to 10,317 previously disease-associated SNPs; we modified our analyses to account for this bias. However, it would be invaluable to extend our analyses to genome-wide *trans*-eQTL data to large sample sizes if those data become available. Sixth, we investigated the ten S2G strategies separately, instead of constructing a single optimal combined strategy. A comprehensive evaluation of S2G strategies, and a method to combine them, will be provided elsewhere (S. Gazal, unpublished data). Seventh, the forward stepwise elimination procedure that we use to identify jointly significant annotations[29] is a heuristic procedure whose choice of prioritized annotations may be close to arbitrary in the case of highly correlated annotations; however, the correlations between the gene scores, S2G strategies, and annotations that we analyzed were modest. Eighth, the potential of the gene scores implicated in this study to aid prioritization of future drug targets—based on observed gene-level enrichments for approved autoimmune disease drug targets—is subject to the limitation that novel drug targets that do not adhere to existing patterns may be missed; encouragingly, we also identify gene-level enrichments of the gene scores implicated in this study for four other gold-standard disease-related gene sets. Despite all these limitations, our findings expand and enhance our understanding of which gene-level

characterizations of enhancer-related and candidate master-regulatory architecture and their corresponding gene-linked regions impact autoimmune diseases.

# STAR★METHODS

## RESOURCE AVAILABILITY

**Lead contact**—Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Kushal K. Dey (kdey@hsph.harvard.edu).

**Materials availability**—This study did not generate new unique reagents.

**Data and code availability**

- All gene scores, S2G links, and SNP annotations analyzed in this study are publicly available here:https://data.broadinstitute.org/alkesgroup/LDSCORE/Dey_Enhancer_MasterReg. Tables S14 and S53 are provided as Excel files in the above link. We have also included annotations for 93 million Haplotype Reference Consortium (HRC) SNPs and 170 million TOPMed SNPs (Freeze 3A).DOIs are listed in the Key resources table.

- All original codes reported in this paper, related to generating SNP annotations from gene sets, and for performing PPI-informed integration of gene sets are publicly available on Github: https://github.com/kkdey/GSSG and also submitted to Zenodo (https://zenodo.org/badge/latestdoi/278143533). DOIs are listed in the Key resources table.

- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

## METHOD DETAILS

**Genomic annotations and the baseline-LD model**—We define an annotation as an assignment of a numeric value to each SNP in a predefined reference panel (e.g., 1000 Genomes Project;[31] see URLs). Binary annotations can have value 0 or one only. Continuous-valued annotations can have any real value; our focus is on continuous-valued annotations with values between 0 and 1. Annotations that correspond to known or predicted function are referred to as functional annotations. The baseline-LD model (v.2.1) contains 86 functional annotations (see URLs). These annotations include binary coding, conserved, and regulatory annotations (e.g., promoter, enhancer, histone marks, TFBS) and continuous-valued linkage disequilibrium (LD)-related annotations.

**Gene scores**—We define a gene score as an assignment of a numeric value between 0 and one to each gene; we primarily focus on binary gene sets defined by the top 10% of genes. We analyze a total of 11 gene scores (Table 1): seven enhancer-related gene scores, two candidate master-regulator gene scores and two PPI-based gene scores (PPI-master, PPI-enhancer) that aggregate information across enhancer-related and candidate master-regulator gene scores. We scored 22,020 genes on chromosomes 1–22 from ref.[7] (see URLs). When

selecting the top 10% of genes for a given score, we rounded the number of genes to 2,200. We used the top 10% of genes in our primary analyses to be consistent with previous work,[9] who also defined gene scores using the top 10% of genes for a given metric, and to ensure that all SNP annotations (gene scores × S2G strategies) analyzed were of reasonable size (0.2% of SNPs or larger).

The seven enhancer-related gene scores are as follows:

- **ABC-G**: A binary gene score denoting genes that are in top 10% of the number of 'intergenic' and 'genic' Activity-by-Contact (ABC) enhancer to gene links in blood cell types, with average HiC score fraction > 0.015[32] (see URLs).

- **ATAC-distal**: A probabilistic gene score denoting the proportion of gene expression variance in 86 immune cell types in mouse, that is explained by the patterns of chromatin covariance of distal enhancer OCRs (open chromatin regions) to the gene, compared to chromatin covariance of OCRs that are near TSS of the gene and unexplained variances (see Figure 2 from[33]). The genes were mapped to their human orthologs using Ensembl biomaRt.[97].

- **EDS-binary**: A binary gene score denoting genes that are in top 10% of the blood-specific Activity-based Enhancer Domain Score (EDS)[24] that reflects the number of conserved bases in enhancers that are linked to genes in blood related cell types as per the Roadmap Epigenomics Project[43,98](see URLs).

- **eQTL-CTS**: A probabilistic gene score denoting the proportion of immune cell-type-specific eQTLs (with FDR adjusted p value < 0.05 in one or two cell-types) across 15 different immune cell-types from the DICEdb project[34] (see URLs). We consider this to be an enhancer-related gene score, as cell type specificity is often characterized by different enhancer activation status in different cell types.[99,100].

- **Expecto-MVP**: A binary gene score denoting genes that are in top 10% in terms of the magnitude of variation potential (MVP) in GTEx Whole Blood, which is the sum of the absolute values of all directional mutation effects within 1kb of the TSS upstream and downstream, as evaluated by the Expecto method[7](see URLs). We consider this to be an enhancer-related gene score, as this score has been reported to be indicative of tissue specificity of expression and activation/repression status.[7].

- **PC-HiC-distal**: A binary gene score denoting genes that are in top 10% in terms of the total number of Promoter-capture HiC connections across 17 primary blood cell-types.

- **SEG-GTEx**: A binary gene score denoting genes that are in top 10% in terms of the SEG t-statistic[9] score in GTEx Whole Blood. We consider this to be an enhancer-related gene score, as tissue specificity is often characterized by different enhancer activation status in different tissues.[99,100]

The two candidate master-regulator gene scores are as follows:

- **Trans-master**: A binary gene score denoting genes with significant trait-associated *cis*-eQTLs in blood that also act as significant *trans*-eQTLs for at least three other genes based on data from eQTLGen Consortium.[56] We used the threshold of *trans*-regulating ≥ 3 genes in our primary analyses because this results in a gene score spanning ≈10% of genes, analogous to other gene scores.

- **TF**: A binary gene score denoting genes that act as human transcription factors.[37]

The two PPI-based gene scores are as follows:

- **PPI-enhancer**: A binary gene score denoting genes in top 10% in terms of closeness centrality measure to the disease informative enhancer-regulated gene scores. To get the closeness centrality metric, we first perform a Random Walk with Restart (RWR) algorithm[54] on the STRING protein-protein interaction (PPI) network[38,101](see URLs) with seed nodes defined by genes in top 10% of the four enhancer-regulated gene scores with jointly significant disease informativeness (ABC-G, ATAC-distal, EDS-binary and SEG-GTEx). The closeness centrality score was defined as the average network connectivity of the protein products from each gene based on the RWR method.

- **PPI-master**: A binary gene score denoting genes in top 10% in terms of closeness centrality measure to the two disease informative candidate master-regulator gene scores (*Trans*-master and TF). The algorithm was same as that of PPI-enhancer.

**S2G strategies**—We define an SNP-to-gene (S2G) linking strategy as an assignment of 0, one or more linked genes to each SNP with minor allele count ≥ 5 in a 1000 Genomes Project European reference panel.[31] We explored 10 SNP-to-gene linking strategies, including both distal and proximal strategies (Table 2). The proximal strategies included gene body ± 5kb; gene body ±100kb; predicted TSS (by Segway[40,41]); coding SNPs; and promoter SNPs (as defined by UCSC[102,103]). The distal strategies included regions predicted to be distally linked to the gene by Activity-by-Contact (ABC) score[32,39] > 0.015 as suggested in ref.[39] (see below); regions predicted to be enhancer-gene links based on Roadmap Epigenomics data (Roadmap);[43,44,98] regions in ATAC-seq peaks that are highly correlated (>50% as recommended in ref.[33]) to expression of a gene in mouse immune cell-types (ATAC);[33] regions distally connected through promoter-capture Hi-C links (PC-HiC);[35] and SNPs with fine-mapped causal posterior probability (CPP)[42] > 0.001 (we chose this threshold to ensure that the SNP annotations generated after combining the gene scores with the eQTL S2G strategy were of reasonable size (0.2% of SNPs or larger) for all gene scores analyzed) in GTEx whole blood (we use this thresholding on CPP to ensure adequate annotation size for annotations resulting from combining this S2G strategy with the gene scores studied in this paper). We considered combined annotations based on S2G strategies related to gene regulation because SNPs that regulate functionally important genes may be important for disease.

**Activity-by-Contact model predictions—**We used the Activity-by-Contact (ABC) model (https://github.com/broadinstitute/ABC-Enhancer-Gene-Prediction) to predict enhancer-gene connections in each cell type, based on measurements of chromatin accessibility (ATAC-seq or DNase-seq) and histone modifications (H3K27ac ChIP-seq), as previously described.[32,39] In a given cell type, the ABC model reports an "ABC score" for each element-gene pair, where the element is within 5 Mb of the TSS of the gene.

For each cell type, we:

- Called peaks on the chromatin accessibility data using MACS2 with a lenient p value cutoff of 0.1.

- Counted chromatin accessibility reads in each peak and retained the top 150,000 peaks with the most read counts. We then resized each of these peaks to be 500bp centered on the peak summit. To this list we added 500bp regions centered on all gene TSS's and removed any peaks overlapping blacklisted regions[104,105] (https://sites.google.com/site/anshulkundaje/projects/blacklists). Any resulting overlapping peaks were merged. We call the resulting peak set candidate elements.

- Calculated element Activity as the geometric mean of quantile normalized chromatin accessibility and H3K27ac ChIP-seq counts in each candidate element region.

- Calculated element-promoter Contact using the average Hi-C signal across 10 human Hi-C datasets as described below.

- Computed the ABC Score for each element-gene pair as the product of Activity and Contact, normalized by the product of Activity and Contact for all other elements within 5 Mb of that gene.

To generate a genome-wide averaged Hi-C dataset, we downloaded KR normalized Hi-C matrices for 10 human cell types (GM12878, NHEK, HMEC, RPE1, THP1, IMR90, HU- VEC, HCT116, K562, KBM7). This Hi-C matrix (5kb) resolution is available here: ftp://ftp.broadinstitute.org/outgoing/lincRNA/average_hic/average_hic.v2.191020.tar.gz.[32,106] For each cell type we performed the following steps.

- Transformed the Hi-C matrix for each chromosome to be doubly stochastic.

- We then replaced the entries on the diagonal of the Hi-C matrix with the maximum of its four neighboring bins.

- We then replaced all entries of the Hi-C matrix with a value of NaN or corresponding to Knight–Ruiz matrix balancing (KR) normalization factors <0.25 with the expected contact under the power-law distribution in the cell type.

- We then scaled the Hi-C signal for each cell type using the power-law distribution in that cell type as previously described.

- We then computed the "average" Hi-C matrix as the arithmetic mean of the 10 cell-type specific Hi-C matrices. In each cell type, we assign enhancers only to genes whose promoters are "active" (i.e., where the gene is expressed and that promoter drives its expression). We defined active promoters as those in the top 60% of Activity (geometric mean of chromatin accessibility and H3K27ac ChIP-seq counts). We used the following set of TSSs (one per gene symbol) for ABC predictions: https://github.com/broadinstitute/ABC-Enhancer-Gene-Prediction/blob/v0.2.1/reference/RefSeqCurated.170308.bed.CollapsedGeneBounds.bed. We note that this approach does not account for cases where genes have multiple TSSs either in the same cell type or in different cell types.

For intersecting ABC predictions with variants, we took the predictions from the ABC Model and applied the following additional processing steps: (i) We considered all distal element-gene connections with an ABC score 0.015, and all distal or proximal promoter-gene connections with an ABC score 0.1. (ii) We shrunk the ~500-bp regions by 150-bp on either side, resulting in a ~200-bp region centered on the summit of the accessibility peak. This is because, while the larger region is important for counting reads in H3K27ac ChIP-seq, which occur on flanking nucleosomes, most of the DNA sequences important for enhancer function are likely located in the central nucleosome-free region. (iii) We included enhancer-gene connections spanning up to 2 Mb.

## QUANTIFICATION AND STATISTICAL ANALYSIS

**Stratified LD score regression**—Stratified LD score regression (S-LDSC) is a method that assesses the contribution of a genomic annotation to disease and complex trait heritability.[28,29] S-LDSC assumes that the per-SNP heritability or variance of effect size (of standardized genotype on trait) of each SNP is equal to the linear contribution of each annotation.

$$\text{var}(\beta_j) = \sum_c a_{cj}\tau_c$$

where $a_{cj}$ is the value of annotation c for SNP j, where $a_{cj}$ may be binary (0/1), continuous or probabilistic, and $\tau_c$ is the contribution of annotation c to per-SNP heritability conditioned on other annotations. S-LDSC estimates the $\tau_c$ for each annotation using the following equation.

$$E(\chi_j^2) = N \sum_c I(j, c)\tau_c + 1$$

Where $I(j, c) = \sum_k a_{ck} r_{jk}^2$ is the stratified LD score of SNP j with respect to annotation c and $r_{jk}$ is the genotypic correlation between SNPs j and k computed using data from 1000 Genomes Project[31] (see URLs); N is the GWAS sample size.

We assess the informativeness of an annotation c using two metrics. The first metric is enrichment (E), defined as follows (for binary and probabilistic annotations only):

$$E = \frac{h_g^2(c)}{h_g^2} X \frac{M}{\sum_j a_{cj}}$$

where h²$_g$ (c) is the heritability explained by the SNPs in annotation c, weighted by the annotation values. The second metric is standardized effect size ($\tau^\star$) defined as follows (for binary, probabilistic, and continuous-valued annotations):

$$\tau_C^* = \frac{\tau_c sd_c}{h_g^2/M}$$

where $sd_c$ is the SE of annotation c, $h_g^2$ the total SNP heritability and M is the total number of SNPs on which this heritability is computed (equal to 5, 961, 159 in our analyses). $\tau_c^\star$ represents the proportionate change in per-SNP heritability associated with a 1 SD increase in the value of the annotation.

Unlike enrichment, $\tau$ quantifies effects that are conditionally informative, i.e. unique to the focal annotation conditional on other annotations included in the model. In our "marginal" analyses, we estimated $\tau^\star$ for each focal annotation conditional on the baseline-LD annotations. In our "joint" analyses, we merged baseline-LD annotations with focal annotations that were marginally significant after Bonferroni correction and performed forward stepwise elimination to iteratively remove focal annotations that had conditionally non-significant $\tau^\star$ values after Bonferroni correction, as in ref.[11,29,42,55,75,107–109]. We did not consider other feature selection methods, as previous research determined that a LASSO-based feature selection method is computationally expensive and did not perform better in predicting off-chromosome $\chi^2$ association statistics (R. Cui and H. Finucane, personal correspondence). The difference between marginal $\tau^\star$ and joint $\tau^\star$ is that marginal $\tau^\star$ assesses informativeness for disease conditional only on baseline-LD model annotations, whereas joint $\tau^\star$ assesses informativeness for disease conditional on baseline-LD model annotations as well as other annotations in the joint model.

**Combined $\tau^\star$**—We defined a new metric quantifying the conditional informativeness of a heritability model (combined $\tau^*$), generalizing the combined $\tau^\star$ metric of ref.[110] to more than two annotations. In detail, given a joint model defined by M annotations (conditional on a published set of annotations such as the baseline-LD model), we define

$$\tau_{comb}^* = \sqrt{\sum_{\{m=1\}}^{\{M\}} \tau_m^{*2} + \sum_{\{m \neq l\}} r_{ml}\tau_m^*\tau_l^*}$$

Here $r_{ml}$ is the pairwise correlation of the annotations m and l, and $r_{ml}\tau_m^\star \tau_l^\star$ is expected to be positive since two positively correlated annotations typically have the same direction of effect (resp. two negatively correlated annotations typically have opposite directions of effect). We calculate standard errors for $\tau^\star$ using a genomic block-jackknife with 200 blocks.

**Evaluating heritability model fit using SumHer logl$_{SS}$**—Given a heritability model (e.g. the baseline-LD model or the combined joint model of Figure 5), we define the $\Delta$logl$_{SS}$ of that heritability model as the logl$_{SS}$ of that heritability model minus the logl$_{SS}$ of a model with no functional annotations (baseline-LD-nofunct; 17 LD and MAF annotations from the baseline-LD model[29]), where logl$_{SS}$[111] is an approximate likelihood metric that has been shown to be consistent with the exact likelihood from restricted maximum likelihood (REML). We compute p values for $\Delta$logl$_{SS}$ using the asymptotic distribution of the Likelihood Ratio Test (LRT) statistic: $-2$ logl$_{SS}$ follows a $\chi^2$ distribution with degrees of freedom equal to the number of annotations in the focal model, so that $-2$ $\Delta$logl$_{SS}$ follows a $\chi^2$ distribution with degrees of freedom equal to the difference in number of annotations between the focal model and the baseline-LD-nofunct model. We used UK10K as the LD reference panel and analyzed 4,631,901 HRC (haplotype reference panel[112]) well-imputed SNPs with MAF $\geq$ 0.01 and INFO $\geq$ 0.99 in the reference panel; We removed SNPs in the MHC region, SNPs explaining >1% of phenotypic variance and SNPs in LD with these SNPs.

We computed $\Delta$logl$_{SS}$ for eight heritability models:

- baseline-LD model: annotations from the baseline-LD model[29] (86 annotations).

- baseline-LD+ model: baseline-LD model plus seven new S2G annotations not included in the baseline-LD model (93 annotations).

- baseline-LD+ Enhancer model: baseline-LD+ model plus six jointly significant S2G annotations c corresponding to enhancer-related gene scores from Figure S11 (99 annotations).

- baseline-LD+ PPI-enhancer model: baseline-LD+ model plus seven jointly significant S2G annotations c corresponding to enhancer-related and PPI-enhancer gene scores from Figure 3D (100 annotations).

- baseline-LD+ cis model: baseline-LD+ plus 20 S2G annotations used to correct for confounding in evaluation of *Trans*-master gene score (see Results) (113 annotations).

- baseline-LD+ Master model: baseline-LD+ cis plus four jointly significant candidate master-regulator S2G annotations from Figure S14 (117 annotations).

- baseline-LD+ PPI-master model: baseline-LD+ cis plus four jointly significant candidate master-regulator and PPI-master S2G annotations from Figure 4D (117 annotations).

- baseline-LD+ PPI-master model: baseline-LD+ cis plus eight jointly significant enhancer-related, candidate master-regulator,

- PPI-enhancer and PPI-master S2G annotations from the final joint model in Figure 5 (121 annotations).

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

## REFERENCES

1. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science 337, 1190–1195. 10.1126/science.1222794. [PubMed: 22955828]

2. Trynka G, Sandor C, Han B, Xu H, Stranger BE, Liu XS, and Raychaudhuri S (2013). Chromatin marks identify critical cell types for fine mapping complex trait variants. Nat. Genet 45, 124–130. 10.1038/ng.2504. [PubMed: 23263488]

3. Pickrell J (2014). Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. Am. J. Hum. Genet 94, 559–573. 10.1016/j.ajhg.2014.03.004. [PubMed: 24702953]

4. Price AL, Spencer CCA, and Donnelly P (2015). Progress and promise in understanding the genetic basis of common diseases. Proc. Biol. Sci 282, 20151684. 10.1098/rspb.2015.1684.

5. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, and Yang J (2017). 10 years of GWAS discovery: biology, function, and translation. Am. J. Hum. Genet 101, 5–22. 10.1016/j.ajhg.2017.06.005. [PubMed: 28686856]

6. Shendure J, Findlay GM, and Snyder MW (2019). Genomic medicine–progress, pitfalls, and promise. Cell 177, 45–57. 10.1016/j.cell.2019.02.003. [PubMed: 30901547]

7. Zhou J, Theesfeld CL, Yao K, Chen KM, Wong AK, and Troyanskaya OG (2018). Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. Nat. Genet 50, 1171–1179. 10.1038/s41588-018-0160-6. [PubMed: 30013180]

8. Zhu X, and Stephens M (2018). Large-scale genome-wide enrichment analyses identify new trait-associated genes and pathways across 31 human phenotypes. Nat. Commun 9, 4361. 10.1038/s41467-018-06805-x. [PubMed: 30341297]

9. Finucane HK, Reshef Y, Anttila V, Slowikowski K, Gusev A, Byrnes A, Gazal S, Loh PR, Lareau C, Shoresh N, et al. (2018). Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. Nat. Genet 50, 621–629. 10.1038/s41588-018-0081-4. [PubMed: 29632380]

10. Fang H, De Wolf H, Knezevic B, Burnham KL, Osgood J, Sanniti A, Lledó Lara A, Kasela S, De Cesco S, Wegner JK, et al. (2019). A genetics-led approach defines the drug target landscape of 30 immune-related traits. Nat. Genet 51, 1082–1091. 10.1038/s41588-019-0456-1. [PubMed: 31253980]

11. Kim SS, Dai C, Hormozdiari F, van de Geijn B, Gazal S, Park Y, O'Connor L, Amariuta T, Loh PR, Finucane H, et al. (2019). Genes with high network connectivity are enriched for disease heritability. Am. J. Hum. Genet 104, 896–913. 10.1016/j.ajhg.2019.03.020. [PubMed: 31051114]

12. Wang Q, Chen R, Cheng F, Wei Q, Ji Y, Yang H, Zhong X, Tao R, Wen Z, Sutcliffe JS, et al. (2019). A Bayesian framework that integrates multi-omics data and gene networks predicts risk genes from schizophrenia GWAS data. Nat. Neurosci 22, 691–699. 10.1038/s41593-019-0382-7. [PubMed: 30988527]

13. Smillie CS, Biton M, Ordovas-Montanes J, Sullivan KM, Burgin G, Graham DB, Herbst RH, Rogel N, Slyper M, Waldman J, et al. (2019). Intra-and inter-cellular rewiring of the human colon during ulcerative colitis. Cell 178, 714–730.e22. 10.1016/j.cell.2019.06.029. [PubMed: 31348891]

14. Wainberg M, Sinnott-Armstrong N, Mancuso N, Barbeira AN, Knowles DA, Golan D, Ermel R, Ruusalepp A, Quertermous T, Hao K, et al. (2019). Opportunities and challenges for

transcriptome-wide association studies. Nat. Genet 51, 592–599. 10.1038/s41588-019-0385-z. [PubMed: 30926968]

15. Sawle A, Kebschull M, Demmer R, and Papapanou P (2016). Identification of master regulator genes in human periodontitis. J. Dent. Res 95, 1010–1017. 10.1177/0022034516653588. [PubMed: 27302879]

16. Boyle EA, Li YI, and Pritchard JK (2017). An expanded view of complex traits: from polygenic to omnigenic. Cell 169, 1177–1186. 10.1016/j.cell.2017.05.038. [PubMed: 28622505]

17. Brynedal B, Choi J, Raj T, Bjornson R, Stranger BE, Neale BM, Voight BF, and Cotsapas C (2017). Large-scale trans-eQTLs affect hundreds of transcripts and mediate patterns of transcriptional co-regulation. Am. J. Hum. Genet 100, 581–591. 10.1016/j.ajhg.2017.02.004. [PubMed: 28285767]

18. Yao C, Joehanes R, Johnson AD, Huan T, Liu C, Freedman JE, Munson PJ, Hill DE, Vidal M, and Levy D (2017). Dynamic role of trans regulation of gene expression in relation to complex traits. Am. J. Hum. Genet 100, 571–580. 10.1016/j.ajhg.2017.02.003. [PubMed: 28285768]

19. Vargas DM, De Bastiani MA, Zimmer ER, and Klamt F (2018). Alzheimer's disease master regulators analysis: search for potential molecular targets and drug repositioning candidates. Alzheimer's Res. Ther 10, 59. 10.1186/s13195-018-0394-7. [PubMed: 29935546]

20. Montefiori L, Sobreira DR, Sakabe NJ, Aneas I, Joslin AC, Hansen GT, Bozek G, Moskowitz IP, McNally EM, Nóbrega MA, and Nóbrega MA (2018). A promoter interaction map for cardiovascular disease genetics. Elife 7, e35788. 10.7554/elife.35788. [PubMed: 29988018]

21. Liu X, Li YI, and Pritchard JK (2019). Trans effects on gene expression can drive omnigenic inheritance. Cell 177, 1022–1034.e6. 10.1016/j.cell.2019.04.014. [PubMed: 31051098]

22. Doostparast Torshizi A, Armoskus C, Zhang H, Forrest MP, Zhang S, Souaiaia T, Evgrafov OV, Knowles JA, Duan J, and Wang K (2019). Deconvolution of transcriptional networks identifies TCF4 as a master regulator in schizophrenia. Sci. Adv 5, eaau4139. 10.1126/sciadv.aau4139. [PubMed: 31535015]

23. Andersson R, and Sandelin A (2019). Determinants of enhancer and promoter activities of regulatory elements. Nat. Rev. Genet 21, 71–87. 10.1038/s41576-019-0173-8. [PubMed: 31605096]

24. Wang X, and Goldstein DB (2020). Enhancer domains predict gene pathogenicity and inform gene discovery in complex disease. Am. J. Hum. Genet 106, 215–233. 10.1016/j.ajhg.2020.01.012. [PubMed: 32032514]

25. Emison ES, McCallion AS, Kashuk CS, Bush RT, Grice E, Lin S, Portnoy ME, Cutler DJ, Green ED, and Chakravarti A (2005). A common sex-dependent mutation in a ret enhancer underlies hirschsprung disease risk. Nature 434, 857–863. 10.1038/nature03467. [PubMed: 15829955]

26. Chatterjee S, Kapoor A, Akiyama JA, Auer DR, Lee D, Gabriel S, Berrios C, Pennacchio LA, and Chakravarti A (2016). Enhancer variants synergistically drive dysfunction of a gene regulatory network in hirschsprung disease. Cell 167, 355–368.e10. 10.1016/j.cell.2016.09.005. [PubMed: 27693352]

27. Kobayashi KS, and van den Elsen PJ (2012). Nlrc5: a key regulator of mhc class i-dependent immune responses. Nat. Rev. Immunol 12, 813–820. 10.1038/nri3339. [PubMed: 23175229]

28. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh PR, Anttila V, Xu H, Zang C, Farh K, et al. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet 47, 1228–1235. 10.1038/ng.3404. [PubMed: 26414678]

29. Gazal S, Finucane HK, Furlotte NA, Loh PR, Palamara PF, Liu X, Schoech A, Bulik-Sullivan B, Neale BM, Gusev A, and Price AL (2017). Linkage disequilibrium–dependent architecture of human complex traits shows action of negative selection. Nat. Genet 49, 1421–1427. 10.1038/ng.3954. [PubMed: 28892061]

30. Gazal S, Marquez-Luna C, Finucane HK, and Price AL (2019). Reconciling S-LDSC and LDAK functional enrichment estimates. Nat. Genet 51, 1202–1204. 10.1038/s41588-019-0464-1. [PubMed: 31285579]

31. Consortium GP, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, and Abecasis GR (2015). A global reference for human genetic variation. Mol. Cell 526, 68–74. 10.1038/nature15393.

32. Fulco CP, Nasser J, Jones TR, Munson G, Bergman DT, Subramanian V, Grossman SR, Anyoha R, Doughty BR, Patwardhan TA, et al. (2019). Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. Nat. Genet 51, 1664–1669. 10.1038/s41588-019-0538-0. [PubMed: 31784727]

33. Yoshida H, Lareau CA, Ramirez RN, Rose SA, Maier B, Wroblewska A, Desland F, Chudnovskiy A, Mortha A, Dominguez C, et al. (2019). The cis-regulatory atlas of the mouse immune system. Cell 176, 897–912.e20. 10.1016/j.cell.2018.12.036. [PubMed: 30686579]

34. Schmiedel BJ, Singh D, Madrigal A, Valdovino-Gonzalez AG, White BM, Zapardiel-Gonzalo J, Ha B, Altay G, Greenbaum JA, McVicker G, et al. (2018). Impact of genetic polymorphisms on human immune cell gene expression. Cell 175, 1701–1715.e16. 10.1016/j.cell.2018.10.022. [PubMed: 30449622]

35. Javierre BM, Burren OS, Wilder SP, Kreuzhuber R, Hill SM, Sewitz S, Cairns J, Wingett SW, Várnai C, Thiecke MJ, et al. (2016). Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. Cell 167, 1369–1384.e19. 10.1016/j.cell.2016.09.037. [PubMed: 27863249]

36. Consortium GTEx; Battle A, Brown CD, Engelhardt BE, and Montgomery SB (2017). Genetic effects on gene expression across human tissues. Nature 550, 204–213. 10.1038/nature24277. [PubMed: 29022597]

37. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, Chen X, Taipale J, Hughes TR, and Weirauch MT (2018). The human transcription factors. Cell 172, 650–665. 10.1016/j.cell.2018.01.029. [PubMed: 29425488]

38. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, et al. (2017). The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. Nucleic Acids Res. 45, D362–D368. 10.1093/nar/gkw937. [PubMed: 27924014]

39. Nasser J, Bergman DT, Fulco CP, Guckelberger P, Doughty BR, Patwardhan TA, Jones TR, Nguyen TH, Ulirsch JC, Lekschas F, et al. (2021). Genome-wide enhancer maps link risk variants to disease genes. Nature 593, 238–243. 10.1038/s41586-021-03446-x. [PubMed: 33828297]

40. Hoffman MM, Ernst J, Wilder SP, Kundaje A, Harris RS, Libbrecht M, Giardine B, Ellenbogen PM, Bilmes JA, Birney E, et al. (2012). Integrative annotation of chromatin elements from ENCODE data. Nucleic Acids Res. 41, 827–841. 10.1093/nar/gks1284. [PubMed: 23221638]

41. Hoffman MM, Buske OJ, Wang J, Weng Z, Bilmes JA, and Noble WS (2012). Unsupervised pattern discovery in human chromatin structure through genomic segmentation. Nat. Methods 9, 473–476. 10.1038/nmeth.1937. [PubMed: 22426492]

42. Hormozdiari F, Gazal S, van de Geijn B, Finucane HK, Ju CJT, Loh PR, Schoech A, Reshef Y, Liu X, O'Connor L, et al. (2018). Leveraging molecular quantitative trait loci to understand the genetic architecture of diseases and complex traits. Nat. Genet 50, 1041–1047. 10.1038/s41588-018-0148-2. [PubMed: 29942083]

43. Liu Y, Sarkar A, Kheradpour P, Ernst J, and Kellis M (2017). Evidence of reduced recombination rate in human regulatory domains. Genome Biol. 18, 193. 10.1186/s13059-017-1308-x. [PubMed: 29058599]

44. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, 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. 10.1038/nature09906. [PubMed: 21441907]

45. Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, Mendez D, Mutowo P, Atkinson F, Bellis LJ, Cibrián-Uhalte E, et al. (2016). The ChEMBL database in 2017. Nucleic Acids Res. 45, D945–D954. 10.1093/nar/gkw1074. [PubMed: 27899562]

46. Freund MK, Burch KS, Shi H, Mancuso N, Kichaev G, Garske KM, Pan DZ, Miao Z, Mohlke KL, Laakso M, et al. (2018). Phenotype-specific enrichment of Mendelian disorder genes near GWAS regions across 62 complex traits. Am. J. Hum. Genet 103, 535–552. 10.1016/j.ajhg.2018.08.017. [PubMed: 30290150]

47. Vuckovic D, Bao EL, Akbari P, Lareau CA, Mousas A, Jiang T, Chen MH, Raffield LM, Tardaguila M, Huffman JE, Ritchie SC, Megy K, Ponstingl H, Penkett CJ, Albers PK, Wigdor EM, Sakaue S, Moscati A, Manansala R, Lo KS, Qian H, Akiyama M, Bartz TM, Ben-Shlomo Y, Beswick A, Bork-Jensen J, Bottinger EP, Brody JA, van Rooij FJA, Chitrala KN, Wilson PWF,

Choquet H, Danesh J, Di Angelantonio E, Dimou N, Ding J, Elliott P, Esko T, Evans MK, Felix SB, Floyd JS, Broer L, Grarup N, Guo MH, Guo Q, Greinacher A, Haessler J, Hansen T, Howson JMM, Huang W, Jorgenson E, Kacprowski T, Kähönen M, Kamatani Y, Kanai M, Karthikeyan S, Koskeridis F, Lange LA, Lehtimäki T, Linneberg A, Liu Y, Lyytikäinen LP, Manichaikul A, Matsuda K, Mohlke KL, Mononen N, Murakami Y, Nadkarni GN, Nikus K, Pankratz N, Pedersen O, Preuss M, Psaty BM, Raitakari OT, Rich SS, Rodriguez BAT, Rosen JD, Rotter JI, Schubert P, Spracklen CN, Surendran P, Tang H, Tardif JC, Ghanbari M, Völker U, Völzke H, Watkins NA, Weiss S, VA Million Veteran Program; Cai N, Kundu K, Watt SB, Walter K, Zonderman AB, Cho K, Li Y, Loos RJF, Knight JC, Georges M, Stegle O, Evangelou E, Okada Y, Roberts DJ, Inouye M, Johnson AD, Auer PL, Astle WJ, Reiner AP, Butterworth AS, Ouwehand WH, Lettre G, Sankaran VG, and Soranzo N (2020). The polygenic and monogenic basis of blood traits and diseases. Cell 182, 1214–1231.e11. 10.1016/j.cell.2020.08.008. [PubMed: 32888494]

48. Wright CF, Fitzgerald TW, Jones WD, Clayton S, McRae JF, van Kogelenberg M, King DA, Ambridge K, Barrett DM, Bayzetinova T, et al. (2015). Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. Lancet 385, 1305–1314. 10.1016/s0140-6736(14)61705-0. [PubMed: 25529582]

49. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, et al. (2016). Analysis of protein-coding genetic variation in 60, 706 humans. Nature 536, 285–291. 10.1038/nature19057. [PubMed: 27535533]

50. Schoech AP, Jordan DM, Loh PR, Gazal S, O'Connor LJ, Balick DJ, Palamara PF, Finucane HK, Sunyaev SR, and Price AL (2019). Quantification of frequency-dependent genetic architectures in 25 UK Biobank traits reveals action of negative selection. Nat. Commun 10, 790. 10.1038/s41467-019-08424-6. [PubMed: 30770844]

51. Farh KKH, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, Shoresh N, Whitton H, Ryan RJH, Shishkin AA, et al. (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature 518, 337–343. 10.1038/nature13835. [PubMed: 25363779]

52. Weissbrod O, Hormozdiari F, Benner C, Cui R, Ulirsch J, Gazal S, Schoech AP, van de Geijn B, Reshef Y, Márquez-Luna C, et al. (2020). Functionally-informed fine-mapping and polygenic localization of complex trait heritability. Nat. Genet 52, 1355–1363. 10.1038/s41588-020-00735-5. [PubMed: 33199916]

53. Kamburov A, Stelzl U, Lehrach H, and Herwig R (2012). The ConsensusPathDB interaction database: 2013 update. Nucleic Acids Res. 41, D793–D800. 10.1093/nar/gks1055. [PubMed: 23143270]

54. Tong H, Faloutsos C, and Pan JY (2008). Random walk with restart: fast solutions and applications. Knowl. Inf. Syst 14, 327–346. 10.1007/s10115-007-0094-2.

55. Hormozdiari F, van de Geijn B, Nasser J, Weissbrod O, Gazal S, Ju CJT, Connor LO, Hujoel MLA, Engreitz J, Hormozdiari F, and Price AL (2019). Functional disease architectures reveal unique biological role of transposable elements. Nat. Commun 10, 4054. 10.1038/s41467-019-11957-5. [PubMed: 31492842]

56. Võsa U, Claringbould A, Westra HJ, Bonder MJ, Deelen P, Zeng B, Kirsten H, Saha A, Kreuzhuber R, Yazar S, Brugge H, Oelen R, de Vries DH, van der Wijst MGP, Kasela S, Pervjakova N, Alves I, Favé MJ, Agbessi M, Christiansen MW, Jansen R, Seppälä I, Tong L, Teumer A, Schramm K, Hemani G, Verlouw J, Yaghootkar H, Sönmez Flitman R, Brown A, Kukushkina V, Kalnapenkis A, Rüeger S, Porcu E, Kronberg J, Kettunen J, Lee B, Zhang F, Qi T, Hernandez JA, Arindrarto W, Beutner F, BIOS Consortium; i2QTL Consortium, Dmitrieva J, Elansary M, Fairfax BP, Georges M, Heijmans BT, Hewitt AW, Kähönen M, Kim Y, Knight JC, Kovacs P, Krohn K, Li S, Loeffler M, Marigorta UM, Mei H, Momozawa Y, Müller-Nurasyid M, Nauck M, Nivard MG, Penninx BWJH, Pritchard JK, Raitakari OT, Rotzschke O, Slagboom EP, Stehouwer CDA, Stumvoll M, Sullivan P, 't Hoen PAC, Thiery J, Tönjes A, van Dongen J, van Iterson M, Veldink JH, Völker U, Warmerdam R, Wijmenga C, Swertz M, Andiappan A, Montgomery GW, Ripatti S, Perola M, Kutalik Z, Dermitzakis E, Bergmann S, Frayling T, van Meurs J, Prokisch H, Ahsan H, Pierce BL, Lehtimäki T, Boomsma DI, Psaty BM, Gharib SA, Awadalla P, Milani L, Ouwehand WH, Downes K, Stegle O, Battle A, Visscher PM, Yang J, Scholz M, Powell J, Gibson G, Esko T, and Franke L (2021). Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. Nat. Genet 53, 1300–1310. 10.1038/s41588-021-00913-z. [PubMed: 34475573]

57. Cai W, Zhou W, Han Z, Lei J, Zhuang J, Zhu P, Wu X, and Yuan W (2020). Master Regulator Genes and Their Impact on Major Diseases. PeerJ 8, e9952. [PubMed: 33083114]

58. Nakamura MC (2014). CIITA: a master regulator of adaptive immunity shows its innate side in the bone. J. Bone Miner. Res 29, 287–289. 10.1002/jbmr.2161. [PubMed: 24343972]

59. Colomer C, Marruecos L, Vert A, Bigas A, and Espinosa L (2017). NF-κB members left home: NF-κB-Independent roles in cancer. Biomedicines 5, 26. 10.3390/biomedicines5020026.

60. Bresnick EH, Katsumura KR, Lee HY, Johnson KD, and Perkins AS (2012). Master regulatory GATA transcription factors: mechanistic principles and emerging links to hematologic malignancies. Nucleic Acids Res. 40, 5819–5831. 10.1093/nar/gks281. [PubMed: 22492510]

61. Paul S, Home P, Bhattacharya B, and Ray S (2017). GATA factors: master regulators of gene expression in trophoblast progenitors. Placenta 60, S61–S66. 10.1016/j.placenta.2017.05.005. [PubMed: 28526138]

62. Chikuma S (2017). Ctla-4, an essential immune-checkpoint for t-cell activation. Curr. Top. Microbiol. Immunol 410, 99–126. 10.1007/82_2017_61. [PubMed: 28900679]

63. Zhao Y, Yang W, Huang Y, Cui R, Li X, and Li B (2018). Evolving roles for targeting ctla-4 in cancer immunotherapy. Cell. Physiol. Bio-chem 47, 721–734. 10.1159/000490025.

64. Liu F, Huang J, Liu X, Cheng Q, Luo C, and Liu Z (2020). Ctla-4 correlates with immune and clinical characteristics of glioma. Cancer Cell Int. 20, 7. 10.1186/s12935-019-1085-6. [PubMed: 31911758]

65. Richer MJ, Lang ML, and Butler NS (2016). T cell fates zipped up: how the bach2 basic leucine zipper transcriptional repressor directs t cell differentiation and function. J. Immunol 197, 1009–1015. 10.4049/jimmunol.1600847. [PubMed: 27496973]

66. Zhang H, Hu Q, Zhang M, Yang F, Peng C, Zhang Z, and Huang C (2019). Bach2 deficiency leads to spontaneous expansion of il-4-producing t follicular helper cells and autoimmunity. Front. Immunol 10, 2050. 10.3389/fimmu.2019.02050. [PubMed: 31552021]

67. Roychoudhuri R, Hirahara K, Mousavi K, Clever D, Klebanoff CA, Bonelli M, Sciumè G, Zare H, Vahedi G, Dema B, et al. (2013). Bach2 represses effector programs to stabilize t reg-mediated immune homeostasis. Nature 498, 506–510. 10.1038/nature12199. [PubMed: 23728300]

68. Cooper JD, Smyth DJ, Smiles AM, Plagnol V, Walker NM, Allen JE, Downes K, Barrett JC, Healy BC, Mychaleckyj JC, et al. (2008). Meta-analysis of genome-wide association study data identifies additional type 1 diabetes risk loci. Nat. Genet 40, 1399–1401. 10.1038/ng.249. [PubMed: 18978792]

69. Ferreira MA, Matheson MC, Duffy DL, Marks GB, Hui J, Le Souëf P, Danoy P, Baltic S, Nyholt DR, Jenkins M, et al. (2011). Identification of il6r and chromosome 11q13. 5 as risk loci for asthma. Lancet 378, 1006–1014. 10.1016/s0140-6736(11)60874-x. [PubMed: 21907864]

70. Morris DL, Sheng Y, Zhang Y, Wang YF, Zhu Z, Tombleson P, Chen L, Cunninghame Graham DS, Bentham J, Roberts AL, et al. (2016). Genome-wide association meta-analysis in Chinese and european individuals identifies ten new loci associated with systemic lupus erythe-matosus. Nat. Genet 48, 940–946. 10.1038/ng.3603. [PubMed: 27399966]

71. Oeckinghaus A, and Ghosh S (2009). The NF- B family of transcription factors and its regulation. Cold Spring Harbor Perspect. Biol 1, a000034. 10.1101/cshperspect.a000034.

72. Grumont RJ, Rourke IJ, O'Reilly LA, Strasser A, Miyake K, Sha W, and Gerondakis S (1998). B lymphocytes differentially use the rel and nuclear factor κB1 (NF-κB1) transcription factors to regulate cell cycle progression and apoptosis in quiescent and mitogen-activated cells. J. Exp. Med 187, 663–674. 10.1084/jem.187.5.663. [PubMed: 9480976]

73. Gerondakis S, and Siebenlist U (2010). Roles of the NF- B pathway in lymphocyte development and function. Cold Spring Harbor Perspect. Biol 2, a000182. 10.1101/cshperspect.a000182.

74. Hujoel ML, Gazal S, Hormozdiari F, van de Geijn B, and Price AL (2019). Disease heritability enrichment of regulatory elements is concentrated in elements with ancient sequence age and conserved function across species. Am. J. Hum. Genet 104, 611–624. 10.1016/j.ajhg.2019.02.008. [PubMed: 30905396]

75. Dey KK, van de Geijn B, Kim SS, Hormozdiari F, Kelley DR, and Price AL (2020). Evaluating the informativeness of deep learning annotations for human complex diseases. Nat. Commun 11, 4703. 10.1038/s41467-020-18515-4. [PubMed: 32943643]

76. de Leeuw CA, Mooij JM, Heskes T, and Posthuma D (2015). MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput. Biol 11, e1004219. 10.1371/journal.pcbi.1004219. [PubMed: 25885710]

77. Daly C, and Rollins B (2003). Monocyte chemoattractant protein-1 (ccl2) in inflammatory disease and adaptive immunity: therapeutic opportunities and controversies. Microcirculation 10, 247–257. 10.1080/713773639. [PubMed: 12851642]

78. Plskova J, Greiner K, Muckersie E, Duncan L, and Forrester JV (2006). Interferon-α: a key factor in autoimmune disease. Microcirculation 47, 3946. 10.1167/iovs.06-0058.

79. Cardinez C, Miraghazadeh B, Tanita K, da Silva E, Hoshino A, Okada S, Chand R, Asano T, Tsumura M, Yoshida K, et al. (2018). Gain-of-function ikbkb mutation causes human combined immune deficiency. J. Exp. Med 215, 2715–2724. 10.1084/jem.20180639. [PubMed: 30337470]

80. Jacobs J, Deschoolmeester V, Zwaenepoel K, Rolfo C, Silence K, Rottey S, Lardon F, Smits E, and Pauwels P (2015). Cd70: an emerging target in cancer immunotherapy. Pharmacol. Therapeut 155, 1–10. 10.1016/j.pharmthera.2015.07.007.

81. Shaffer DR, Savoldo B, Yi Z, Chow KKH, Kakarla S, Spencer DM, Dotti G, Wu MF, Liu H, Kenney S, and Gottschalk S (2011). T cells redirected against CD70 for the immunotherapy of CD70-positive malignancies. Blood 117, 4304–4314. 10.1182/blood-2010-04-278218. [PubMed: 21304103]

82. Verhoeven Y, Tilborghs S, Jacobs J, De Waele J, Quatannens D, Deben C, Prenen H, Pauwels P, Trinh XB, Wouters A, et al. (2020). The potential and controversy of targeting stat family members in cancer. Semin. Cancer Biol 60, 41–56. 10.1016/j.semcancer.2019.10.002. [PubMed: 31605750]

83. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, et al. (2020). The mutational constraint spectrum quantified from variation in 141, 456 humans. Nature 581, 434–443. 10.1038/s41586-020-2308-7. [PubMed: 32461654]

84. Minikel EV, Karczewski KJ, Martin HC, Cummings BB, Whiffin N, Rhodes D, Alfoldi J, Trembath RC, van Heel DA, Daly MJ, Schreiber SL, and MacArthur DG (2020). Evaluating drug targets through human loss-of-function genetic variation. Nature 581, 459–464. 10.1038/s41586-020-2267-z. [PubMed: 32461653]

85. Jagadeesh K, Dey K, Montoro DT, Gazal S, Engreitz JM, Xavier RJ, Price AL, and Regev A (2021). Identifying disease-critical cell types and cellular processes across the human body by integration of single-cell profiles and human genetics. Preprint at bioRxiv. 10.1101/2021.03.19.436212.

86. Mancuso N, Freund MK, Johnson R, Shi H, Kichaev G, Gusev A, and Pasaniuc B (2019). Probabilistic fine-mapping of transcriptome-wide association studies. Nat. Genet 51, 675–682. 10.1038/s41588-019-0367-1. [PubMed: 30926970]

87. Weeks E, Ulirsch JC, Cheng NY, Trippe BL, Fine RS, Miao J, Patwardhan TA, Kanai M, Nasser J, Fulco CP, and Tashman KC (2020). Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. Preprint at medRxiv. 10.1101/2020.09.08.20190561.

88. Kichaev G, Yang WY, Lindstrom S, Hormozdiari F, Eskin E, Price AL, Kraft P, and Pasaniuc B (2014). Integrating functional data to prioritize causal variants in statistical fine-mapping studies. PLoS Genet. 10, e1004722. 10.1371/journal.pgen.1004722. [PubMed: 25357204]

89. Chen W, McDonnell SK, Thibodeau SN, Tillmans LS, and Schaid DJ (2016). Incorporating functional annotations for fine-mapping causal variants in a Bayesian framework using summary statistics. Genetics 204, 933–958. 10.1534/genetics.116.188953. [PubMed: 27655946]

90. Kichaev G, Roytman M, Johnson R, Eskin E, Lindström S, Kraft P, and Pasaniuc B (2017). Improved methods for multi-trait fine mapping of pleiotropic risk loci. Bioinformatics 33, 248–255. 10.1093/bioinformatics/btw615. [PubMed: 27663501]

91. Ray JP, de Boer CG, Fulco CP, Lareau CA, Kanai M, Ulirsch JC, Tewhey R, Ludwig LS, Reilly SK, Bergman DT, et al. (2020). Prioritizing disease and trait causal variants at the TNFAIP3 locus using functional and genomic features. Nat. Commun 11, 1237. 10.1038/s41467-020-15022-4. [PubMed: 32144282]

92. Hu Y, Lu Q, Powles R, Yao X, Yang C, Fang F, Xu X, and Zhao H (2017). Leveraging functional annotations in genetic risk prediction for human complex diseases. PLoS Comput. Biol 13, e1005589. 10.1371/journal.pcbi.1005589. [PubMed: 28594818]

93. Márquez-Luna C, Gazal S, Loh PR, Kim SS, Furlotte N, Auton A, Agee M, Alipanahi B, Bell RK, Bryc K, et al. (2021). Incorporating functional priors improves polygenic prediction accuracy in UK Biobank and 23andMe data sets. Nat. Commun 12, 6052. 10.1038/s41467-021-25171-9. [PubMed: 34663819]

94. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature 562, 203–209. 10.1038/s41586-018-0579-z. [PubMed: 30305743]

95. Consortium, 1000G, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, and Abecasis GR (2015). A global reference for human genetic variation. Nature 526, 74. 10.1038/nature15393.

96. Loh PR, Tucker G, Bulik-Sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, Chasman DI, Ridker PM, Neale BM, Berger B, et al. (2015). Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat. Genet 47, 284–290. 10.1038/ng.3190. [PubMed: 25642633]

97. Kinsella R, Kähäri A, Haider S, Zamora J, Proctor G, Spudich G, Almeida-King J, Staines D, Derwent P, Kerhornou A, and Kersey P (2011). Ensembl BioMarts: A hub for data retrieval across taxonomic space. Database 2011, bar030. [PubMed: 21785142]

98. Kundaje A, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature 518, 317–330. 10.1038/nature14248. [PubMed: 25693563]

99. Ong CT, and Corces VG (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat. Rev. Genet 12, 283–293. 10.1038/nrg2957. [PubMed: 21358745]

100. Ko JY, Oh S, and Yoo KH (2017). Functional enhancers as master regulators of tissue-specific gene regulation and cancer development. Mol. Cell 40, 169–177. 10.14348/molcells.2017.0033.

101. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, et al. (2014). STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 43, D447–D452. 10.1093/nar/gku1003. [PubMed: 25352553]

102. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, and Haussler D (2002). The human genome browser at ucsc. Genome Res. 12, 996–1006. 10.1101/gr.229102. [PubMed: 12045153]

103. Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, and Kent WJ (2004). The ucsc table browser data retrieval tool. Nucleic Acids Res. 32, D493–D496. 10.1093/nar/gkh103. [PubMed: 14681465]

104. Amemiya HM, Kundaje A, and Boyle AP (2019). The ENCODE blacklist: identification of problematic regions of the genome. Sci. Rep 9, 9354. 10.1038/s41598-019-45839-z. [PubMed: 31249361]

105. Consortium EP (2012). An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74. 10.1038/nature11247. [PubMed: 22955616]

106. Moonen J-RA, Chappell J, Shi M, Shinohara T, Li D, Mumbach MR, Zhang F, Nasser J, Mai DH, Taylor S, and Wang L (2020). KLF4 recruits SWI/SNF to increase chromatin accessibility and reprogram the endothelial enhancer landscape under laminar shear stress. Preprint at bioRxiv. 10.1101/2020.07.10.195768.

107. Gazal S, Loh PR, Finucane HK, Ganna A, Schoech A, Sunyaev S, and Price AL (2018). Functional architecture of low-frequency variants highlights strength of negative selection across coding and non-coding annotations. Nat. Genet 50, 1600–1607. 10.1038/s41588-018-0231-8. [PubMed: 30297966]

108. Palamara PF, Terhorst J, Song YS, and Price AL (2018). High-throughput inference of pairwise coalescence times identifies signals of selection and enriched disease heritability. Nat. Genet 50, 1311–1317. 10.1038/s41588-018-0177-x. [PubMed: 30104759]

109. Kim SS, Dey KK, Weissbrod O, Márquez-Luna C, Gazal S, and Price AL (2020). Improving the informativeness of Mendelian disease-derived pathogenicity scores for common disease. Nat. Commun 11, 6258. 10.1038/s41467-020-20087-2. [PubMed: 33288751]

110. van de Geijn B, Finucane H, Gazal S, Hormozdiari F, Amariuta T, Liu X, Gusev A, Loh PR, Reshef Y, Kichaev G, et al. (2020). Annotations capturing cell-type-specific TF binding explain a large fraction of disease heritability. Hum. Mol. Genet 29, 1057–1067. 10.1093/hmg/ddz226. [PubMed: 31595288]

111. Speed D, Holmes J, and Balding DJ (2020). Evaluating and improving heritability models using summary statistics. Nat. Genet 52, 458–462. 10.1038/s41588-020-0600-y. [PubMed: 32203469]

112. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, Kang HM, Fuchsberger C, Danecek P, Sharp K, et al. (2016). A reference panel of 64, 976 haplotypes for genotype imputation. Nat. Genet 48, 1279–1283. 10.1038/ng.3643. [PubMed: 27548312]

**Highlights**

- Enhancer-related and master-regulator genes provide a unique heritability signal

- Integration with protein-protein interaction networks further magnifies the signal

- The identified specific gene programs are highly enriched for immune drug target genes

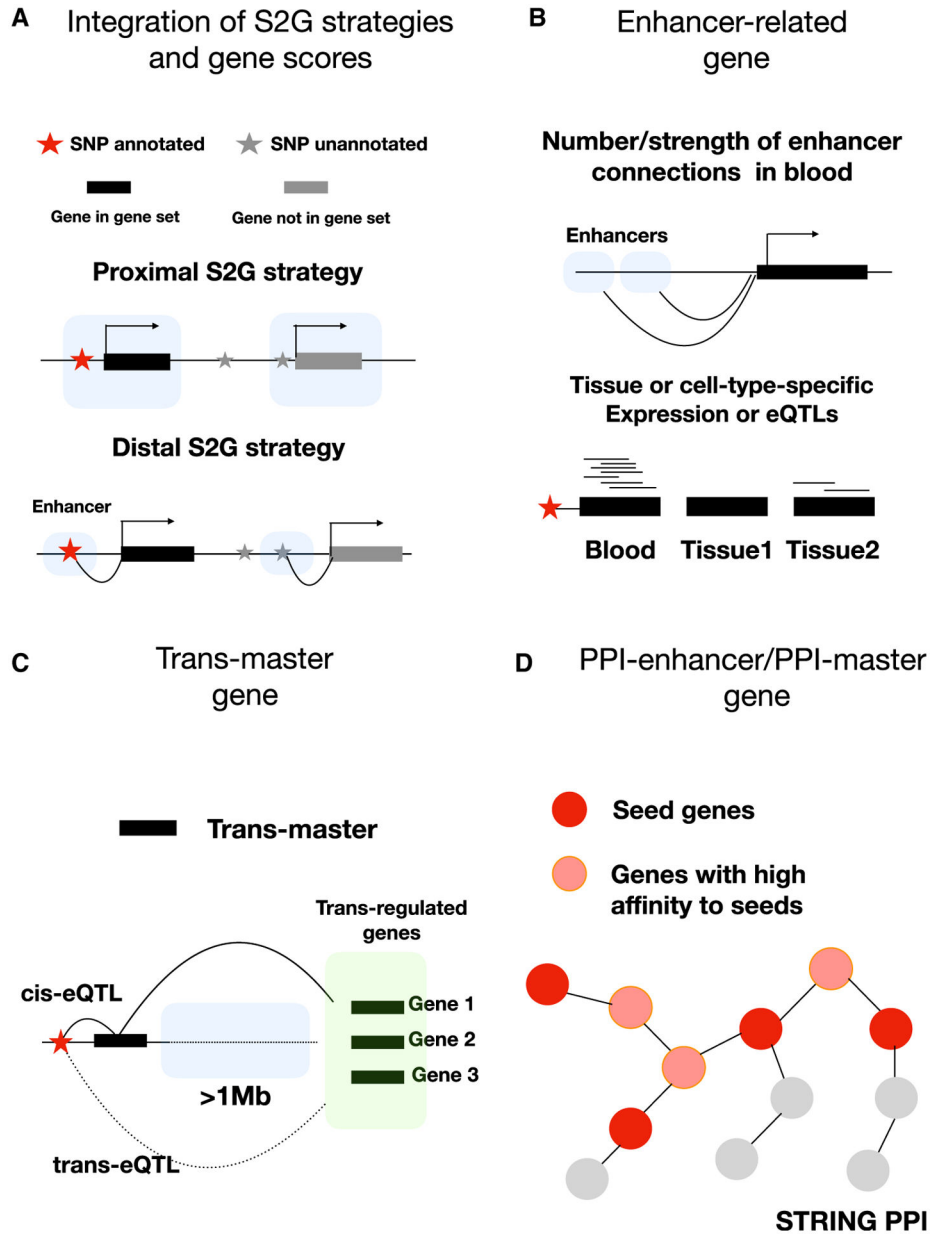- Functional SNP-to-gene (ABC, Roadmap, etc.) linking strategies drive the signal

**Figure 1. Illustration of S2G strategies and gene scores**

(A) SNP annotations defined by integration of genes in gene set with proximal (close to gene body) and distal S2G strategies.

(B) Examples of approaches used to define enhancer-related genes.

(C) A *Trans*-master gene regulates multiple distal genes via a *cis*-eQTL that is a *trans*-eQTL of the distal genes.

(D) PPI-enhancer genes have high connectivity to enhancer-related genes in a PPI network.

**A** ENR, meta-analyzed across 11 blood-related traits
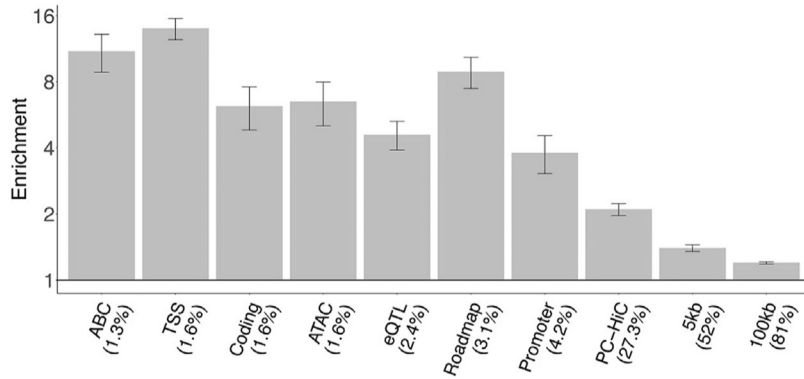
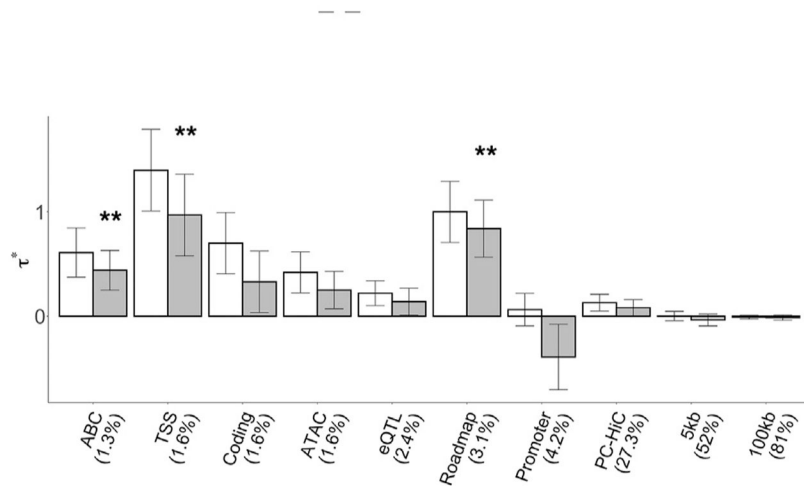**B** $\tau^*$, meta-analyzed across 11 blood-related traits

**Figure 2. Disease informativeness of S2G annotations**

We evaluated ten S2G annotations defined from the corresponding S2G strategies by SNPs linked to the set of all genes.

(A) Heritability enrichment (log scale), conditional on the baseline-LD model. Horizontal line denotes no enrichment.(B) Standardized effect size ($\tau^*$), conditional on either the baseline-LD model (marginal analyses: left column, white) or the baseline-LD+ model, which includes all ten S2G annotations (right column, dark shading).

Results are meta-analyzed across 11 blood-related traits. **$p < 0.05/10$. Error bars denote 95% confidence intervals. Numerical results are reported in Tables S2 and S4.
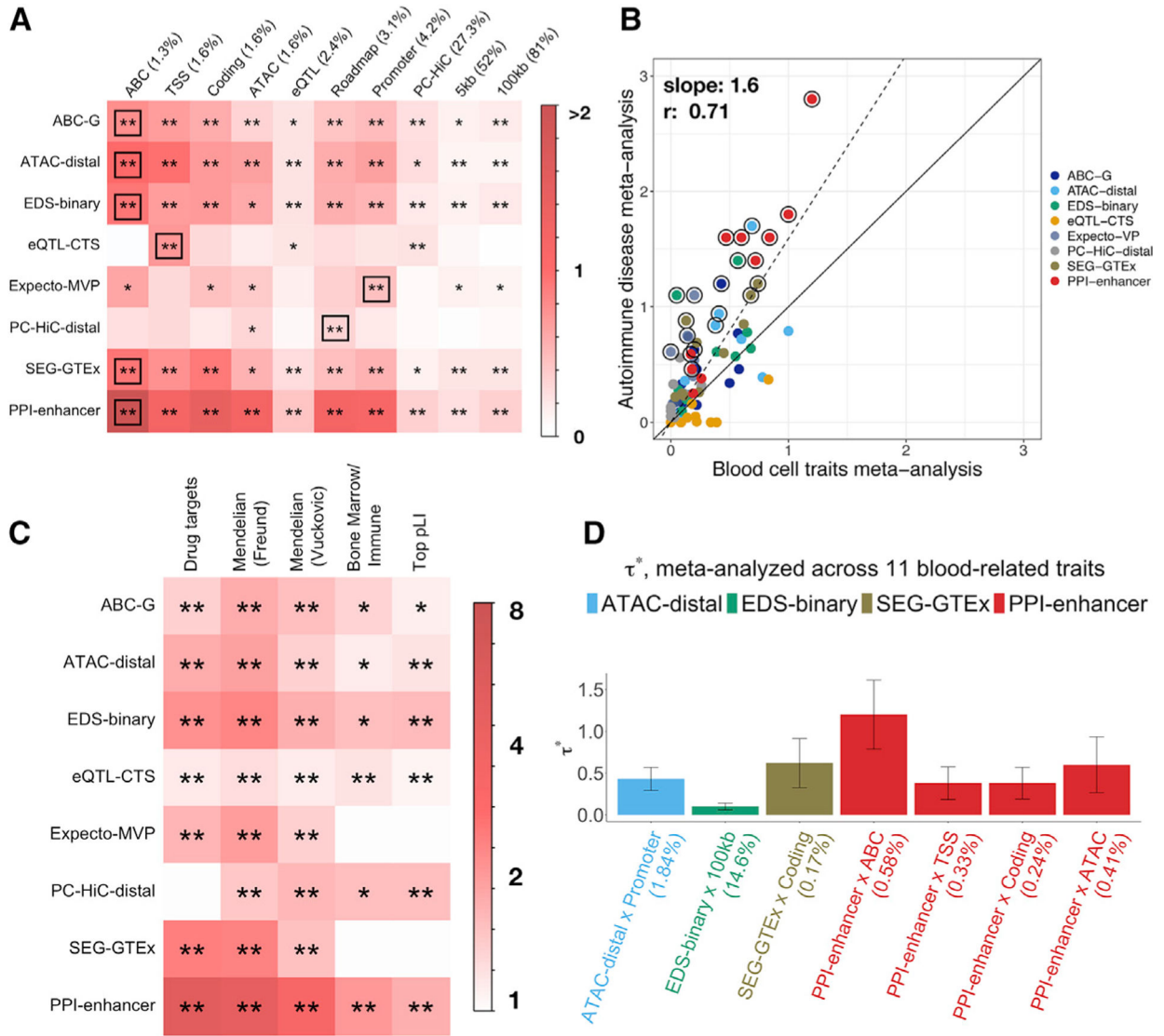
**Figure 3. Disease informativeness of enhancer-related and PPI-enhancer annotations**

We evaluated 80 annotations constructed by combining seven enhancer-related + 1 PPI-enhancer gene scores with ten S2G strategies.

(A) Standardized effect size (τ*), conditional on the baseline-LD+ model.

(B) Comparison of meta-analyzed standardized effect size (τ*) across six autoimmune diseases versus five blood cell traits.

(C) Enrichment of enhancer-related and PPI-enhancer genes in five "gold-standard" disease-related gene sets.

(D) Standardized effect size (τ*), conditional on the baseline-LD+ model plus seven jointly significant enhancer-related + PPI-enhancer annotations.

In (A) and (D), results are meta-analyzed across 11 blood-related traits. In (A) and (C), double asterisks denote Bonferroni-significant p values (**p < 0.05/110 in A and **p < 0.05/55 in C) and single asterisk (*) denotes FDR < 0.05. In (A), the black box in each row denotes the S2G strategy with highest τ*. In (B), circled dots denote annotations

with significant (FDR < 5%) difference in effect size between the two meta-analyses, the solid line denotes y = x, and the dashed line denotes the regression slope. We report the slope of the regression and the Pearson correlation for enhancer-related and PPI-enhancer annotations (slope = 1.3, r = 0.57 for enhancer-related annotations only). Error bars in (D) denote 95% confidence intervals. Numerical results are reported in Tables S6, S8, S10, S11, S23, and S26.
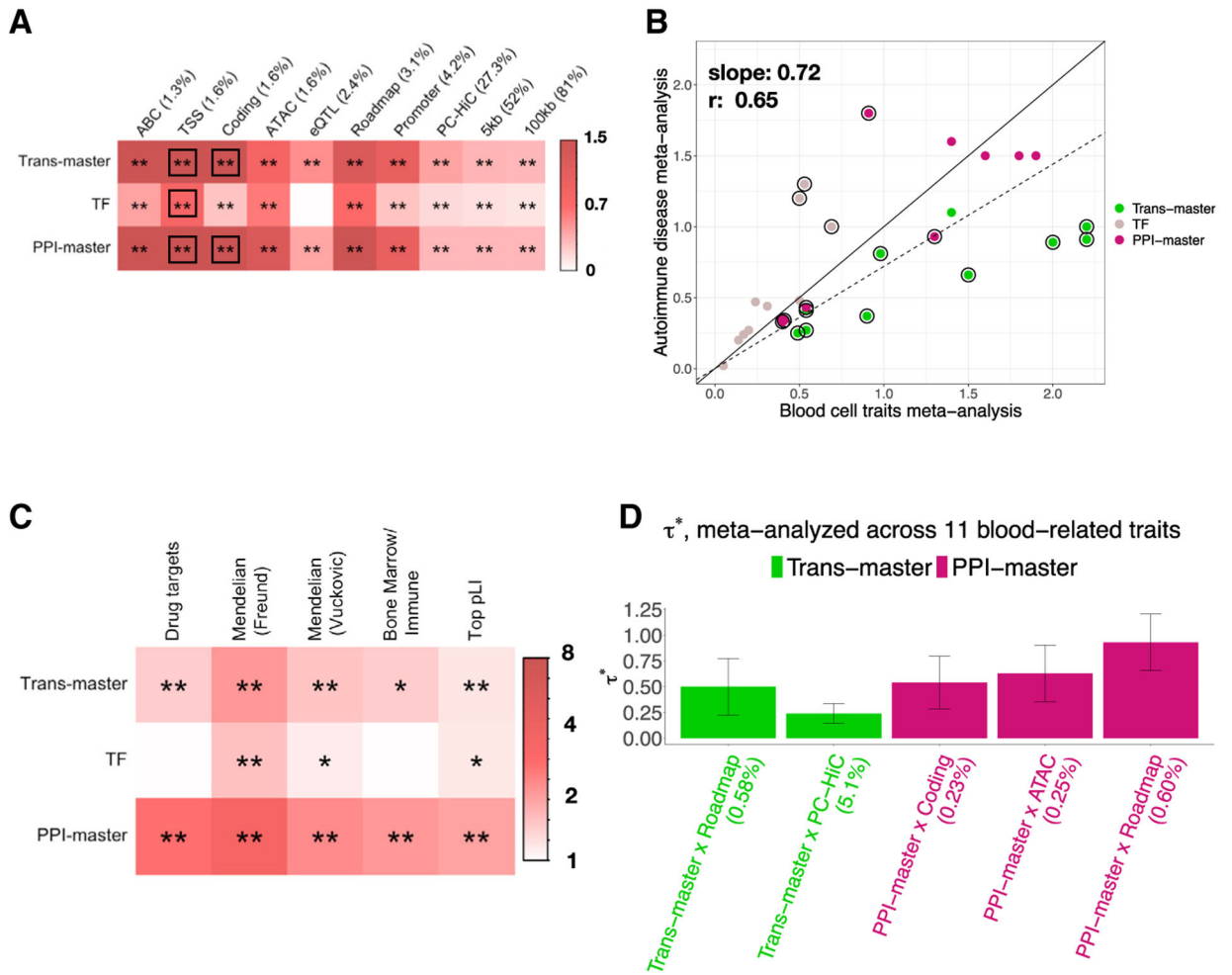
**Figure 4. Disease informativeness of master-regulator and PPI-master annotations**

We evaluated 30 annotations constructed by combining two master-regulator + 1 PPI-master gene scores with ten S2G strategies.

(A) Standardized effect size ($\tau^*$), conditional on the 113 baseline-LD+ *cis* model annotations.

(B) Comparison of meta-analyzed standardized effect size ($\tau^*$) across six autoimmune diseases versus five blood cell traits.

(C) Enrichment of master-regulator and PPI-master genes in five "gold-standard" disease-related gene sets.

(D) Standardized effect size ($\tau^*$), conditional on the baseline-LD+ *cis* model plus five jointly significant master-regulator + PPI-master annotations.

In (A) and (D), results are meta-analyzed across 11 blood-related traits. In (A) and (C), double asterisks denote Bonferroni-significant p values (**p < 0.05/110 in A and **p < 0.05/55 in C), and single asterisk (*) denotes FDR < 0.05. In (A), the black box in each row denotes the S2G strategy with highest $\tau^*$. In (B), circled dots denote annotations with significant (FDR < 5%) difference in effect size between the two meta-analyses, the solid line denotes y = x, and the dashed line denotes the regression slope. We report the slope of the regression and the Pearson correlation for master-regulator and PPI-master annotations

(slope = 0.57, r = 0.56 for master-regulator annotations only). Error bars in (D) denote 95% confidence intervals. Numerical results are reported in Tables S6, S29, S38, S39, S47, and S51.
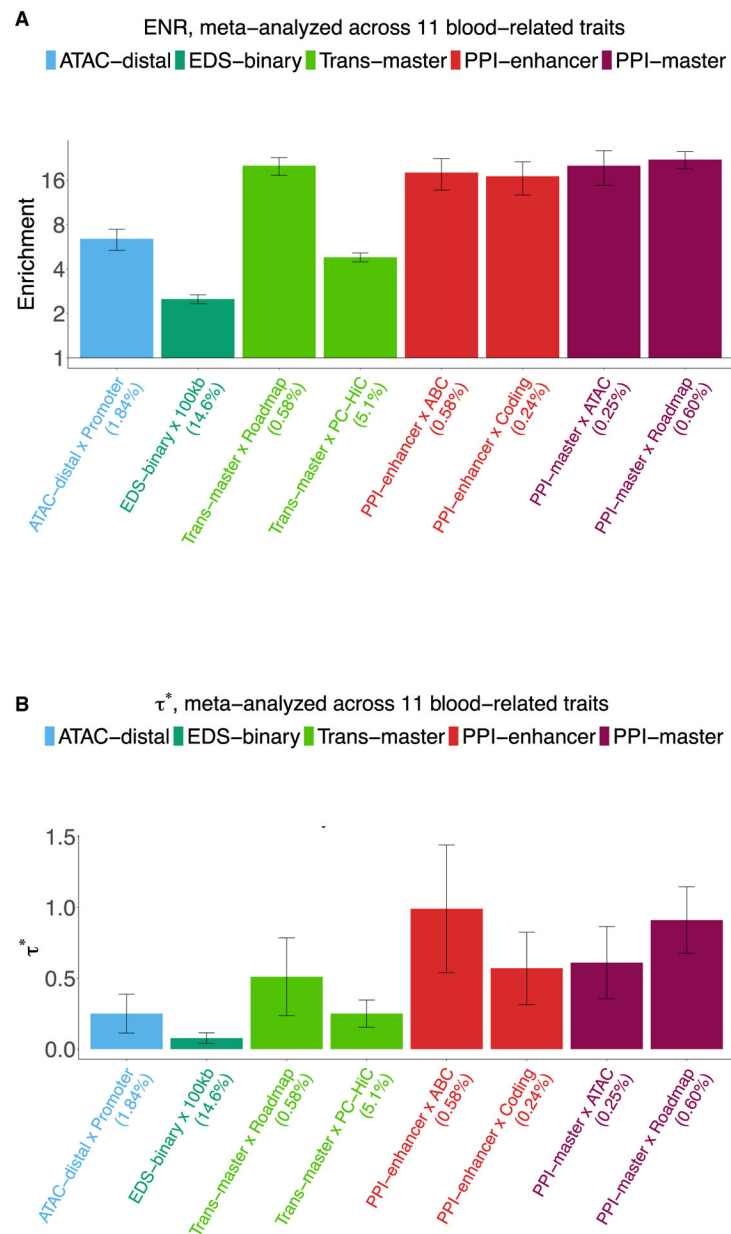
**Figure 5. Combined joint model**

A) Heritability enrichment (log scale) of the eight jointly significant enhancer-related, master-regulator, PPI-enhancer-related, and PPI-master-regulator annotations, conditional on the baseline-LD+ *cis* model. Horizontal line denotes no enrichment.

(B) Standardized effect size (τ*) conditional on the baseline-LD+ *cis* model plus the eight jointly significant annotations.

Significance is corrected for multiple testing by Bonferroni correction (p < 0.05/110). Errors bars denote 95% confidence intervals. Numerical results are reported in Table S52.

**Table 1.**

List of 11 gene scores

| Gene score | Description | Size (%) |
|---|---|---|
| Enhancer-related gene scores | | |
| ABC-G | genes in top 10% genes of number of genic and intergenic enhancer-gene connections in blood, assessed using Activity-By-Contact[32] | 10 |
| ATAC-distal | proportion of mouse gene expression variability across immune cell types explained by distal ATAC-seq peaks[33] | 29 |
| EDS-binary | genes in top 10% of blood-specific enhancer domain score, reflecting the number of bases in enhancers linked to a gene[24] | 10 |
| eQTL-CTS | proportion of eQTLs[34] (FDR < 0.05) that are specific to a single cell type (union across blood cell types) | 32 |
| Expecto-MVP | genes in top 10% of magnitude of variation potential, based on Expecto predictions of regions surrounding the transcription start site[7] | 10 |
| PC-HiC-distal | genes in top 10% genes of number of distal promoter-capture HiC connections in blood cell types[35] | 10 |
| SEG-GTEx | specifically expressed genes[9] in GTEx whole blood[36] | 10 |
| Candidate master-regulator gene scores | | |
| *Trans*-master | genes that significantly *trans*-regulate ≥ 3 genes by any significant *cis*-eQTL of the focal gene | 10 |
| TF | curated list of human transcription factor genes[37] | 7.4 |
| PPI-network-based gene score | | |
| PPI-enhancer | genes with high network connectivity to enhancer-related genes in STRING[38] PPI network | 10 |
| PPI-master | genes with high network connectivity to candidate master-regulator genes in STRING[38] PPI network | 10 |

For each gene score, including seven enhancer-related genes scores, two candidate master-regulator gene scores, and two PPI-network informed gene scores (corresponding to enhancer-related and candidate master-regulator gene scores), we provide a brief description and report its size (average gene score across 22,020 genes; equal to percentage of genes for binary gene scores). Gene scores are listed alphabetically within each category. All gene scores are binary except ATAC-distal and eQTL-CTS, which are probabilistic. Density plots of the distribution of the metric underlying each gene score are provided in Figure S1. Further details are provided in the STAR Methods.

**Table 2.**

List of ten S2G strategies

| S2G strategy | Description | Distal/proximal | Size (%) |
|---|---|---|---|
| ABC | intergenic SNPs with distal enhancer-gene connections, assessed by Activity-By-Contact[32,39] across blood cell types | Distal | 1.4 |
| TSS | SNPs in predicted transcription start sites[40,41] overlapping Ensembl gene ±5 kb window | Proximal | 1.6 |
| Coding | SNPs in coding regions | proximal | 1.6 |
| ATAC | SNPs in ATAC-seq peaks >50% correlated to mouse expression across blood cell types[33] (mapped to human) | distal | 1.6 |
| eQTL | SNPs with fine-mapped causal posterior probability[42] >0.001 in GTEx whole blood | distal + proximal | 2.4 |
| Roadmap | SNPs in predicted enhancer-gene links, assessed using Roadmap Epigenomics Project data[43,44] | distal | 3.2 |
| Promoter | SNPs in promoter regions | proximal | 4.3 |
| PC-HiC | distal SNPs with promoter-capture HiC[35] connections to promoter regions in blood cell types | distal | 27 |
| 5 kb | SNPs in ±5 kb window around gene body | proximal | 53 |
| 100 kb | SNPs in ±100 kb window around gene body[8,9,11] | distal | 81 |

For each S2G strategy, we provide a brief description, indicate whether the S2G strategy prioritizes distal or proximal SNPs relative to the gene, and report its size (% of SNPs linked to genes). S2G strategies are listed in order of increasing size. Further details are provided in STAR Methods.

KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Deposited data | | |
| Gene scores, S2G links, annotations | This paper | https://data.broadinstitute.org/alkesgroup/LDSCORE/Dey_Enhancer_MasterReg |
| UK Biobank summary statistics | Bycroft et al. 2018 Nat Genet[94] | https://data.broadinstitute.org/alkesgroup/UKBB/ |
| 1000 Genomes Project Phase 3 data | 1000G Consortium 2015 Nature[95] | ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502 |
| Baseline-LD model annotations | Finucane et al. 2015 Nat Genet[28] | https://data.broadinstitute.org/alkesgroup/LDSCORE/ |
| Software and algorithms | | |
| GSSG algorithm | This paper | https://github.com/kkdey/GSSG Zenodo link:https://zenodo.org/badge/latestdoi/278143533 |
| Stratified LD score regression algorithm | Finucane et al. 2015 Nat Genet[28] | https://github.com/bulik/ldsc |
| Activity-By-Contact S2G links | Nasser et al. 2021 Nature[39] | https://www.engreitzlab.org/resources |
| BOLT-LMM | Lohetal. 2015 Nat Genet[96] | https://data.broadinstitute.org/alkesgroup/BOLT-LMM |