bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Integrative approaches to improve the informativeness of deep learning models for human complex diseases

Kushal K. Dey[1], Samuel S. Kim[1, 2], Steven Gazal[1], Joseph Nasser[3], Jesse M. Engreitz[3], Alkes L. Price [1,3,4]

1 Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA, USA
2 Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA
3 Broad Institute of MIT and Harvard, Cambridge, MA, USA
4 Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA

Abstract

Deep learning models have achieved great success in predicting genome-wide regulatory effects from DNA sequence, but recent work has reported that SNP annotations derived from these predictions contribute limited unique information for human complex disease. Here, we explore three integrative approaches to improve the disease informativeness of allelic-effect annotations (predicted difference between reference and variant alleles) constructed using two previously trained deep learning models, DeepSEA and Basenji. First, we employ gradient boosting to learn optimal combinations of deep learning annotations, using (off-chromosome) fine-mapped SNPs and matched control SNPs for training. Second, we improve the specificity of these annotations by restricting them to SNPs implicated by (proximal and distal) SNP-to-gene (S2G) linking strategies, e.g. prioritizing SNPs involved in gene regulation. Third, we predict gene expression (and derive allelic-effect annotations) from deep learning annotations at SNPs implicated by S2G linking strategies — generalizing the previously proposed ExPecto approach, which incorporates deep learning annotations based on distance to TSS. We evaluated these approaches using stratified LD score regression, using functional data in blood and focusing on 11 autoimmune diseases and blood-related traits (average N=306K). We determined that the three approaches produced SNP annotations that were uniquely informative for these diseases/traits, despite the fact that linear combinations of the underlying DeepSEA and Basenji blood annotations were not uniquely informative for these diseases/traits. Our results highlight the benefits of integrating SNP annotations produced by deep learning models with other types of data, including data linking SNPs to genes.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Deep learning models have shown considerable promise in predicting regulatory marks from DNA sequence[1–8], motivated by the well-documented role of non-coding variation in complex disease[9–15]. However, we recently showed that existing deep learning models provide limited unique information about complex disease when conditioned on a broad set of coding, conserved, regulatory and LD-related annotations[16]. Thus, further ideas are required in order for deep learning models to achieve their full potential in contributing to our understanding of complex disease.

Here, we explore three approaches for integrating different types of functional data to improve the disease informativeness of allelic-effect SNP annotations (predicted difference between reference and variant alleles) constructed using two previously trained deep learning models, DeepSEA[4] and Basenji[5]. First, we employ gradient boosting[17] to learn optimal combinations of deep learning annotations, integrating deep learning annotations with (off-chromosome) fine-mapped SNPs from previous studies[18–20]. Second, we improve the specificity of deep learning annotations by restricting them to SNPs linked to genes; we consider a broad set of proximal and distal SNP-to-gene (S2G) linking strategies, e.g. prioritizing SNPs involved in gene regulation[21–30]. Third, we predict gene expression (and derive allelic-effect annotations) from deep learning annotations at SNPs implicated by S2G linking strategies, generalizing the previously proposed ExPecto approach[4], which incorporates deep learning annotations based on distance to TSS. We consider either SNPs linked to all genes, or SNPs linked to genes in biologically important gene sets[22,31]. We assessed the informativeness of the resulting annotations for disease heritability by applying stratified LD score regression (S-LDSC)[13] to 11 autoimmune diseases and blood-related traits (average N=306K), conditional on a broad set of coding, conserved, regulatory and LD-related annotations from the baseline-LD model[32,33]. We focused our analyses on functional data in blood, and on blood-related diseases/traits.

Results

Overview of Methods

We define an *annotation* as an assignment of a numeric value to each SNP with minor allele count $\geq$5 in a 1000 Genomes Project European reference panel[34], as in our previous work[13]; we primarily focus on annotations with values between 0 and 1. Our annotations are derived from allelic-effect deep learning annotations (predicted difference between reference and variant alleles of sequence-based deep learning predictions of functional annotations) from two recently developed deep learning models, DeepSEA[4] and Basenji[5]. DeepSEA employs a multi-class classification model to predict chromatin features by analyzing sequence data in a 1kb of human reference sequence around a SNP. Basenji employs a Poisson likelihood model to analyze 130kb of human reference sequence around each SNP using dilated convolutional layers. Our previous work[16] focused on unsigned (absolute) allelic-effect annotations for DNase and three histone marks, H3K27ac, H3K4me1 and H3K4me3 (associated with active enhancers and promoters). Here, we integrate signed allelic-effect annotations with other types of data - fine-mapped SNPs, SNPs linked to genes, and gene expression - to generate more disease-informative unsigned annotations. We have publicly released all new annotations analyzed in this study, along with open-source software for constructing the new annotations (see URLs).

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

First, we employ gradient boosting[17] to integrate deep learning annotations with (off-chromosome) fine-mapped SNPs for blood-related traits from previous studies[18–20] to generate boosted annotations representing an optimal combination of deep-learning annotations across features. Second, we improve the specificity of these annotations by restricting them to SNPs linked to genes using 10 (proximal and distal) SNP-to-gene (S2G) strategies[21–30] (Table 1). Third, we predict gene expression (and derive allelic-effect annotations) from deep learning annotations at SNPs implicated by S2G linking strategies, generalizing the previously proposed ExPecto approach[4], which incorporates deep learning annotations based on distance to TSS.

We assessed the informativeness of the resulting annotations for disease heritability by applying stratified LD score regression (S-LDSC)[13] to 11 independent blood-related diseases and traits (5 autoimmune diseases and 6 blood cell traits; average N=306K, Table S1) and meta-analyzing S-LDSC results across traits; we restricted our analyses to blood-related traits due to our focus on functional data in blood. We conservatively conditioned all analyses on a "baseline-LD-deep model" defined by 86 coding, conserved, regulatory and LD-related annotations from the baseline-LD model (v2.1)[32,33] and 14 additional jointly significant annotations from ref.[16]: 1 non-tissue-specific allelic-effect Basenji annotation, 3 Roadmap annotations, 5 ChromHMM annotations, and 5 other annotations (100 annotations total) (Table S2 and Table S3).

We used two metrics to evaluate informativeness for disease heritability: enrichment and standardized effect size ($\tau^\star$). Enrichment is defined as the proportion of heritability explained by SNPs in an annotation divided by the proportion of SNPs in the annotation[13], and generalizes to annotations with values between 0 and 1[25]. Standardized effect size ($\tau^\star$) is defined as the proportionate change in per-SNP heritability associated with a 1 standard deviation increase in the value of the annotation, conditional on other annotations included in the model[32]; unlike enrichment, $\tau^\star$ quantifies effects that are unique to the focal annotation. In our "marginal" analyses, we estimated $\tau^\star$ for each focal annotation conditional on the baseline-LD-deep annotations. In our "joint" analyses, we merged baseline-LD-deep 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.[32].

DeepBoost deep learning annotations restricted to SNPs implicated by functionally informed S2G linking strategies are uniquely informative for autoimmune disease heritability

We developed a gradient boosting approach, DeepBoost, to learn optimal combinations of deep learning annotations (sequence-based deep learning predictions of functional annotations), using off-chromosome fine-mapped SNPs for blood-related diseases/traits[18–20] and matched control SNPs for training (Figure S1 and Methods). The input deep learning annotations consisted of either 2,002 DeepSEA allelic-effect annotations[4] or 4,229 Basenji allelic-effect annotations[5] (see Methods), each representing a particular chromatin mark in a particular tissue/cell type; these annotations spanned 127 tissues and cell types from Roadmap[35], but the 27 blood cell types provided the bulk of the information (see below). The fine-mapped SNPs consisted of 8,741 fine-mapped autoimmune disease SNPs[18] with causal probability > 0.0275 (in secondary analyses, we also considered other sets of fine-mapped SNPs[19,20]). DeepBoost uses decision trees to distinguish fine-mapped SNPs from matched control SNPs (with similar MAF and

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

LD structure and local GC content) using an optimal combination of deep learning annotations; the DeepBoost model is trained using the XGBoost gradient boosting software[17] (see URLs). DeepBoost attained an AUROC of up to 0.67 in distinguishing fine-mapped SNPs from control SNPs, an encouraging result given the fundamental difficulty of this task (Table S4). The boosted allelic-effect annotations derived from DeepBoost (DeepSEAΔ-boosted and BasenjiΔ-boosted; we use Δ to denote allelic-effect annotations) were only moderately correlated with allelic-effect annotations derived from a simple maximum of published deep learning annotations across 4 chromatin marks (DNase, H3K27ac, H3K4me1 and H3K4me3) and 27 blood cell types, as in ref.[16] (DeepSEAΔ-published and BasenjiΔ-published) (average $r = 0.35$) (Figure S2). We broadly investigated which features of the DeepSEA and Basenji models contributed the most to the DeepSEAΔ-boosted and BasenjiΔ-boosted annotations by applying Shapley Additive Explanation (SHAP)[36], a widely used tool for interpreting machine-learning models. For DeepSEAΔ-boosted, the top features included TF features in K562 and GM12878, two immune-related cell lines; for BasenjiΔ-boosted, top features included activating histone marks (H3K27ac and H3K4me3) in blood cell types (Figure S3, Figure S4). In secondary analyses, we determined that (i) using logistic classification instead of XGBoost (using the same features as the DeepBoost model) attained only slightly lower AUROC (Table S5); and (ii) using only features from 27 blood cell types (524 DeepSEA features or 479 Basenji features) attained only slightly lower AUROC (Table S6).

In our marginal analysis of disease heritability (across 11 autoimmune diseases and blood-related traits) using S-LDSC conditional on the baseline-LD-deep model, all 4 annotations (DeepSEAΔ-published, BasenjiΔ-published, DeepSEAΔ-boosted, BasenjiΔ-boosted) were significantly enriched for heritability, with larger enrichments for the boosted annotations (Figure 1A, left panel and Table S7); values of standardized enrichment (defined as enrichment scaled by the standard deviation of the annotation) are reported in Figure S5 and Table S8. However, none of these annotations attained Bonferroni-significant $\tau^\star$ values (correcting for 90 annotations tested; see Methods) (although the BasenjiΔ-boosted annotation was FDR-significant) (Figure 1B, left panel and Table S7). We constructed analogs of the DeepSEAΔ-boosted and BasenjiΔ-boosted annotations using three other sets of fine-mapped SNPs: 4,312 fine-mapped inflammatory bowel disease SNPs[19], 1,429 functionally fine-mapped SNPs for 14 blood-related UK Biobank traits[20,37], or the union of all 14,482 fine-mapped SNPs. The resulting annotations produced less disease signal than those constructed using the 8,741 fine-mapped autoimmune disease SNPs[18] (Table S7).

We sought to improve the specificity of these annotations by restricting them to SNPs implicated by SNP-to-gene (S2G) linking strategies, e.g. prioritizing SNPs that may play a role in gene regulation; we define an S2G strategy as an assignment of 0, 1 or more linked genes to each SNP. We considered 10 S2G strategies capturing both proximal and distal gene regulation in blood, as in our previous work[22] (see Methods and Table 1), and constructed 10 corresponding binary S2G annotations defined by SNPs linked to the set of all genes; the S2G annotations were only mildly positively correlated with each other (average $r = 0.09$; Figure S6). We defined *restricted* allelic-effect annotations as a simple product of allelic-effect annotations and S2G annotations. Due to correlations between allelic-effect annotations and S2G annotations (average $r = 0.16$; Figure S7), the size of a restricted allelic-effect annotation was generally larger than the product of the respective sizes of the underlying allelic-effect and S2G annotations. We evaluated 40 restricted allelic-effect annotations (4 allelic-effect annotations (2 published + 2 boosted) x 10 S2G annotations). We analyzed the restricted allelic-effect annotations conditional on a "baseline-LD-deep-S2G model" defined by 100 baseline-LD-deep annotations and 7 new S2G annotations from Table 1 that were not already included in the baseline-LD model

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(107 annotations total) (Table S2 and Table S9), to ensure that heritability enrichments that are entirely due to S2G annotations would not produce conditional signals.

In our marginal analysis of disease heritability using S-LDSC conditional on the baseline-LD-deep-S2G model, all 40 annotations were significantly enriched for heritability, with larger enrichments for smaller annotations (Figure 1A, right panel and Table S10); values of standardized enrichment were more similar across annotations (Table S11). Although published and boosted allelic-effect annotations were of similar size, the enrichments for boosted-restricted annotations across S2G strategies were on average 1.3x (DeepSEA) and 1.5x (Basenji) higher than the enrichments for published-restricted annotations. 3 of the boosted-restricted annotations (DeepSEAΔ-boosted $\times$ eQTL, BasenjiΔ-boosted $\times$ ABC and BasenjiΔ-boosted $\times$ TSS) attained Bonferroni-significant $\tau^\star$ values (Figure 1B, right panel and Table S10). (In comparison, when we conditioned only on the baseline-LD-deep model, 15 of the 40 annotations attained Bonferroni-significant $\tau^\star$ values (Table S12).

We jointly analyzed the 3 marginally significant annotations from Figure 1B, right panel by performing forward stepwise elimination to iteratively remove annotations that had conditionally non-significant τ^* values after Bonferroni correction. All 3 annotations were jointly significant in the resulting joint model, with joint effect sizes very similar to the conditional effect sizes from Figure 1B, right panel (Figure S8 and Table S13). All three annotations had joint $\tau^\star > 0.5$; annotations with $\tau^* > 0.5$ are unusual, and considered to be important[38].

We investigated whether the boosted-restricted annotations would detect gene set-specific signals by further restricting them to SNPs linked to two biologically important gene sets: genes intolerant to loss-of-function (LoF) mutations[31] (pLI) and genes with high PPI network connectivity to Enhancer-driven genes in blood[22] (PPI-enhancer). We defined *gene set-specific* boosted-restricted annotations by replacing the S2G annotations (containing SNPs linked to all genes) with annotations containing SNPs linked to genes in the input gene set (Methods); we primarily focused on boosted-restricted annotations (instead of published-restricted annotations) because these were the restricted annotations that produced significant conditional signals in Figure 1B, right panel. We evaluated 40 gene set-specific boosted-restricted annotations (2 gene sets (pLI, PPI-enhancer) x 2 boosted allelic-effect annotations (BasenjiΔ-boosted, DeepSEAΔ-boosted) x 10 S2G strategies). We analyzed the gene set-specific boosted-restricted annotations conditional on a "baseline-LD-deep-S2G-geneset" model defined by 107 baseline-LD-deep-S2G annotations and 8 jointly significant gene set-specific S2G annotations (Table S14 and Table S2), to ensure that heritability enrichments that are entirely due to the gene set-specific S2G annotations would not produce conditional signals.

In our marginal analysis of disease heritability using S-LDSC conditional on the baseline-LD-deep-S2G-geneset model, all 40 annotations were significantly enriched for heritability, with larger enrichment for smaller annotations (Figure 2A and Table S15); values of standardized enrichment were more similar across annotations (Figure S9 and Table S16). 7 of the 40 annotations (3 BasenjiΔ-boosted (PPI-enhancer) and 4 DeepSEAΔ-boosted (PPI-enhancer)) attained conditionally Bonferroni-significant $\tau^\star$ values (Figure 2B and Table S15). We jointly analyzed these 7 annotations by performing forward stepwise elimination. The resulting joint model contained 2 jointly significant annotations, BasenjiΔ-boosted (PPI-enhancer) $\times$ ABC and BasenjiΔ-boosted (PPI-enhancer) $\times$ Roadmap (Figure 2C and Table S17); both annotations had joint $tau^\star > 0.5$. Both annotations remained jointly significant, with very similar τ^* values, when further conditioned on the 3 jointly significant boosted-restricted annotations from Figure 1B,

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

right panel (including the underlying BasenjiΔ-boosted $\times$ ABC annotation) and the underlying BasenjiΔ-boosted $\times$ Roadmap annotation (Table S17).

We performed 2 secondary analyses. First, we repeated the analysis of restricted annotations using local GC-content (proportion of G and C nucleotides in a 1000bp window around each SNP) in addition to the S2G strategies, conditioning on the baseline-LD-deep-S2G model and the unweighted local GC-content annotation. The τ^* values for all 3 jointly significant restricted annotations from Figure 1B, right panel were nearly unchanged and remained Bonferroni-significant (Table S18); this implies that the unique disease signal in our restricted annotations cannot be explained by local GC-content. Second, we repeated the analysis of gene set-specific restricted annotations using published-restricted annotations instead of boosted-restricted annotations. Marginal results were only slightly worse than Figure 2B (6 annotations with Bonferroni-significant τ^* values; Table S19), but none of the gene set-specific published-restricted annotations annotations were significant conditional on the 2 jointly significant gene set-specific boosted-restricted from Figure 2C (Table S20).

We conclude that boosted deep learning allelic-effect annotations restricted to SNPs implicated by functionally informed S2G linking strategies are uniquely informative for autoimmune diseases and blood-related traits.

Sequence-based deep learning predictions of gene expression informed by S2G linking strategies are uniquely informative for autoimmune disease heritability

We developed a new approach, Imperio, to predict gene expression from DNA sequence by using S2G strategies to prioritize deep learning annotations (sequence-based deep learning predictions of functional annotations) as features (Figure S10 and Methods). Imperio generalizes the ExPecto approach[4], which prioritizes deep learning annotations as features based on distance to TSS. Specifically, Imperio uses regularized linear regression to fit optimal combinations of features predicting gene expression across 22,020 genes based on 2,002 DeepSEA or 4,229 Basenji deep learning annotations restricted to relatively common SNPs (MAF $> 1\%$) linked to the target gene by 5 S2G strategies that are suitably large in size and generalizable to tissues beyond blood (5kb, 100kb, TSS, ABC and Roadmap Enhancer; Table 1) (2,002 $\times$ 5 or 4,229 $\times$ 5 features); the feature weights are independent of the target gene but dependent on the deep learning annotation and the S2G strategy (see Methods). In contrast, ExPecto fits optimal combinations of features based on 2,002 DeepSEA annotations restricted to 10 different functions of distance to TSS (using exponential decay), for a total of 2,002 $\times$ 10 features. We focused on predicting gene expression in blood, due to the larger amount of data currently available for ABC and Roadmap Enhancer in blood cell types (however, our approach is generalizable to other tissues). We evaluated the accuracy of Imperio in predicting gene expression *across genes* on chromosome 8, which was withheld from Imperio training data (analogous to ref.[4]). We determined that Imperio attained similar predictive accuracy as ExPecto (Spearman correlation $\rho = 0.76$ (Basenji) and $\rho = 0.72$ (DeepSEA) with log RPKM expression, vs. $\rho = 0.79$ for ExPecto; Figure S11). The expression predictions were highly correlated between the Imperio and ExPecto models (average $\rho = 0.82$) (Figure S12), but the resulting allelic-effect annotations were less correlated, such that Imperio may contribute unique information (see below). The top significant features driving the Imperio model fit included Transcription Factor (TF) features for DeepSEA and CAGE features for Basenji (Table S21). When we compared

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the 5 Imperio models utilizing a single S2G strategy, TSS outperformed the other S2G strategies, but the resulting model fit (Spearman correlation $\rho = 0.69$ (Basenji) and $\rho = 0.71$ (DeepSEA) with log RPKM expression) was substantially worse than the model fit of the Imperio model utilizing all 5 S2G strategies (Table S22).

We used the Imperio allelic effects (signed predicted difference in expression between reference and variant alleles) to predict GTEx blood gene expression *across individuals* for each gene (see Methods). For each gene, we compared the Imperio prediction r^2 to the total cis-SNP heritability of that gene, which represents an upper bound on the prediction r^2 that can be attained using DNA sequence. Averaging across all 22,020 genes, Imperio predictions captured 81% of the total cis-SNP heritability on average (vs. 75% for ExPecto; this analysis was not considered in ref.[4]) (Table S23). The Imperio prediction r^2 closely tracked cis-SNP heritability ($\rho = 0.83$ for DeepSEA, $\rho = 0.84$ for Basenji across genes, vs. $\rho = 0.81$ for ExPecto) (Figure S13).

We used the gene expression predictions from Imperio (DeepSEA and Basenji models) and ExPecto[4] (DeepSEA model) to construct expression allelic-effect annotations (absolute value of the predicted difference in expression between reference and variant alleles) by summing allelic effects across genes linked by S2G strategies to the annotated SNP (see Methods). The Imperio training data excluded chromosome 8 (analogous to ref.[4]; see above), but did not exclude the target chromosomes on which allelic-effect annotations were constructed. However, this does not constitute overfitting, because the Imperio model was trained using reference sequence only. The Imperio-DeepSEA and Imperio-Basenji annotations were moderately correlated with each other ($r = 0.54$) and with ExPecto-DeepSEA (average $r = 0.48$) (Figure S14), such that each may contribute unique information. Furthermore, Imperio-DeepSEA and Imperio-Basenji annotations showed only mild correlation (average r=0.11) with boosted-restricted allelic effect annotations from previous section (Table S24). We analyzed the Imperio-DeepSEA, Imperio-Basenji and ExPecto-DeepSEA allelic-effect annotations conditional on the baseline-LD-deep-S2G-geneset model (see above; Table S2 and Table S14), for consistency with analyses of gene set-specific allelic-effect annotations (see below).

In our marginal analysis of disease heritability using S-LDSC conditional on the baseline-LD-deep-S2G-geneset model, all 3 annotations were significantly enriched for disease heritability, with larger enrichments for smaller annotations annotations (Figure 3A and Table S25); values of standardized enrichment were more similar across annotations (Table S26). One annotation, Imperio-Basenji, attained a Bonferroni-significant $\tau^\star$ value (Figure 3B and Table S25); the $\tau^\star$ value was very close to 0.5. This implies that Imperio provides unique information about autoimmune diseases and blood-related traits.

We investigated whether the Imperio approach would detect gene-set specific signals by restricting Imperio to two biologically important gene sets: pLI[31] and PPI-enhancer[22] (see above). We defined *gene set-specific* allelic-effect annotations by restricting both the fitting of feature weights and the gene expression predictions to genes in the input gene set. Pairwise correlations between the 4 gene set-specific allelic-effect annotations ([Imperio-DeepSEA or Imperio-Basenji] x [pLI or PPI-enhancer]) (and the 3 non-gene set-specific allelic-effect annotations) are reported in Figure S14. We analyzed the gene set-specific allelic-effect annotations conditional on the baseline-LD-deep-S2G-geneset model (see above; Table S2 and Table S14).

In our marginal analysis of disease heritability using S-LDSC conditional on the baseline-LD-deep-S2G-geneset model, all 4 annotations were significantly enriched for dis-

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ease heritability, with larger enrichments for smaller annotations annotations (Figure 3C and Table S27); values of standardized enrichment were more similar across annotations (Table S28). Two annotations, Imperio-DeepSEA (PPI-enhancer) and Imperio-Basenji (PPI-enhancer), attained Bonferroni-significant $\tau^\star$ values (Figure 3D and Table S27). In a joint analysis of both annotations, only Imperio-DeepSEA (PPI-enhancer) remained significant (Figure 3D and Table S29); the $\tau^\star$ value was larger than 0.5. Imperio-DeepSEA (PPI-enhancer) remained significant (with $\tau^\star > 0.5$) when further conditioned on the Imperio-Basenji annotation from Figure 3B (Table S30).

We performed 5 secondary analyses. First, we fit an Imperio+ExPecto model using both Imperio (DeepSEA or Basenji) and ExPecto features. The Imperio+ExPecto allelic-effect annotations did not produce a significant disease signal conditional on the baseline-LD-deep-S2G-geneset model plus the Imperio-Basenji annotation from Figure 3B (Table S31). Second, we investigated a partially restricted gene set-specific Imperio approach by restricting either (a) the fitting of feature weights *or* (b) the gene expression predictions (but not both) to genes in the input gene set. None of the partially restricted gene set-specific annotations produced a significant disease signal conditional on the baseline-LD-deep-S2G-geneset model plus the two significant annotations from Figure 3B,D (Table S32 and Table S33). Third, we assessed whether the disease informativeness of Imperio could be explained by annotations defined by the number of genes linked to each SNP by each S2G strategy (see Methods). However, none of these annotations produced a significant disease signal conditional on the baseline-LD-deep-S2G-geneset model, either for all genes (Table S34) or when restricted to PPI-enhancer genes (Table S35). Fourth, we modified Imperio by constructing allelic-effect annotations using the *maximum* across genes proximal to the annotated SNPs, instead of the sum (see Methods). None of the modified annotations produced a significant disease signal conditional on the baseline-LD-deep-S2G-geneset model plus the two significant annotations from Figure 3B,D (Table S36). Fifth, we compared the Imperio annotations to MaxCPP-blood (Maximum across genes of fine-mapped eQTL Causal Posterior Probability) annotation[25] constructed using GTEx whole blood gene expression data[39]. The MaxCPP-blood annotation was only weakly correlated with Imperio annotations (average $r = 0.09$) and did not produce a significant disease signal conditioned on the baseline-LD-deep-S2G-geneset model (Table S37), consistent with the fact that a related MaxCPP annotation based on a meta-analysis across tissues[25] is already included in the baseline-LD model.

We conclude that allelic-effect annotations based on predictions of gene expression from DNA sequence using S2G linking strategies to prioritize deep learning annotations as features are uniquely informative for autoimmune diseases and blood-related traits.

Combined joint model

We constructed a combined joint model containing annotations from the above analyses that were jointly significant, contributing unique information conditional on all other annotations. In detail, we merged the baseline-LD-deep-S2G-geneset model with 3 DeepBoost boosted-restricted allelic-effect annotations from Figure 1, 2 gene-set specific DeepBoost annotations from Figure 2, 1 Imperio gene expression prediction allelic-effect annotation from Figure 3B, and 1 gene-set specific Imperio annotation from Figure 3D, and performed forward stepwise elimination to iteratively remove annotations that had conditionally non-significant $\tau^\star$ values after Bonferroni correction. The resulting combined joint model contained 3 new annotations, including 1 DeepBoost annotation (BasenjiΔ-boosted $\times$ TSS) and the 2 Imperio annotations (Imperio-Basenji and Imperio-

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

DeepSEA (PPI-enhancer)) (Figure 4 and Table S38). 2 of these annotations attained $\tau^* > 0.5$: BasenjiΔ-boosted $\times$ TSS (1.1 ± 0.29) and Imperio-DeepSEA (PPI-enhancer) (0.67 ± 0.15); as noted above, annotations with $\tau^* > 0.5$ are unusual, and considered to be important[38]. The combined τ^*[22,40] of the 3 annotations was high (1.7 ± 0.3).

We evaluated the combined joint model of Figure 4 (and other models) by computing $\log l_{SS}$[41] (an approximate likelihood metric) relative to a model with no functional annotations ($\Delta\log l_{SS}$), averaged across a subset of 6 blood-related traits (1 autoimmune disease and 5 blood cell traits) from the UK Biobank[37] (Table S1). The combined joint model attained a +20.3% larger $\Delta\log l_{SS}$ than the baseline-LD model (Table S39); +2.5% of the improvement derived from the 3 new annotations from Figure 4. The combined joint model also attained a +14.2% larger $\Delta\log l_{SS}$ than the baseline-LD model (+2.2% of the improvement derived from the 3 new annotations from Figure 4) in a separate analysis of 24 non-blood-related traits from the UK Biobank (see Table S40 for list of traits) that had lower absolute $\log l_{SS}$ values (Table S39), implying that the value of the annotations introduced in this paper is not restricted to autoimmune diseases and blood-related traits.

We conclude that two types of allelic-effect annotations informed by S2G strategies—DeepBoost boosted-restricted annotations and Imperio gene expression prediction annotations—are jointly informative for autoimmune diseases and blood-related traits.

Discussion

We have evaluated the contribution to autoimmune disease of SNP annotations constructed by integrating deep learning models with different types of functional data, including fine-mapped SNPs, SNP-to-gene linking strategies, gene expression data, and biologically important gene sets, using our DeepBoost and Imperio frameworks. We determined that boosted deep allelic-effect annotations restricted to SNPs implicated by functionally informed S2G linking strategies are uniquely informative for disease. We also determined that allelic-effect annotations based on prediction of gene expression from DNA sequence that were informed by S2G linking strategies are uniquely informative for disease, outperforming allelic-effect annotations from ExPecto[4]. We further determined that both DeepBoost and Imperio allelic-effect annotationswere jointly informative for disease, resulting in an improved heritability model.

Our work has several downstream implications. First, the DeepBoost and Imperio frameworks can be applied to deep learning models beyond DeepSEA and Basenji, and we anticipate that future deep learning models will benefit from these frameworks. . Second, the accuracy of the Imperio framework in capturing cis-SNP heritability in blood suggests that it may be valuable to integrate Imperio gene expression predictions in other settings, such as transcriptome-wide association studies (TWAS)[?,42,43] or mediated expression score regression (MESC)[44]. Third, our findings have immediate potential for improving functionally informed fine-mapping[20,45–47] (including experimental follow-up[48]), polygenic localization[20], and polygenic risk prediction[49,50].

Our work has several limitations, representing important directions for future research. First, we focused our analyses on functional data in blood, and on blood-related diseases/traits; this choice was motivated by (i) the better representation of some S2G strategies, such as ABC and Roadmap Enhancer, in blood cell types than in other tissues, and (ii) the particularly large functional enrichments observed in autoimmune

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

diseases and blood-related traits[13,22,25,32]. However, it will be of interest to apply the DeepBoost and Imperio frameworks to other tissues and corresponding diseases/traits, once richer functional data becomes available. Second, we investigated the 10 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). Third, our S-LDSC analyses are inherently focused on common variants, but deep learning models have also shown promise in prioritizing rare pathogenic variants[4,8,51]. The value of deep learning models for prioritizing rare pathogenic variants has been questioned in a recent analysis focusing on Human Gene Mutation Database (HGMD) variants[52], meriting further investigation. Fourth, we focused here on deep learning models trained using human data, but models trained using data from other species may also be informative for human disease[24,53]. Fifth, the forward stepwise elimination procedure that we use to identify jointly significant annotations[32] is a heuristic procedure whose choice of prioritized annotations may be close to arbitrary in the case of highly correlated annotations. Nonetheless, our framework does impose rigorous criteria for conditional informativeness. Despite all these limitations, our findings improve the informativeness of deep learning models for autoimmune diseases and blood-related traits, and enhance our understanding of the sequence-mediated regulatory processes impacting these diseases/traits.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Methods

Genomic annotations and the baseline-LD model

We define a functional annotation as an assignment of a numeric value to each SNP with minor allele count ≥ 5 in a predefined reference panel (e.g., 1000 Genomes Project[34]; see URLs). Annotations can be either binary or continuous-valued (Methods). Our focus is on continuous-valued annotations (with values between 0 and 1) that are obtained by integrating deep learning models with functional data, including fine-mapped SNPs, SNP-to-gene linking strategies, gene expression data, and biologically important gene sets. 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.

DeepSEA and Basenji functional annotations

Tissue-specific deep learning annotations were derived using two pre-trained Convolutional Neural Net (CNN) models: Basenji[5] and DeepSEA[2,4] (architecture from ref.[4]) (see URLs). Basenji is a Poisson likelihood model trained on original count data from $4,229$ cell-type specific histone mark, chromatin accessibility and FANTOM5 CAGE[54,55] annotations. DeepSEA is a classification based model trained on binary peak call data from $2,002$ cell-type specific TFBS, histone mark and chromatin accessibility annotations from the ENCODE[56] and Roadmap Epigenomics[35] projects. Additionally, Basenji uses dilated convolutional layers that allow scanning much larger contiguous sequence around a variant ($\approx$ 130kb) compared to DeepSEA (1kb). For each SNP with minor allele count ≥ 5 in 1000 Genomes, we applied the pre-trained DeepSEA and Basenji models to the surrounding DNA sequence to compute both the prediction (at reference allele) and the predicted difference in probability between the reference and the alternate alleles. We call these the *variant-level annotations* and *allelic-effect annotations* respectively; this naming convention has been used previously[16]. The allelic-effect annotations are more interesting from a biological perspective as they are specific to a sequence-based predictive model like these deep learning models. The Basenji allelic-effect annotations were quantile-matched to corresponding DeepSEA allelic-effect annotations to ensure a fair comparison of the two approaches.

Our previous work focused on DeepSEA and Basenji unsigned allelic-effect annotations, aggregated by maximum across tissue-specific features, for DNase, H3K27ac, H3K4me1 and H3K4me3. Here we define 2 allelic-effect annotations, DeepSEAΔ-published and BasenjiΔ-published, as the maximum over these 4 chromatin marks (DNase, H3K27ac, H3K4me1 and H3K4me3).

Boosted deep learning annotations using DeepBoost

DeepSEAΔ-published and BasenjiΔ-published represent a simple maximum of allelic-effect annotations across tissues and chromatin features. Here we introduce a gradient boosting approach to combine allelic-effect annotations across tissues and chromatin

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

features. In detail, we train a classification model using decision trees, where each node in a tree splits SNPs into 2 classes (fine-mapped and control) using deep learning allelic-effect annotations from DeepSEA and Basenji models. The features in this classification model comprise of either allelic-effect annotations for 2,002 DeepSEA features or allelic-effect annotations for 4,229 Basenji features. We choose the control SNPs from non-finemapped SNPs matched for MAF, LD, local GC-content and the number of repeats distribution. MAF is based on the same reference panel (European samples from 100 Genomes Phase 3[34]), and LD is estimated by applying S-LDSC on all SNPs annotation ('base'). The number of control SNPs were chosen equal to the number of fine-mapped SNPs. We used fine-mapped SNPs data related to blood traits from three sources[18–20].

We used the Extreme gradient boosting method implemented in the XGBoost software[17,57] with following model parameters: the number of estimators (200, 250, 300), depth of the tree (25, 30, 35), learning rate (0.05), gamma (minimum loss reduction required before additional partitioning on a leaf node; 10), minimum child weight (6, 8 ,10), and subsample (0.6, 0.8, 1); we optimized parameters by tuning hyper-parameters (a randomized search) with five-fold cross-validation. Two important parameters to avoid over-fitting are gamma and learning rate; we chose these values consistent with previous studies[58].

The gradient boosting predictor is based on T additive estimators (T=200, 250, 300) and it minimizes the loss objective function $\mathcal{L}^t$ at iteration t.

$$\mathcal{L}^{(t)} = \sum_{i=1}^{N} l(y_i, \hat{y}_i^{(t)}) + \gamma(f_t) \qquad \hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i) \tag{1}$$

f_t is an independent tree structure and $\gamma(f_t)$ is the complexity parameter. The final prediction from the gradient boosting model therefore is

$$\hat{y}_i = \sum_{t=1}^{T} f_t(x_i) \tag{2}$$

In order to avoid a winner's curse problem and overfitting problem, we break down the training and testing into two separate groups. Our annotations are generated by out-of-sample predictions by performing the training on odd (resp. even) chromosomes and then predicting on even (resp. odd) chromosomes.The output of the classifier is probabilistic in nature because of the logistic loss and so the annotations are automatically probabilistic. Eventually, we generate 2 annotations, DeepSEAΔ-boosted and BasenjiΔ-boosted for each of 4 sets of fine-mapped SNPs, comprising of 8,741 fine-mapped autoimmune disease SNPs[18], 4,312 fine-mapped inflammatory bowel disease SNPs[19], 1,429 functionally fine-mapped SNPs for 14 blood-related UK Biobank traits[20,37], or the union of these 14,482 fine-mapped SNPs.

Boosted-restricted deep learning annotations using S2G strategies

We define a SNP-to-gene (S2G) linking strategy as an assignment of 0, 1 or more linked genes to each SNP with minor allele count ≥ 5 in a 1000 Genomes Project

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

European reference panel[34]. We intersect the 4 allelic-effect annotations from the previous subsections (DeepSEAΔ-published, BasenjiΔ-published, DeepSEAΔ-boosted, BasenjiΔ-boosted) with 10 S2G strategies used in ref.[22]to generate 40 restricted allelic-effect annotations

We explored 10 SNP-to-gene linking strategies in blood (Table 1). We use either gene proximity based linking strategies (SNPs in gene body $\pm$ 5kb, gene body $\pm$ 100kb, predicted TSS (by Segway[28,29]) around a protein-coding gene, gene coding SNPs, gene promoter SNPs as defined by UCSC[59,60]) or SNPs that lie in regions that are distally connected to a gene in blood cell-types. The distal connection strategies studied here include - regions estimated to be distally connected to the gene by Activity-by-Contact (ABC) score[23]; regions predicted to be Enhancer-Gene links based on Roadmap Epigenomics data (Roadmap)[26,30,35]; regions in ATAC-seq peaks that are highly correlated to expression of a gene in mouse immune cell-types (ATAC)[24]; regions distally connected through promoter-capture HiC links (PC-HiC)[27].

The boosted-restricted allelic-effect annotations were further restricted to SNPs linked to genes in two biologically important gene sets - pLI[31] and PPI-enhancer[22].

- **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 as defined in ref.[22]. To get the closeness centrality metric, we first perform a Random Walk with Restart (RWR) algorithm[61] on the STRING protein-protein interaction (PPI network[62,63](see URLs) with seed nodes defined by genes in top 10% of the 4 enhancer-regulated gene scores defined in ref.[22] with jointly significant disease informativeness (ABC-G[21,23], ATAC-distal[24], EDS-binary[64] and SEG-GTEx[65]). The closeness centrality score was defined as the average network connectivity of the protein products from each gene based on the RWR method.
- **pLI** : A probabilistic gene score with each gene graded by the probability of intolerance to loss of function mutations[31].

We generate an additional 40 annotations by combining the 2 gene scores (pLI, PPI-enhancer) with 20 restricted boosted allelic-effect annotations for DeepSEA and Basenji models and 10 S2G strategies.

Imperio deep learning annotations using gene expression predictions informed by S2G strategies

We propose a new framework, Imperio, that for predicting gene expression from DNA sequence by using S2G strategies to prioritize deep learning annotations (sequence-based deep learning predictions of functional annotations) as features. This approach is analogous to the recent ExPecto framework[4], but focuses on sequences around common variants linked to genes—either proximally or distally via enhancers, as in the Roadmap and ABC distal S2G strategies. We selected these two distal S2G strategies because they outperformed other distal strategies in blood in our previous work[22]. We integrate both DeepSEA and Basenji models with S2G strategies to predict gene expression. We consider a reduced set of 5 classes of S2G strategies: 5kb, 100kb, TSS, ABC and Roadmap Enhancer. We fit a elastic net regularized linear regression model to log RPKM expression data for gene g, Y_g.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$Y_g = \sum_f \sum_d \beta_{fd} \sum_{s \in \mathcal{C}_g^d} p_{sf} + \epsilon_g; \qquad \epsilon_g \sim N(0, \sigma^2) \tag{3}$$

where f represents the chromatin mark features for the deep learning model (2,002 for the DeepSEA model and 4,229 for the Basenji model), s represents SNPs that are at least 1Kb apart ensuring relatively weaker correlation in their variant-effect or allelic-effect annotations, d represents a SNP-to-gene linking strategy, and $\mathcal{C}_g^d$ represents the set of all SNPs linked to gene g by the S2G strategy d. The 5 types of $\mathcal{C}_g^d$ are:

- $\mathcal{C}_g^{5kb}$: SNPs in a window of 5kb around gene g
- $\mathcal{C}_g^{100kb}$: SNPs in a window of 100kb around gene g
- $\mathcal{C}_g^{TSS \pm 5kb}$: SNPs in a window of of $\pm 5kb$ around the TSS of gene g.
- $\mathcal{C}_g^{ABC}$: SNPs in regions linked to gene g by aggregation of Hi-C and enhancer marks data in 56 blood cell-types with a Acitivity-by-Contact (ABC) score of > 0.03.
- $\mathcal{C}_g^{Roadmap}$: SNPs in Roadmap Enhancers linked to gene g in 27 blood cell-types.

β_{fd} represents the model coefficient capturing the effect of each chromatin feature f and each S2G strategy d on the gene expression. p_{sf} represents the variant-level prediction for chromatin feature f around SNP i. ϵ_g represents white noise in the regression model. The model in Equation 3 is fitted by using Extreme gradient boosting (XGBoost) method. Following the training procedure in ExPecto, all genes except the ones in chromosome chr8 were used for training. The predictive performance of this approch is assessed on the holdout chromosome chr8.

We define the signed Imperio effect of each SNP as the sequence mediated effect on expression of a variant s and S2G strategy d.

$$\mathcal{J}_{sd} := \sum_f \hat{\beta}_{fd} \left(p_{sf}^{alt} - p_{sf}^{ref} \right) \tag{4}$$

$\mathcal{J}_{sd}$ is the per-allele estimated change in expression caused by SNP s for any gene it is linked to through S2G strategy d. $\mathcal{J}_{sd}$ is treated as the atom for any Imperio based annotations we investigate.

The total absolute change in expression of gene g caused by SNP s and strategy d is given as follows.

$$|\Delta Y_g^{(s,d)}| := |\mathcal{J}_{sd}| 1_{s \in \mathcal{C}_g^d} \tag{5}$$

The total sequence mediated absolute predicted change by SNP s and S2G strategy d across all genes g is given by

$$\Delta_{sd} := \sum_g |\mathcal{J}_{sd}| 1_{s \in \mathcal{C}_g^d} = |\mathcal{J}_{sd}| N_{sd} \tag{6}$$

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

where $N_s d$ is the number of genes linked to SNP s by S2G strategy d.

$$\Delta_s := \sum_d \sum_g |\mathcal{J}_{sd}| N_{sd} = \sum_d \sum_g \left| \sum_f \hat{\beta}_{fd} \left(p_{sf}^{alt} - p_{sf}^{ref} \right) \right| 1_{s \in \mathcal{C}_g^d} \tag{7}$$

We adjust for the minor allele frequency (MAF) p_s for each SNP s to adjust for per-allele effect sizes, as per ref[66].

$$\Delta_s := p_s(1-p_s) \sum_d \sum_g \left| \sum_f \hat{\beta}_{fd} \left(p_{sf}^{alt} - p_{sf}^{ref} \right) \right| 1_{s \in \mathcal{C}_g^d} \tag{8}$$

These Δ_s scores were normalized to convert them to a probabilistic scale.

For a supplementary analysis, we also consider annotations that do not include the information of the number of genes linked to a SNP (ξ_s).

$$\xi_s := p_s(1-p_s) \sum_d \sum_g \left| \sum_f \hat{\beta}_{fd} \left(p_{sf}^{alt} - p_{sf}^{ref} \right) \right| \tag{9}$$

We analyze 3 annotations, 2 Imperio annotations, Imperio-DeepSEA and Imperio-Basenji, and ExPecto-DeepSEA.

Predicting gene expression *across individuals* using Imperio

We use the Imperio effect of each SNP s in S2G strategy d,$\mathcal{J}_{sd}$ from Equation 4 (for either DeepSEA or the Basenji model) to define a gene specific Imperio score for each individual n and S2G strategy d as follows.

$$\mathcal{I}_{nd}^{(g,1)} = \sum_{s \in \mathcal{C}_g^d} G_{ns} \mathcal{J}_{sd} \tag{10}$$

$$\mathcal{I}_{nd}^{(g,2)} = G_{ns^\star} \mathcal{J}_{s^\star d} \qquad s^\star = \underset{argmax}{s \in \mathcal{C}_g^d} \mathcal{J}_{s^\star d} \tag{11}$$

where G_{ns} represents the number of risk alleles for individual n and the commonly varying SNP s.

Next we perform a regression on the normalized gene expression log RPKM data for individual n and gene g, Y_{ng} with predictors giveny by $\mathcal{I}_{nd}^{(g,1)}$ and $\mathcal{I}_{nd}^{(g,2)}$.

$$Y_{ng} = \sum_d \gamma_{gd}^{(1)} \mathcal{I}_{nd}^{(g,1)} + \sum_d \gamma_{gd}^{(2)} \mathcal{I}_{nd}^{(g,2)} + \sum_k \beta_k B_{nk} + \epsilon_{ng} \qquad \epsilon_{ng} \sim N(0, \eta^2) \tag{12}$$

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

B denotes the covariates that are adjusted for in the model. We consider a total of 68 covariates including 5 principal components across samples, platform, gender and PCR amplification. In cases where there is only one SNP s in $\mathcal{C}_g^d$, $\mathcal{I}_{nd}^{(g,1)} = \mathcal{I}_{nd}^{(g,2)}$ and only one of these predictors is used. This model provides an insight on relative contribution of different S2G strategies in explaining the inter-individual gene expression variation.

We compute R_g^2, the proportion of variance explained by the predictor variables $\mathcal{I}_{nd}^{(g,1)}$ and $\mathcal{I}_{nd}^{(g,2)}$ for all S2G strategies d and for gene g.

Gene set-specific Imperio deep learning annotations

The Imperio model coefficients β_{fd} in the previous section are fitted across all genes. However, different genes may have distinct sequence-mediated regulatory characteristics. Additionally, not all genes in blood are equally important. Therefore, we propose a gene-set specific Imperio model, where we perform the training of the model in Equation 3 over all genes g in a particular gene set $\mathcal{G}$. We consider two gene sets, pLI[31] and PPI-enhancer[22] (see above).

The sequence-mediated expression effect of a variant s corresponding to gene set $\mathcal{G}$ is given by

$$\Delta_s(\mathcal{G}) := p_s(1-p_s) \sum_d \sum_{g \in \mathcal{G}} \left| \sum_f \hat{\beta}_{fd}(\mathcal{G}) \left(p_{sf}^{alt} - p_{sf}^{ref} \right) \right| 1_{s \in \mathcal{C}_g^d} \tag{13}$$

where $\hat{\beta}_{fd}(\mathcal{G})$ are the estimated model coefficients of β_{fd} in Equation 3 fitted for genes in gene set $\mathcal{G}$.

We analyze 4 annotations, combining Imperio models for DeepSEA and Basenji models with the PPI-enhancer and pLI gene sets.

We further define intermediate Imperio annotations by restricting either (a) the fitting of feature weights *or* (b) the gene expression predictions (but not both) to genes in the input gene set.

We define Imperio-sub-1 annotations generated by using all genes for fitting the model and gene sets for computing the expression allelic effects.

$$\Delta_s^{(2)}(\mathcal{I}) := p_s(1-p_s) \sum_d \sum_{g \in \mathcal{G}} \left| \sum_f \hat{\beta}_{fd} \left(p_{sf}^{alt} - p_{sf}^{ref} \right) \right| 1_{s \in \mathcal{C}_g^d} \tag{14}$$

We define Imperio-sub-2 annotations generated by using genes in a geneset for fitting the model and all genes for computing the expression allelic effects

$$\Delta_s^{(2)}(\mathcal{I}) := p_s(1-p_s) \sum_d \sum_{g} \left| \sum_f \hat{\beta}_{fd}(\mathcal{G}) \left(p_{sf}^{alt} - p_{sf}^{ref} \right) \right| 1_{s \in \mathcal{C}_g^d} \tag{15}$$

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Activity-by-Contact S2G strategy

The Activity-by-Contact (ABC)[21,23] (https://github.com/broadinstitute/ABC-Enhancer-Gene-Prediction) S2G strategy is determined by a predictive model for 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[21,23]. Since this S2G strategy is currently unpublished, we provide a brief summary of this approach. 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[67,68] (`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, HUVEC, HCT116, K562, KBM7). This Hi-C matrix (5 Kb) resolution is available here: `ftp://ftp.broadinstitute.org/outgoing/lincRNA/average_hic/average_hic.v2.191020.tar.gz`[23,69]. 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.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.

Number of new annotations analyzed

Fpr the Bonferroni correction, we corrected for 90 annotations tested in all of our analyses, based on 90 new annotations analyzed in our primary analyses (4 + 40 + 40 + 2 + 4 = 90):

- 4 genome-wide deep learning allelic-effect annotations: 2 published (DeepSEAΔ-published and BasenjiΔ-published) and 2 boosted (DeepSEAΔ-boosted and BasenjiΔ-boosted) annotations constructed using fine-mapped SNPs from ref.[18] (vs. matched control SNPs). [Figure 1]
- 40 restricted deep learning allelic-effect annotations integrating DeepSEAΔ-published, BasenjiΔ-published, DeepSEAΔ-boosted and BasenjiΔ-boosted annotations with 10 S2G strategies from Table 1. [Figure 1]
- 40 gene set-specific restricted deep learning allelic-effect annotations, integrating , DeepSEAΔ-boosted and BasenjiΔ-boosted annotations, 10 S2G strategies from Table 1, and SNPs linked to genes specific to 2 gene scores(pLI and PPI-enhancer). [Figure 2]
- 2 Imperio annotations (Imperio-DeepSEA, Imperio-Basenji) (we also analyzed 1 ExPecto-DeepSEA annotation from ref.[4], but this is not a new annotation). [Figure 3]
- 4 gene-set specific Imperio annotations combining Imperio-DeepSEA and Imperio-Basenji models with genes from 2 gene sets (pLI and PPI-enhancer). [Figure 3]

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[13,32]. S-LDSC assumes that

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

$$var\left(\beta_j\right) := \sum_c a_{cj}\tau_c, \tag{16}$$

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

$$E\left[\chi_j^2\right] = N\sum_c l(j,c)\tau_c + 1, \tag{17}$$

where $l(j,c) = \sum_k a_{ck} r_{jk}^2$ 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[34] (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):

$$E = \frac{\frac{h_g^2(c)}{h_g^2}}{\frac{\sum_j a_{cj}}{M}}, \tag{18}$$

where $h_g^2(c)$ 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):

$$\tau_c^\star = \frac{\tau_c sd_c}{\frac{h_g^2}{M}}, \tag{19}$$

where sd_c is the standard error 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 to a 1 standard deviation increase in the value of the annotation.

Combined $\tau^\star$

We use the combined $tau^\star$ metric of ref.[22], quantifying the conditional informativeness of a heritability model (combined τ^*, generalizing the combined $\tau^\star$ metric of ref.[40] 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

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$\tau^{\star}_{comb} = \sqrt{\sum_{m=1}^{M} \tau^{\star 2}_{m} + \sum_{m \neq l} r_{ml} \tau^{\star}_{m} \tau^{\star}_{l}} \tag{20}$$

Here r_{ml} is the pairwise correlation of the annotations m and l, and $r_{ml}\tau^{\star}_{m}\tau^{\star}_{l}$ 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 $\tau^{\star}_{comb}$ 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 4), we define the $\Delta\text{log}l_{\text{SS}}$ 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[32]), where logl_{SS}[41] 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 $\Delta\text{log}l_{\text{SS}}$ using the asymptotic distribution of the Likelihood Ratio Test (LRT) statistic: -2 logl_{SS} follows a χ^2 distribution with degrees of freedom equal to the number of annotations in the focal model, so that $-2\Delta\text{log}l_{\text{SS}}$ follows a χ^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[70]) 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.

Data Availability

All DeepBooost and Imperio annotations are available at `https://alkesgroup.broadinstitute.org/LDSCORE/DeepLearning/Dey_DeepBoost_Imperio/`. The deep learning allelic effect SNP level annotations for DeepSEA and Basenji models are available at `https://alkesgroup.broadinstitute.org/LDSCORE/DeepLearning/`. This work used summary statistics from the UK Biobank study (`http://www.ukbiobank.ac.uk/`). The summary statistics for UK Biobank is available online (`https://data.broadinstitute.org/alkesgroup/UKBB/`). The 1000 Genomes Project Phase 3 data are available at `ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502`. The baseline-LD annotations are available at `https://data.broadinstitute.org/alkesgroup/LDSCORE/`.

Code Availability

The codes for generating DeepBoost and Imperio annotations are available in the Github repository `https://github.com/kkdey/Imperio`.This work primarily uses the S-LDSC software (`https://github.com/bulik/ldsc`). We used publicly available software for DeepSEA (`https://github.com/FunctionLab/ExPecto`) and Basenji (`https://github.com/calico/basenji`) to generate annotations for these respective models.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgments

We thank David Kelley, Steven Gazal, Alexander Gusev and Armin Schoech for helpful discussions. This research was funded by NIH grants U01 HG009379, R01 MH101244, R37 MH107649, R01 MH115676 and R01 MH109978. This research was conducted using the UK Biobank Resource under application 16549.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Tables

Table 1. List of 10 S2G strategies: 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 the Methods section.

S2G strategy	Description	Distal/ Proximal	Size (%)
ABC	Inter-genic SNPs with distal enhancer-gene connections, assessed by Activity-By-Contact[21,23] across blood cell types.	Distal	1.4
TSS	SNPs in predicted Transcription start sites[28,29] overlapping Ensembl gene±5kb 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[24] (mapped to human).	Distal	1.6
eQTL	SNPs with fine-mapped causal posterior probability[25] (CPP) >0.001 in GTEx whole blood.	Distal +Proximal	2.4
Roadmap	SNPs in predicted enhancer-gene links, assessed using Roadmap Epigenomics data[26,30].	Distal	3.2
Promoter	SNPs in promoter regions.	Proximal	4.3
PC-HiC	Distal SNPs with Promoter Capture HiC (PC-HiC)[27] connections to promoter regions in blood cell-types.	Distal	27
5kb	SNPs in ±5kb window around gene body.	Proximal	53
100kb	SNPs in ±100kb window around gene body.	Distal	81

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figures

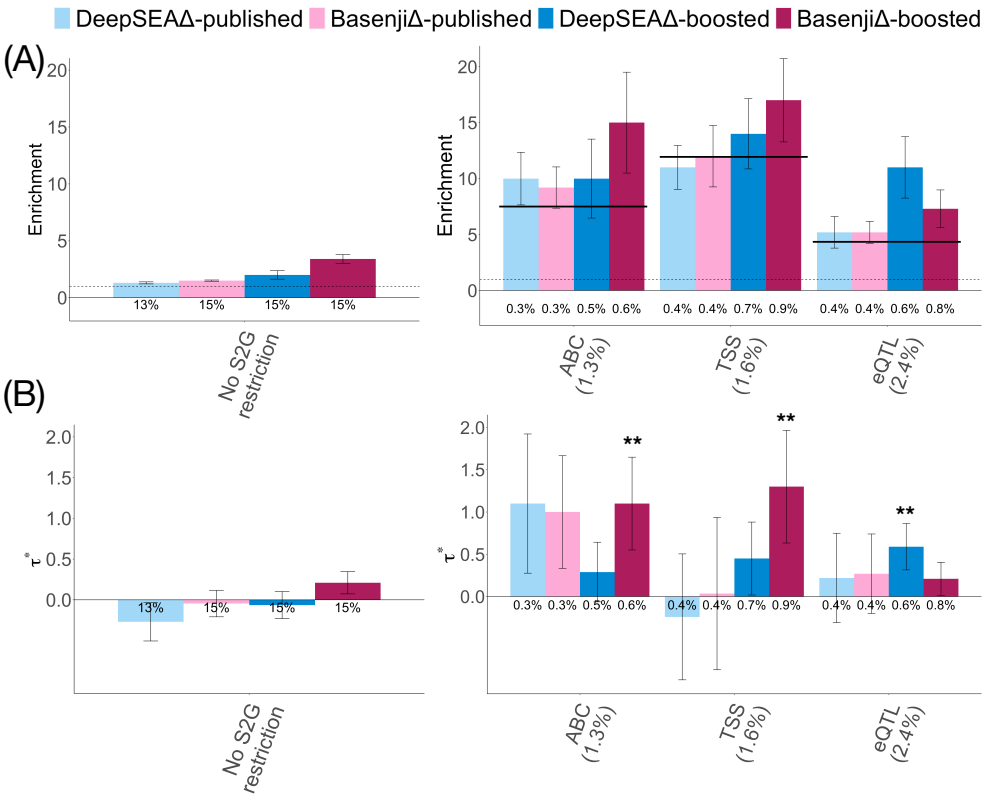

Figure 1. Disease informativeness of published and boosted allelic-effect deep learning annotations restricted to SNPs implicated by functionally informed S2G strategies: (A, Left panel) Heritability enrichment of published and boosted annotations, conditional on the baseline-LD-deep model. Dashed horizontal line denotes no enrichment. (B, Left panel) Standardized effect size ($\tau^\star$) of published and boosted annotations, conditional on the baseline-LD-deep model. (A, Right panel) Heritability enrichment of published-restricted and boosted-restricted annotations, conditional on the baseline-LD-deep-S2G model. Dashed horizontal line denotes no enrichment, solid horizontal lines denote enrichments of underlying S2G annotations. (B, Right panel) Standardized effect size ($\tau^\star$) of published-restricted and boosted-restricted annotations, conditional on the baseline-LD-deep-S2G model. Solid horizontal lines denote $\tau^\star$ values of underlying S2G annotations conditional on the baseline-LD-deep model. Results are meta-analyzed across 11 blood-related traits. ** denotes $P < 0.05/90$. Error bars denote 95% confidence intervals. Numerical results, including results for all 10 S2G strategies analyzed, are reported in Table S7 and Table S10.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

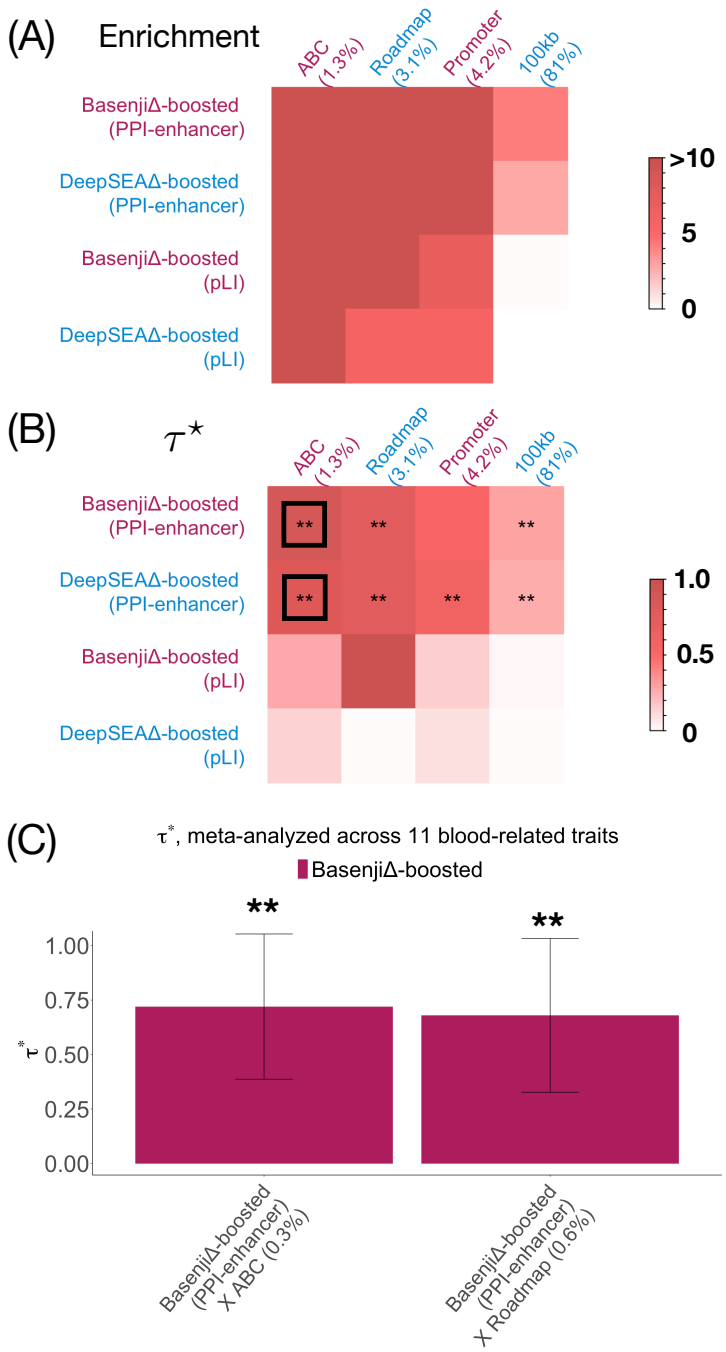

Figure 2. Disease informativeness of gene set-specific boosted-restricted annotations: (A) Heritability enrichment of gene set-specific boosted-restricted annotations, conditional on the baseline-LD-deep-S2G-geneset model. (B) Standardized effect size ($\tau^\star$) of gene set-specific boosted-restricted annotations, conditional on the baseline-LD-deep-S2G-geneset model. (C) Standardized effect size ($\tau^\star$) of the two jointly significant annotations, conditional on the baseline-LD-deep-S2G-geneset model plus both annotations. Results are meta-analyzed across 11 blood-related traits. $\tau^\star$ values less than 0 are displayed as 0 for visualization purposes. ** denotes $P < 0.05/90$. Error bars denote 95% confidence intervals. In panel B, the black box in each row denotes the S2G strategy with highest $\tau^\star$. Numerical results, including results for all 10 S2G strategies analyzed, are reported in Table S15 and Table S17.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

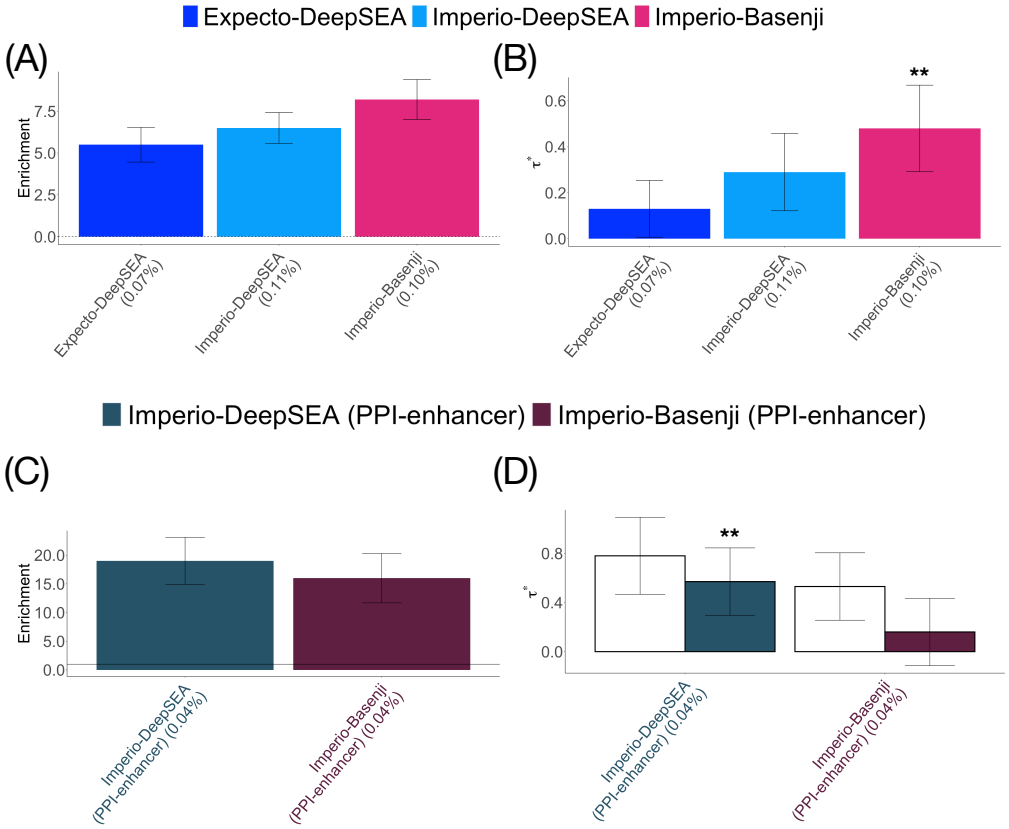

Figure 3. Disease informativeness of allelic-effect annotations based on predictions of gene expression from DNA sequence using S2G linking strategies to prioritize deep learning annotations as features: (A) Heritability enrichment of Imperio allelic-effect annotations, conditional on the baseline-LD-deep-S2G-geneset model. Dashed horizontal line denotes no enrichment. (B) Standardized effect size ($\tau^\star$) of Imperio allelic-effect annotations, conditional on the baseline-LD-deep-S2G-geneset model. (C) Heritability enrichment of gene set-specific Imperio allelic-effect annotations, conditional on the baseline-LD-deep-S2G-geneset model. Dashed horizontal line denotes no enrichment. (D) Standardized effect size ($\tau^\star$) of gene set-specific Imperio allelic-effect annotations, conditional on the baseline-LD-deep-S2G-geneset model. Results are meta-analyzed across 11 blood-related traits. ** denotes $P < 0.05/90$. Error bars denote 95% confidence intervals. Numerical results, including results for both pLI and PPI-enhancer gene sets, are reported in Table S25, Table S27 and Table S29.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

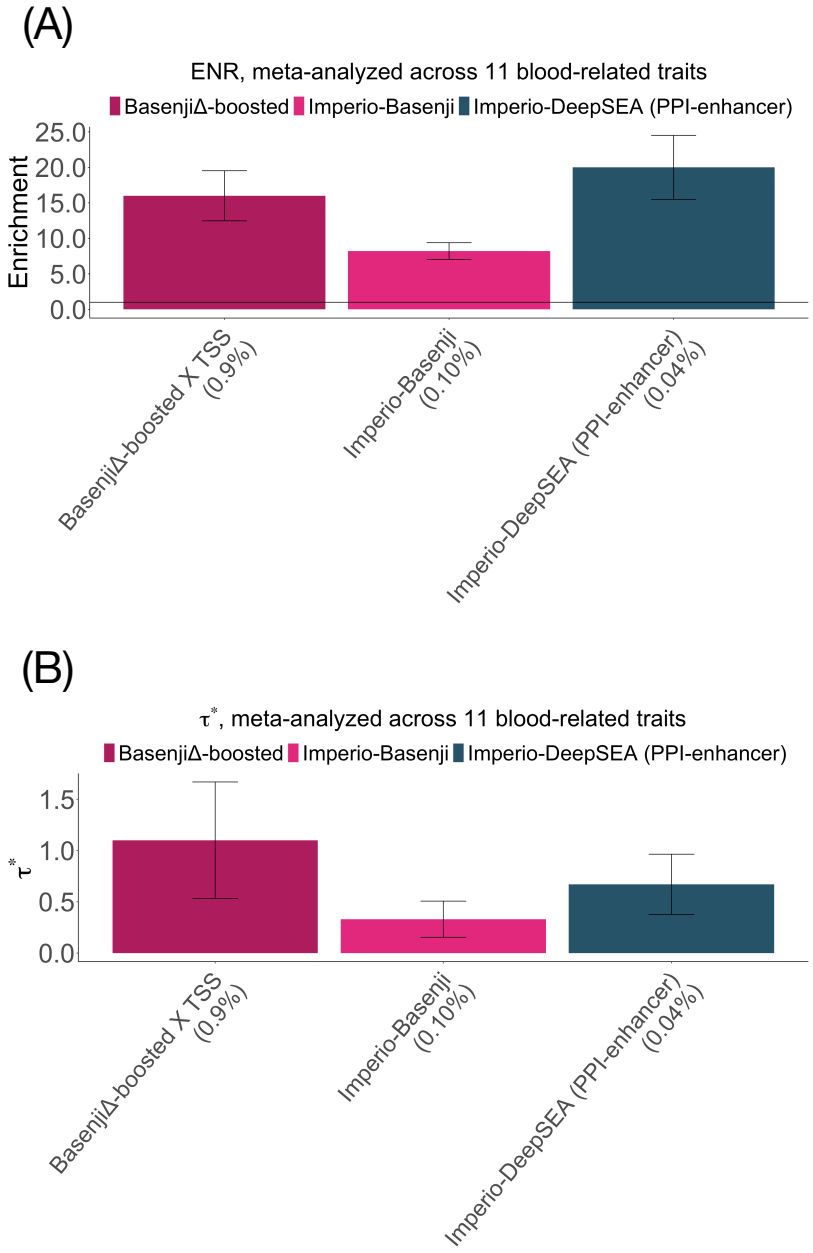

Figure 4. Combined joint model: (A) Heritability enrichment of 3 jointly significant annotations, conditional on the baseline-LD-deep-S2G-geneset model. (B) Standardized effect size ($\tau^{\star}$) conditional on the baseline-LD-deep-S2G-geneset model plus the 3 jointly significant annotations. Results are meta-analyzed across 11 blood-related traits. Error bars denote 95% confidence intervals. Numerical results are reported in Table S38.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Tables

Table S1. List of all blood-related traits: List of 11 blood-related traits (6 autoimmune diseases and 5 blood cell traits) analyzed in this paper.

Trait	Source	N
Auto Immune Traits (Sure)	UKBiobank[37]	459324
Crohn's Disease	Jostins et al., 2012 Nature[71]	20883
Rheumatoid Arthritis	Okada et al., 2014 Nature[72]	37681
Ulcerative Colitis	Jostins et al., 2012 Nature[71]	27432
Lupus	Bentham et al., 2015[73]	14267
Celiac	Dubois et al., 2010[74]	15283
Platelet Count	UKBiobank[37]	444382
Red Blood Cell Count	UKBiobank[37]	445174
Red Blood Cell Distribution Width	UKBiobank[37]	442700
Eosinophil Count	UKBiobank[37]	439938
White Blood Cell Count	UKBiobank[37]	444502

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S2. List of baseline models used in this paper: We report the 6 baseline models or joint models discussed in this paper, along with number of annotations and a brief description.

Models	No. of annotations	Description
baseline-LD	86	Publicly available baseline-LD model[32,33]
baseline-LD-deep	100	baseline-LD plus 14 genome-wide deep-learning, Roadmap, ChromHMM and competing annotations from ref[16].
baseline-LD-deep-S2G	107	baseline-LD-deep + 7 new S2G annotations from Table 1[22].
baseline-LD-deep-PPI-enhancer	114	baseline-LD-deep-S2G + 7 jointly significant Enhancer and PPI-enhancer S2G strategies from joint model in ref[22].
baseline-LD-deep-S2G-geneset	115	baseline-LD-deep-PPI-enhancer + 1 significant pLI S2G annotation (pLI+Roadmap)
Final joint model	118	baseline-LD-deep-S2G-geneset + 3 jointly significant annotations from our analysis in Figure 4.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S3. Additional annotations of baseline-LD-deep model:List of 14 jointly significant annotations from ref.[16] added to the baseline-LD model to create the baseline-LD-deep model. They include 1 non-tissue-specific Basenji allelic-effect annotation, 3 Roadmap annotations, 5 ChromHMM annotations and 5 other annotations.

Annotations
BasenjiΔ-H3K4me3-Max (0.7%)
Roadmap-H3K4me1-Avg (4.4%)
Roadmap-H3K4me1-bloodMax (11.2%)
Roadmap-H3K4me3-bloodAvg (1.8%)
ChromHMM-ActiveEnhancer-1-Avg (0.3%)
ChromHMM-ActiveEnhancer-2-bloodMax (3.5%)
ChromHMM-PromoterDownstream-1-bloodMax (1.1%)
ChromHMM-TranscriptionalRegulator-blood-Avg (0.4%)
ChromHMM-RepressedPolycomb-bloodAvg (1.4%)
pLI-Coding (0.4%)
pLI-TSS (0.3%)
LocalGCcontent (41.0%)
LocalGCcontent-TSS (1.0%)
LocalGCcontent-Coding (0.8%)

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S4. AUROC attained by DeepBoost. We report the AUROC for a gradient boosting model distinguishing fine-mapped SNPs from matched control SNPs using allelic-effect deep-learning annotations from the DeepSEA and Basenji models as features. We consider four sets of fine-mapped SNPs - 8,741 fine-mapped autoimmune disease SNPs[18] (Farh), 4,312 fine-mapped inflammatory bowel disease SNPs[19] (Huang), 1,429 functionally fine-mapped SNPs for 14 blood-related UK Biobank traits[20,37] (Weissbrod), or union of all 14,482 fine-mapped SNPs (Union).

Model	Fine-mapped SNPs	AUROC
DeepSEA	Farh	0.62
	Huang	0.57
	Weissbrod	0.72
	Union	0.64
Basenji	Farh	0.67
	Huang	0.68
	Weissbrod	0.73
	Union	0.69

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S5. AUROC attained by logistic classification instead of XGBoost. We report the AUROC for a logistic regression model distinguishing fine-mapped SNPs from matched control SNPs using allelic-effect deep-learning annotations from the DeepSEA and Basenji models as features. We consider four sets of fine-mapped SNPs - 8,741 fine-mapped autoimmune disease SNPs[18] (Farh), 4,312 fine-mapped inflammatory bowel disease SNPs[19] (Huang), 1,429 functionally fine-mapped SNPs for 14 blood-related UK Biobank traits[20,37] (Weissbrod), or union of all 14,482 fine-mapped SNPs (Union).

Model	Fine-mapped SNPs	AUROC
DeepSEA	Farh	0.60
	Huang	0.56
	Weissbrod	0.69
	Union	0.62
Basenji	Farh	0.65
	Huang	0.67
	Weissbrod	0.69
	Union	0.67

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S6. AUROC attained by DeepBoost using 27 blood cell types only. We report the AUROC for a gradient boosting model distinguishing fine-mapped SNPs from matched control SNPs using using blood-specific allelic-effect deep-learning annotations from the DeepSEA and Basenji models as features. We consider four sets of fine-mapped SNPs - 8,741 fine-mapped autoimmune disease SNPs[18] (Farh), 4,312 fine-mapped inflammatory bowel disease SNPs[19] (Huang), 1,429 functionally fine-mapped SNPs for 14 blood-related UK Biobank traits[20,37] (Weissbrod), or union of all 14,482 fine-mapped SNPs (Union).

Model	Fine-mapped SNPs	AUROC
DeepSEA	Farh	0.60
	Huang	0.56
	Weissbrod	0.67
	Union	0.62
Basenji	Farh	0.64
	Huang	0.65
	Weissbrod	0.70
	Union	0.67

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S7. S-LDSC results for published and boosted deep learning allelic-effect annotations: Standardized Effect sizes ($\tau^\star$) and Enrichment (E) of 2 published allelic-effect deep-learning annotations for 2 deep learning models, DeepSEA and Basenji (DeepSEAΔ-published, BasenjiΔ-published) and 8 boosted allelic-effect deep-learning annotations for the same 2 deep learning models and 4 sets of finemapped SNPs for blood-related traits - 8,741 fine-mapped autoimmune disease SNPs[18] (DeepSEAΔ-boosted, BasenjiΔ-boosted), 4,312 fine-mapped inflammatory bowel disease SNPs[19] (DeepSEAΔ-boosted-Huang, BasenjiΔ-boosted-Huang), 1,429 functionally fine-mapped SNPs for 14 blood-related UK Biobank traits[20] (DeepSEAΔ-boosted-Weissbrod, BasenjiΔ-boosted-Weissbroad), or union of these fine-mapped SNPs (DeepSEAΔ-boosted-Union, BasenjiΔ-boosted-Union). Results are conditioned on 100 baseline-LD-deep annotations. Reports are meta-analyzed across 11 blood and autoimmune traits.

	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
DeepSEAΔ-published	-0.27	0.12	0.02	1.3	0.046	9.5e-05
BasenjiΔ-published	-0.046	0.083	0.58	1.5	0.028	2.2e-08
DeepSEAΔ-boosted	-0.065	0.085	0.44	2	0.19	0.00052
DeepSEAΔ-boosted-Huang	0.14	0.071	0.049	3	0.18	2.1e-09
DeepSEAΔ-boosted-Weissbrod	-0.16	0.1	0.12	2.6	0.2	5e-05
DeepSEAΔ-boosted-Union	-0.13	0.088	0.06	2.6	0.2	0.0001
BasenjiΔ-boosted	0.21	0.07	0.0032	3.4	0.2	3.8e-09
BasenjiΔ-boosted-Huang	0.14	0.071	0.049	3	0.18	2.1e-09
BasenjiΔ-boosted-Weissbrod	-0.094	0.092	0.3	2.6	0.19	6e-05
BasenjiΔ-boosted-Union	0.13	0.066	0.02	3.3	0.19	6.7e-07

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S8. Standardized enrichment of SNP annotations for published and boosted deep learning allelic-effect annotations: Standardized enrichment of 2 published allelic-effect deep-learning annotations for 2 deep learning models, DeepSEA and Basenji (DeepSEAΔ-published, BasenjiΔ-published) and 8 boosted allelic-effect deep-learning annotations for the same 2 deep learning models and 4 sets of finemapped SNPs for blood-related traits - 8,741 fine-mapped autoimmune disease SNPs[18] (DeepSEAΔ-boosted, BasenjiΔ-boosted), 4,312 fine-mapped inflammatory bowel disease SNPs[19] (DeepSEAΔ-boosted-Huang, BasenjiΔ-boosted-Huang), 1,429 functionally fine-mapped SNPs for 14 blood-related UK Biobank traits[20] (DeepSEAΔ-boosted-Weissbrod, BasenjiΔ-boosted-Weissbroad), or union of these fine-mapped SNPs (DeepSEAΔ-boosted-Union, BasenjiΔ-boosted-Union). Results are conditioned on 100 baseline-LD-deep annotations. Reports are meta-analyzed across 11 blood and autoimmune traits.

	$StdE$	se($StdE$)	p($StdE$)
DeepSEAΔ-published	0.1	0.0036	9.5e-05
BasenjiΔ-published	0.098	0.0019	2.2e-08
DeepSEAΔ-boosted	0.71	0.066	0.00052
DeepSEAΔ-boosted-Huang	0.66	0.067	0.0033
DeepSEAΔ-boosted-Weissbrod	0.93	0.071	5e-05
DeepSEAΔ-boosted-Union	0.87	0.072	0.00013
BasenjiΔ-boosted	1.2	0.07	3.8e-09
BasenjiΔ-boosted-Huang	1.1	0.063	2.1e-09
BasenjiΔ-boosted-Weissbrod	1.2	0.054	6.1e-08
BasenjiΔ-boosted-Union	1.3	0.063	1.4e-09

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S9. Additional annotations of baseline-LD-deep-S2G model:List of 7 annotations corresponding to 7 S2G strategies linked to all genes from ref.[16] added to the baseline-LD model to create the baseline-LD-deep-S2G model.

Annotations
ABC (1.4%)
ATAC (1.6%)
eQTL (2.4%)
Roadmap (3.2%)
PC-HiC (27%)
5kb (53%)
100kb (81%)

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S10. S-LDSC results for published-restricted and boosted-restricted deep learning allelic-effect annotations restricted using S2G strategies, conditional on the baseline-LD-deep-S2G model annotations: Standardized Effect sizes ($\tau^\star$) and Enrichment (E) of SNP annotations corresponding to each of DeepSEAΔ-published, BasenjiΔ-published, DeepSEAΔ-boosted and BasenjiΔ-boosted annotations restricted using 10 S2G strategies conditional on 107 baseline-LD-deep-S2G annotations (100 baseline-LD-deep and 7 additional annotations from Table S9). Reports are meta-analyzed across 11 blood-related traits.

DeepSEAΔ-published						
	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
ABC (0.29%)	1.1	0.42	0.0092	10	1.2	3.6e-06
TSS (0.40%)	-0.24	0.38	0.52	11	1	6.3e-06
Coding (0.27%)	0.35	0.28	0.21	8.7	1	7.5e-05
ATAC (0.26%)	0.75	0.22	0.00089	9.1	1.1	6e-07
eQTL (0.37%)	0.22	0.27	0.43	5.2	0.72	0.00028
Roadmap (0.56%)	0.34	0.37	0.36	8.3	1.2	3.4e-07
Promoter (0.79%)	-0.06	0.23	0.8	5.3	0.55	8.1e-05
PC-HiC (3.94%)	-0.28	0.18	0.12	2.7	0.18	6.1e-08
5kb (7.15%)	-0.3	0.12	0.014	1.8	0.067	1.8e-06
100kb (11.0%)	-0.27	0.1	0.0064	1.5	0.051	1.4e-05
BasenjiΔ-published						
	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
ABC (0.32%)	1.0	0.34	0.002	9.2	0.94	7e-06
TSS (0.44%)	0.036	0.46	0.94	12	1.4	1.2e-05
Coding (0.33%)	0.81	0.33	0.014	9.2	1.1	8.6e-05
ATAC (0.31%)	0.58	0.29	0.045	8.4	1.2	1.5e-06
eQTL (0.44%)	0.27	0.24	0.26	5.2	0.5	1.5e-05
Roadmap (0.67%)	0.85	0.56	0.13	8.4	1.1	1.6e-08
Promoter (0.92%)	0.2	0.34	0.55	5.5	0.55	2.3e-05
PC-HiC (4.80%)	0.091	0.25	0.72	2.8	0.15	2.8e-10
5kb (8.58%)	-0.25	0.13	0.05	1.9	0.047	2.4e-08
100kb (13.1%)	-0.091	0.1	0.37	1.6	0.032	1.6e-08
DeepSEAΔ-boosted						
	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
ABC (0.50%)	0.29	0.18	0.1	10	1.8	0.0007
TSS (0.75%)	0.45	0.22	0.037	14	1.6	6.3e-05
Coding (0.47%)	0.44	0.18	0.013	12	1.9	0.00032
ATAC (0.46%)	0.18	0.14	0.2	8.6	1.6	0.00087
eQTL (0.59%)	0.59	0.14	1.5e-05	11	1.4	0.00022
Roadmap (1.0%)	0.2	0.18	0.28	8.3	1.4	0.00038
Promoter (1.33%)	0.29	0.2	0.16	6.3	1.1	0.001
PC-HiC (5.89%)	-0.1	0.13	0.42	3.1	0.49	0.00038
5kb (9.89%)	-0.12	0.088	0.16	2.3	0.22	0.00052
100kb (13.7%)	-0.079	0.087	0.37	2.1	0.2	0.00062
BasenjiΔ-boosted						
	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
ABC (0.63%)	1.1	0.28	0.00016	15	2.3	8.8e-06
TSS (0.88%)	1.3	0.34	0.00011	17	1.9	1.8e-06
Coding (0.59%)	0.55	0.21	0.01	11	1.6	0.00015
ATAC (0.60%)	0.35	0.19	0.059	11	2.1	2.4e-05
eQTL (0.79%)	0.21	0.1	0.036	7.3	0.86	0.00015
Roadmap (1.54%)	0.76	0.26	0.0036	11	1.6	1.1e-07
Promoter (1.63%)	0.19	0.14	0.18	6.6	0.86	4.4e-05
PC-HiC (7.67%)	0.22	0.12	0.082	4.4	0.42	2.4e-09
5kb (11.2%)	0.01	0.07	0.88	3.1	0.15	6.8e-08
100kb (14.5%)	0.15	0.071	0.034	3.3	0.19	1.4e-08

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S11. Standardized enrichment of published-restricted and boosted-restricted deep learning allelic-effect annotations restricted using S2G strategies, conditional on the baseline-LD-deep-S2G model annotations: Standardized enrichment of restricted SNP annotations corresponding to each of DeepSEAΔ-published, BasenjiΔ-published, DeepSEAΔ-boosted and BasenjiΔ-boosted annotations restricted using 10 S2G strategies conditional on 107 baseline-LD-deep-S2G annotations (100 baseline-LD-deep and 7 additional annotations from Table S9). Reports are meta-analyzed across 11 blood-related traits.

DeepSEAΔ-published

	StdE	se(*StdE*)	p(*StdE*)
ABC (0.29%)	0.28	0.034	3.6e-06
TSS (0.40%)	0.38	0.036	6.3e-06
Coding (0.27%)	0.21	0.025	7.5e-05
ATAC (0.26%)	0.22	0.027	6e-07
eQTL (0.37%)	0.14	0.02	0.00028
Roadmap (0.56%)	0.29	0.042	3.4e-07
Promoter (0.79%)	0.23	0.024	8.1e-05
PC-HiC (3.94%)	0.21	0.014	6.1e-08
5kb (7.15%)	0.16	0.0061	1.8e-06
100kb (11.0%)	0.13	0.0044	1.4e-05

BasenjiΔ-published

	StdE	se(*StdE*)	p(*StdE*)
ABC (0.32%)	0.27	0.028	7e-06
TSS (0.44%)	0.44	0.051	1.2e-05
Coding (0.33%)	0.26	0.03	8.6e-05
ATAC (0.31%)	0.23	0.033	1.5e-06
eQTL (0.44%)	0.16	0.016	1.5e-05
Roadmap (0.67%)	0.34	0.044	1.6e-08
Promoter (0.92%)	0.26	0.026	2.3e-05
PC-HiC (4.80%)	0.25	0.013	2.8e-10
5kb (8.58%)	0.18	0.0045	2.4e-08
100kb (13.1%)	0.14	0.0027	1.6e-08

DeepSEAΔ-boosted

	StdE	se(*StdE*)	p(*StdE*)
ABC (0.29%)	0.74	0.13	0.0007
TSS (0.40%)	1.2	0.14	6.3e-05
Coding (0.27%)	0.8	0.13	0.00032
ATAC (0.26%)	0.58	0.11	0.00087
eQTL (0.37%)	0.86	0.11	0.00022
Roadmap (0.56%)	0.83	0.14	0.00038
Promoter (0.79%)	0.72	0.12	0.001
PC-HiC (3.94%)	0.74	0.11	0.00038
5kb (7.15%)	0.68	0.066	0.00052
100kb (11.0%)	0.71	0.067	0.00062

BasenjiΔ-boosted

	StdE	se(*StdE*)	p(*StdE*)
ABC (0.63%)	1.2	0.18	8.8e-06
TSS (0.88%)	1.5	0.18	1.8e-06
Coding (0.59%)	0.84	0.12	0.00015
ATAC (0.60%)	0.86	0.16	2.4e-05
eQTL (0.79%)	0.65	0.076	0.00015
Roadmap (1.54%)	1.4	0.19	1.1e-07
Promoter (1.63%)	0.84	0.11	4.4e-05
PC-HiC (7.67%)	1.2	0.11	2.4e-09
5kb (11.2%)	0.99	0.048	6.8e-088
100kb (14.5%)	1.2	0.067	1.4e-08

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S12. S-LDSC results for published-restricted and boosted-restricted deep learning allelic-effect annotations restricted using S2G strategies, conditional on the baseline-LD-deep model annotations : Standardized Effect sizes ($\tau^\star$) and Enrichment (E) of 40 restricted SNP annotations corresponding to DeepSEAΔ-published, BasenjiΔ-published, DeepSEAΔ-boosted and BasenjiΔ-boosted, annotations restricted using 10 S2G strategies. Results are conditional on 100 baseline-LD-deep annotations. Reports are meta-analyzed across 11 blood-related traits.

AllSNPs						
	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
ABC (1.3%)	0.2	0.096	0.034	7.4	0.83	5.5e-06
TSS (1.6%)	0.49	0.19	0.012	12	0.81	4.3e-07
Coding (1.6%)	0.39	0.13	0.0026	6.9	0.63	1.3e-05
ATAC (1.6%)	0.4	0.11	0.00018	6.6	0.79	5.3e-08
eQTL (2.4%)	0.15	0.06	0.012	4.1	0.35	4.5e-06
Roadmap (3.1%)	0.42	0.12	0.00038	7	0.67	6e-10
Promoter (4.2%)	-0.093	0.097	0.33	4.3	0.34	3.1e-05
PC-HiC (27.3%)	0.02	0.035	0.57	2.1	0.062	1.2e-10
5kb (52%)	0.007	0.02	0.73	1.4	0.028	5.5e-08
100kb (81%)	0.0091	0.0091	0.32	1.2	0.0067	7.4e-10
DeepSEAΔ-published						
	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
ABC (0.29%)	0.32	0.1	0.0013	9.3	1.1	3.6e-06
TSS (0.40%)	0.35	0.2	0.086	13	0.95	4.5e-07
Coding (0.27%)	0.63	0.15	3.9e-05	9.8	0.99	2.3e-05
ATAC (0.26%)	0.58	0.14	3.7e-05	9.5	1.2	2.3e-08
eQTL (0.37%)	0.17	0.074	0.02	5.6	0.46	2.8e-06
Roadmap (0.56%)	0.55	0.15	0.00021	8.7	0.96	1.3e-09
Promoter (0.79%)	-0.019	0.11	0.86	6.5	0.49	2e-06
PC-HiC (3.94%)	0.02	0.049	0.68	2.8	0.1	1.7e-10
5kb (7.15%)	-0.016	0.033	0.63	2	0.038	1.1e-08
100kb (11.0%)	-0.016	0.023	0.48	1.6	0.024	1.9e-09
BasenjiΔ-published						
	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
ABC (0.32%)	0.31	0.1	0.0023	9.1	0.96	3.3e-06
TSS (0.44%)	0.5	0.22	0.022	13	1	6.6e-07
Coding (0.33%)	0.78	0.17	3.5e-06	10	1	1.8e-05
ATAC (0.31%)	0.52	0.13	0.0001	8.8	1.1	2.5e-08
eQTL (0.44%)	0.17	0.071	0.017	5.4	0.41	1.1e-06
Roadmap (0.67%)	0.53	0.14	0.00026	8.3	0.86	3.7e-10
Promoter (0.92%)	0.019	0.11	0.86	6.4	0.47	1.4e-06
PC-HiC (4.80%)	0.034	0.052	0.51	2.8	0.093	6.5e-11
5kb (8.58%)	-0.007	0.031	0.82	2	0.033	5.2e-09
100kb (13.1%)	-0.001	0.021	0.96	1.6	0.022	8.1e-10
DeepSEAΔ-boosted						
	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
ABC (0.50%)	0.36	0.12	0.003	11	1.6	0.00016
TSS (0.75%)	0.58	0.18	0.001	16	1.5	6e-06
Coding (0.47%)	0.69	0.17	3.1e-05	13	1.9	0.00013
ATAC (0.46%)	0.65	0.18	0.00025	14	2.5	3.1e-06
eQTL (0.59%)	0.39	0.1	0.00014	9.7	1.2	7e-05
Roadmap (1.0%)	0.66	0.18	0.00019	12	1.8	1e-06
Promoter (1.33%)	0.13	0.12	0.27	8.5	1.1	1.4e-05
PC-HiC (5.89%)	0.053	0.11	0.62	3.6	0.41	1.2e-06
5kb (9.89%)	-0.056	0.077	0.46	2.5	0.2	2.9e-05
100kb (13.7%)	-0.045	0.072	0.53	2.2	0.16	0.00011
BasenjiΔ-boosted						
	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
ABC (0.63%)	0.5	0.12	5.5e-05	13	1.8	6.8e-06
TSS (0.88%)	1.2	0.27	9.1e-06	18	1.8	2.8e-07
Coding (0.59%)	0.82	0.21	5.9e-05	12	1.6	2.9e-05
ATAC (0.60%)	0.66	0.18	0.00017	13	2.2	2.7e-07
eQTL (0.79%)	0.26	0.087	0.0027	7.9	0.87	2.3e-05
Roadmap (1.54%)	0.71	0.18	4.6e-05	11	1.4	3.5e-09
Promoter (1.63%)	0.23	0.12	0.047	8.1	0.83	1.7e-06
PC-HiC (7.67%)	0.2	0.11	0.07	4.4	0.39	5.4e-10
5kb (11.2%)	0.07	0.068	0.3	3.3	0.15	2.1e-08
100kb (14.5%)	0.17	0.069	0.013	3.4	0.2	6.4e-09

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S13. S-LDSC results for joint model of published-restricted and boosted-restricted deep learning allelic-effect annotations restricted using S2G strategies, conditional on the baseline-LD-deep-S2G model annotations. Standardized Effect sizes ($\tau^\star$) and Enrichment (E) of the significant SNP annotations in a joint model comprising of the marginally significant published-restricted and boosted-restricted SNP annotations corresponding to published and boosted deep learning allelic-effect annotations combined with S2G strategies. Results are conditional on 107 baseline-LD-deep-S2G model annotations (100 baseline-LD-deep and 7 additional annotations from Table S9). Results are meta-analyzed across 11 blood-related traits.

Annotation	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se (E)	p(E)
DeepSEAΔ-boosted $\times$ eQTL (0.6%)	0.54	0.13	3.3e-05	11	1.4	0.00044
BasenjiΔ-boosted $\times$ ABC (0.6%)	0.83	0.23	2e-04	14	2.1	1.4e-05
BasenjiΔ-boosted $\times$ TSS (0.9%)	1.1	0.29	1e-04	16	1.8	2.7e-06

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S14. Additional annotations of baseline-LD-deep-S2G-geneset model:List of 8 jointly significant gene set-specific S2G annotations from ref.[22] added to the baseline-LD-deep-S2G model to create the baseline-LD-deep-S2G-geneset model. They include 7 annotations from the Enhancer-driven+PPI-enhancer joint model in ref[22] and 1 jointly significant pLI S2G annotation.

Annotations
ATAC-distal × Promoter (1.8%)
EDS-binary × 100kb (14.6%)
SEG-GTEx × Coding (0.17%)
PPI-enhancer × ABC (0.58%)
PPI-enhancer × TSS (0.33%)
PPI-enhancer × Coding (0.24%)
PPI-enhancer × ATAC (0.41%)
pLI × Roadmap (0.56%)

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S15. S-LDSC results for gene set-specific boosted-restricted annotations, conditional on baseline-LD-deep-S2G-geneset model annotations: Standardized Effect sizes ($\tau^\star$) and Enrichment (E) of 40 restricted SNP annotations corresponding to 2 deep-learning prioritizations (DeepSEAΔ-boosted and BasenjiΔ-boosted), 2 gene scores (PPI-enhancer and pLI) and 10 S2G strategies, conditional on 115 baseline-LD-deep-S2G-geneset annotations. Reports are meta-analyzed across 11 blood-related traits.

DeepSEAΔ-boosted (PPI-enhancer)						
	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
ABC (0.23%)	0.8	0.18	6.1e-06	21	2.6	1.2e-07
TSS (0.16%)	0.59	0.25	0.016	28	5.1	9.5e-05
Coding (0.08%)	0.66	0.19	0.00061	36	5.7	7.9e-05
ATAC (0.12%)	0.38	0.15	0.013	28	6.1	0.00019
eQTL (0.09%)	0.32	0.13	0.014	17	4.1	0.001
Roadmap (0.31%)	0.75	0.21	0.00026	23	3.9	2.1e-06
Promoter (0.22%)	0.61	0.17	0.00032	23	3.4	1.3e-05
PC-HiC (1.98%)	0.17	0.076	0.03	5.4	0.63	6e-070
5kb (1.47%)	0.053	0.057	0.35	5.7	0.41	3.6e-07
100kb (3.54%)	0.25	0.056	1.1e-05	5.2	0.26	3.2e-08
BasenjiΔ-boosted (PPI-enhancer)						
	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
ABC (0.32%)	0.85	0.18	1.6e-06	19	2.2	2.5e-07
TSS (0.21%)	0.73	0.34	0.034	25	4.3	9.6e-06
Coding (0.11%)	0.75	0.25	0.0025	36	6.1	8.9e-06
ATAC (0.18%)	0.19	0.16	0.24	21	4.8	0.00017
eQTL (0.13%)	0.17	0.1	0.093	12	2.8	0.00019
Roadmap (0.56%)	0.74	0.19	0.00011	18	2.6	1.3e-08
Promoter (0.31%)	0.5	0.19	0.0072	19	3.1	3.4e-06
PC-HiC (2.79%)	0.2	0.077	0.0099	5.6	0.55	3.9e-09
5kb (2.04%)	0.06	0.055	0.27	6.1	0.35	1.4e-08
100kb (4.70%)	0.28	0.054	1.5e-07	5.6	0.3	8.1e-10
DeepSEAΔ-boosted (pLI)						
	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
ABC (0.32%)	0.14	0.15	0.35	11	2	0.00077
TSS (0.41%)	0.088	0.17	0.6	13	1.7	0.00016
Coding (0.24%)	0.26	0.15	0.074	14	2.4	0.00032
ATAC (0.16%)	0.041	0.12	0.73	9.2	2.6	0.022
eQTL (0.18%)	0.14	0.098	0.14	8.4	2.1	0.005
Roadmap (0.39%)	-0.19	0.21	0.37	7.4	2.2	0.066
Promoter (0.68%)	0.091	0.12	0.46	7.2	1.2	0.00055
PC-HiC (3.55%)	0.0045	0.085	0.96	3.6	0.42	1.3e-05
5kb (4.91%)	-0.015	0.049	0.77	3	0.18	9.3e-06
100kb (11.1%)	-0.079	0.063	0.2	2.3	0.16	2.6e-05
BasenjiΔ-boosted (pLI)						
	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
ABC (0.42%)	0.26	0.15	0.075	12	1.7	3.6e-05
TSS (0.49%)	0.19	0.15	0.21	14	1.6	1.8e-05
Coding (0.31%)	0.28	0.14	0.042	14	2	3.7e-05
ATAC (0.20%)	0.11	0.13	0.4	12	3	0.0015
eQTL (0.23%)	0.053	0.099	0.59	7	1.8	0.0031
Roadmap (0.59%)	0.96	0.31	0.0017	16	2.4	5.2e-07
Promoter (0.83%)	0.15	0.14	0.27	8.1	1.2	3.1e-05
PC-HiC (4.78%)	0.061	0.068	0.37	4.1	0.32	1.6e-08
5kb (5.70%)	-0.019	0.049	0.69	3.5	0.16	5.6e-08
100kb (12.5%)	0.015	0.064	0.81	3.1	0.17	1.1e-08

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S16. Standardized enrichment of gene set-specific boosted-restricted annotations, conditional on baseline-LD-deep-S2G-geneset model annotations: Standardized enrichment of 40 restricted SNP annotations corresponding to 2 deep-learning prioritizations (DeepSEAΔ-boosted and BasenjiΔ-boosted), 2 gene scores (PPI-enhancer and pLI) and 10 S2G strategies, conditional on 115 baseline-LD-deep-S2G-geneset annotations. Reports are meta-analyzed across 11 blood-related traits.

DeepSEAΔ-boosted (PPI-enhancer)			
	StdE	se(*StdE*)	p(*StdE*)
ABC (0.23%)	1.2	0.15	1.6e-05
TSS (0.16%)	1.1	0.2	9.5e-05
Coding (0.08%)	1	0.16	7.9e-05
ATAC (0.12%)	0.97	0.21	0.00019
eQTL (0.09%)	0.52	0.12	0.001
Roadmap (0.31%)	1.3	0.22	2.1e-06
Promoter (0.22%)	1.1	0.16	1.3e-05
PC-HiC (1.98%)	0.75	0.088	6e-07
5kb (1.47%)	0.69	0.05	3.6e-07
100kb (3.54%)	0.96	0.049	3.2e-08
BasenjiΔ-boosted (PPI-enhancer)			
	StdE	se(*StdE*)	p(*StdE*)
ABC (0.32%)	1.3	0.2	1.6e-06
TSS (0.21%)	1.1	0.2	9.6e-06
Coding (0.11%)	1.2	0.2	8.9e-06
ATAC (0.18%)	0.91	0.21	0.00017
eQTL (0.13%)	0.46	0.1	0.00019
Roadmap (0.56%)	1.3	0.2	1.3e-08
Promoter (0.31%)	1.1	0.17	3.4e-06
PC-HiC (2.79%)	0.93	0.09	3.9e-09
5kb (2.04%)	0.86	0.05	1.4e-08
100kb (4.70%)	1.2	0.063	8.1e-10
DeepSEAΔ-boosted (pLI)			
	StdE	se(*StdE*)	p(*StdE*)
ABC (0.32%)	0.61	0.11	0.00077
TSS (0.41%)	0.8	0.1	0.00016
Coding (0.24%)	0.66	0.11	0.00032
ATAC (0.16%)	0.33	0.092	0.022
eQTL (0.18%)	0.33	0.082	0.005
Roadmap (0.39%)	0.42	0.13	0.066
Promoter (0.68%)	0.56	0.089	0.00055
PC-HiC (3.55%)	0.63	0.073	1.3e-055
5kb (4.91%)	0.6	0.037	9.3e-06
100kb (11.1%)	0.7	0.049	2.6e-05
BasenjiΔ-boosted (pLI)			
	StdE	se(*StdE*)	p(*StdE*)
ABC (0.42%)	0.76	0.1	3.6e-05
TSS (0.49%)	0.95	0.1	1.8e-05
Coding (0.31%)	0.73	0.11	3.7e-05
ATAC (0.20%)	0.49	0.12	0.0015
eQTL (0.23%)	0.3	0.08	0.0031
Roadmap (0.59%)	1.1	0.17	5.2e-07
Promoter (0.83%)	0.68	0.1	3.1e-05
PC-HiC (4.78%)	0.83	0.066	1.6e-08
5kb (5.70%)	0.76	0.035	5.6e-08
100kb (12.5%)	1	0.055	1.1e-08

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S17. S-LDSC results for joint model of gene set-specific boosted-restricted annotations, conditional on baseline-LD-deep-S2G-geneset model annotations: Standardized Effect sizes ($\tau^\star$) and Enrichment (E) of the Bonferroni significant boosted-restricted S2G annotations linked to PPI-enhancer genes that were marginally significant in Figure 2. The results were conditioned either on 115 baseline-LD-deep-S2G-geneset annotations, or baseline-LD-deep-S2G-geneset plus 3 annotations from Figure 1, or baseline-LD-deep-S2G-geneset plus 3 annotations from Figure 1 plus BasenjiΔ-boosted$\times$Roadmap. Reports are meta-analyzed across 11 blood-related traits.

	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
baseline-LD-deep-S2G-geneset model						
BasenjiΔ-boosted (PPI-enhancer) $\times$ ABC (0.32%)	0.72	0.17	3.2e-05	17	1.8	9.5e-07
BasenjiΔ-boosted (PPI-enhancer) $\times$ Roadmap (0.56%)	0.68	0.18	9.7e-05	17	2.6	1.9e-08
baseline-LD-deep-S2G-geneset + 3 annotations from Figure 1						
BasenjiΔ-boosted (PPI-enhancer) $\times$ ABC (0.32%)	0.69	0.18	0.0001	17	1.8	9.7e-07
BasenjiΔ-boosted (PPI-enhancer) $\times$ Roadmap (0.56%)	0.66	0.18	0.0002	18	2.7	1.4e-08
baseline-LD-deep-S2G-geneset + 3 annotations from Figure 1 + BasenjiΔ-boosted$\times$Roadmap						
BasenjiΔ-boosted (PPI-enhancer) $\times$ ABC (0.32%)	0.63	0.18	0.0003	16	1.5	8.2e-06
BasenjiΔ-boosted (PPI-enhancer) $\times$ Roadmap (0.56%)	0.64	0.18	0.0003	18	2.8	3.3e-08

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S18. S-LDSC results for published-restricted and boosted-restricted deep learning allelic-effect annotations restricted using S2G strategies, conditional on the baseline-LD-deep-S2G model annotations plus the local GC-content annotation and annotations restricted using the local GC-content annotation: Standardized Effect sizes ($\tau^\star$) and Enrichment (E) of restricted SNP annotations corresponding to each of DeepSEAΔ-published, BasenjiΔ-published, DeepSEAΔ-boosted and BasenjiΔ-boosted annotations restricted using the local GC-content and 10 S2G strategies conditional on 100 baseline-LD-deep annotations and unrestricted S2G annotations and S2G annotations restricted using local GC-content annotation. Reports are meta-analyzed across 11 blood-related traits.

DeepSEAΔ-published						
	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
ABC (0.29%)	0.66	0.36	0.069	9.8	1.3	4.4e-06
TSS (0.40%)	-0.36	0.37	0.33	12	0.99	1.8e-06
Coding (0.27%)	0.51	0.28	0.076	9	0.98	4.5e-05
ATAC (0.26%)	0.57	0.24	0.019	9.8	1.3	2.5e-07
eQTL (0.37%)	0.0015	0.29	1	5.1	0.72	0.00046
Roadmap (0.56%)	0.22	0.38	0.56	9.1	1.3	6.1e-08
Promoter (0.79%)	0.13	0.25	0.6	5.7	0.53	1.8e-05
PC-HiC (3.94%)	-0.67	0.2	0.00097	2.5	0.19	2.2e-07
5kb (7.15%)	-0.36	0.13	0.0056	1.8	0.071	1.9e-06
100kb (11.0%)	-0.37	0.13	0.004	1.5	0.053	1.8e-05
BasenjiΔ-published						
	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
ABC (0.32%)	0.93	0.34	0.0068	9.4	1	6e-06
TSS (0.44%)	0.23	0.45	0.6	13	1.1	2.7e-06
Coding (0.33%)	1.1	0.32	0.0004	9.5	1	3.6e-05
ATAC (0.31%)	0.65	0.37	0.081	8.9	1.2	4.8e-07
eQTL (0.44%)	0.16	0.24	0.5	5.4	0.49	9.5e-06
Roadmap (0.67%)	0.91	0.58	0.11	9.4	1.2	2.9e-09
Promoter (0.92%)	0.52	0.31	0.097	5.9	0.52	3.4e-06
PC-HiC (4.80%)	-0.05	0.26	0.85	2.8	0.16	1.8e-10
5kb (8.58%)	-0.23	0.13	0.077	1.9	0.047	2.1e-08
100kb (13.1%)	-0.14	0.1	0.18	1.6	0.032	1.2e-08
DeepSEAΔ-boosted						
	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
ABC (0.50%)	0.22	0.18	0.21	10	1.8	0.00082
TSS (0.75%)	0.5	0.22	0.023	15	1.7	2.4e-05
Coding (0.47%)	0.6	0.2	0.0023	13	1.9	9.2e-05
ATAC (0.46%)	0.11	0.14	0.42	9.4	1.9	0.00078
eQTL (0.59%)	0.57	0.14	8.3e-05	11	1.4	0.00025
Roadmap (1.0%)	0.21	0.18	0.25	9.3	1.4	0.0001
Promoter (1.33%)	0.35	0.21	0.092	7.6	1.2	0.00019
PC-HiC (5.89%)	-0.25	0.13	0.047	2.9	0.5	0.0021
5kb (9.89%)	-0.11	0.09	0.2	2.4	0.23	0.00028
100kb (13.7%)	-0.1	0.089	0.25	2.1	0.2	0.00047
BasenjiΔ-boosted						
	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
ABC (0.63%)	1.1	0.29	0.00018	15	2.3	6.1e-06
TSS (0.88%)	1.4	0.36	6.5e-05	18	1.9	3.7e-07
Coding (0.59%)	0.76	0.23	0.0011	12	1.7	3.7e-05
ATAC (0.60%)	0.38	0.22	0.08	12	2.2	1.1e-05
eQTL (0.79%)	0.17	0.1	0.11	7.4	0.87	0.00011
Roadmap (1.54%)	0.76	0.27	0.0051	12	1.7	3e-08
Promoter (1.63%)	0.31	0.14	0.025	7.6	0.83	4.6e-06
PC-HiC (7.67%)	0.18	0.14	0.19	4.4	0.45	2.4e-09
5kb (11.2%)	0.031	0.068	0.64	3.2	0.15	2.9e-08
100kb (14.5%)	0.13	0.07	0.064	3.3	0.2	6.6e-09

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S19. S-LDSC results for gene set-specific published-restricted annotations, conditional on baseline-LD-deep-S2G-geneset model annotations: Standardized Effect sizes ($\tau^\star$) and Enrichment (E) of 40 published-restricted SNP annotations corresponding to 2 deep-learning models (DeepSEAΔ-published and BasenjiΔ-published), 2 gene scores (PPI-enhancer and pLI) and 10 S2G strategies, conditional on 115 baseline-LD-deep-S2G-geneset annotations. Reports are meta-analyzed across 11 blood-related traits.

DeepSEAΔ-published (PPI-enhancer)						
	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
ABC (0.13%)	0.69	0.18	0.00013	24	3.1	1.6e-05
TSS (0.08%)	0.54	0.22	0.015	22	2.6	2.5e-06
Coding (0.04%)	0.74	0.21	0.00055	27	3.1	4.1e-06
ATAC (0.07%)	0.66	0.22	0.002	23	3.9	2.4e-07
eQTL (0.06%)	0.16	0.082	0.049	9.6	1.5	3e-06
Roadmap (0.18%)	0.61	0.16	0.00015	15	1.9	1.2e-09
Promoter (0.12%)	0.34	0.13	0.006	15	1.5	1.9e-06
PC-HiC (1.28%)	0.13	0.044	0.0041	4	0.25	1.6e-10
5kb (0.93%)	0.018	0.03	0.56	4.5	0.17	5.4e-10
100kb (2.43%)	0.14	0.03	3.4e-06	3.6	0.13	5.3e-11
BasenjiΔ-published (PPI-enhancer)						
	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
ABC (0.14%)	0.78	0.21	0.0001	23	3.5	1.6e-06
TSS (0.09%)	0.65	0.26	0.014	23	2.6	2.6e-06
Coding (0.05%)	0.73	0.22	0.001	26	3	2.6e-06
ATAC (0.08%)	0.51	0.21	0.017	19	3.3	5e-07
eQTL (0.07%)	0.17	0.08	0.038	9.3	1.4	8.7e-07
Roadmap (0.23%)	0.55	0.16	0.00061	13	1.7	1.1e-09
Promoter (0.15%)	0.37	0.14	0.0062	15	1.4	6.1e-07
PC-HiC (1.59%)	0.12	0.046	0.0083	3.9	0.23	5.7e-11
5kb (1.16%)	0.028	0.03	0.35	4.5	0.18	2.9e-10
100kb (2.99%)	0.14	0.029	8.8e-07	3.5	0.13	3.2e-11
DeepSEAΔ-published (pLI)						
	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
ABC (0.17%)	0.064	0.2	0.74	9.6	1.3	8.8e-06
TSS (0.22%)	-0.11	0.15	0.48	12	1.1	9e-06
Coding (0.13%)	0.11	0.13	0.4	11	1.2	1.1e-05
ATAC (0.09%)	0.17	0.094	0.07	10	1.8	4.8e-05
eQTL (0.11%)	0.063	0.075	0.4	5.8	0.93	0.00013
Roadmap (0.21%)	-0.24	0.34	0.47	8	1	2.1e-05
Promoter (0.38%)	-0.0007	0.11	0.99	6.5	0.63	1.3e-05
PC-HiC (2.29%)	-0.0093	0.052	0.86	3	0.12	7.1e-10
5kb (3.25%)	-0.012	0.022	0.59	2.4	0.052	5.7e-09
100kb (8.12%)	-0.048	0.025	0.054	1.9	0.038	3.3e-10
BasenjiΔ-published (pLI)						
	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
ABC (0.19%)	-0.058	0.16	0.72	9.6	1.1	9.6e-06
TSS (0.24%)	-0.13	0.14	0.35	12	1.1	1.4e-05
Coding (0.16%)	0.1	0.14	0.44	11	1.2	7.8e-06
ATAC (0.10%)	0.14	0.091	0.13	9.6	1.6	2.7e-05
eQTL (0.13%)	0.054	0.067	0.41	5.7	0.79	5.1e-05
Roadmap (0.25%)	0.16	0.5	0.75	10	1.4	3e-07
Promoter (0.44%)	-0.0033	0.11	0.98	6.5	0.59	7e-06
PC-HiC (2.79%)	-0.02	0.049	0.68	3	0.11	1.7e-10
5kb (3.89%)	-0.016	0.021	0.46	2.4	0.047	1.9e-09
100kb (9.76%)	-0.045	0.023	0.053	1.9	0.035	8.2e-11

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S20. S-LDSC results for gene set-specific published-restricted annotations, conditional on baseline-LD-deep-S2G-geneset model annotations plus the 2 jointly significant gene set-specific boosted-restricted annotations from Figure 2 Panel C: Standardized Effect sizes ($\tau^\star$) and Enrichment (E) of 40 restricted SNP annotations corresponding to 2 deep-learning models (DeepSEAΔ-published and BasenjiΔ-published), 2 gene scores (PPI-enhancer and pLI) and 10 S2G strategies, conditional on 115 baseline-LD-deep-S2G-geneset annotations and 2 jointly significant annotations from Figure 2 Panel C. Reports are meta-analyzed across 11 blood-related traits.

DeepSEAΔ-published (PPI-enhancer)						
	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
ABC (0.13%)	0.46	0.22	0.035	18	2.1	3.7e-07
TSS (0.08%)	0.55	0.18	0.002	19	3.3	2.6e-06
Coding (0.04%)	0.5	0.21	0.016	23	2.8	3.8e-05
ATAC (0.07%)	0.66	0.22	0.002	23	3.9	2.4e-07
eQTL (0.06%)	0.092	0.078	0.24	8.8	1.5	1.2e-05
Roadmap (0.18%)	-0.086	0.21	0.69	13	1.6	1.4e-08
Promoter (0.12%)	0.13	0.14	0.36	14	1.4	2.5e-05
PC-HiC (1.28%)	0.082	0.038	0.029	3.9	0.23	2.4e-10
5kb (0.93%)	-0.033	0.036	0.36	4.5	0.17	5.4e-10
100kb (2.43%)	0.078	0.04	0.053	3.6	0.13	5.9e-11
BasenjiΔ-published (PPI-enhancer)						
	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
ABC (0.14%)	0.52	0.29	0.073	17	1.8	1.4e-06
TSS (0.09%)	0.099	0.21	0.64	19	2	7.1e-05
Coding (0.05%)	0.43	0.21	0.037	22	2.6	2.7e-05
ATAC (0.08%)	0.19	0.18	0.28	16	2.8	5.2e-06
eQTL (0.07%)	0.095	0.075	0.21	8.6	1.3	3.3e-06
Roadmap (0.23%)	-0.27	0.2	0.18	12	1.6	3.5e-09
Promoter (0.15%)	0.13	0.14	0.38	13	1.4	8.6e-06
PC-HiC (1.59%)	0.073	0.038	0.054	3.8	0.21	7.7e-11
5kb (1.16%)	-0.028	0.037	0.45	4.5	0.19	2.4e-10
100kb (2.99%)	0.083	0.038	0.031	3.6	0.13	3e-11
DeepSEAΔ-published (pLI)						
	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
ABC (0.17%)	-0.013	0.2	0.95	9.1	1.2	1.4e-05
TSS (0.22%)	-0.041	0.14	0.77	12	1.1	8.7e-06
Coding (0.13%)	0.12	0.13	0.36	11	1.2	9.5e-06
ATAC (0.09%)	0.18	0.094	0.054	10	1.8	4.6e-05
eQTL (0.11%)	0.074	0.073	0.31	5.9	0.92	0.00011
Roadmap (0.21%)	-0.3	0.34	0.37	7.7	1	3.8e-05
Promoter (0.38%)	0.017	0.099	0.86	6.5	0.63	1.2e-05
PC-HiC (2.29%)	-0.01	0.052	0.85	3	0.12	7.4e-10
5kb (3.25%)	-0.006	0.022	0.78	2.4	0.053	5.8e-09
100kb (8.12%)	-0.047	0.025	0.058	1.9	0.037	3.7e-10
BasenjiΔ-published (pLI)						
	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
ABC (0.19%)	-0.14	0.17	0.42	9.2	1.1	1.7e-05
TSS (0.24%)	-0.081	0.14	0.56	12	1.1	1.6e-05
Coding (0.16%)	0.11	0.13	0.43	11	1.2	6.9e-06
ATAC (0.10%)	0.15	0.091	0.11	9.5	1.6	2.8e-05
eQTL (0.13%)	0.063	0.065	0.33	5.8	0.78	4.3e-05
Roadmap (0.25%)	-0.42	0.35	0.22	9	1.2	8.9e-07
Promoter (0.44%)	0.013	0.11	0.9	6.5	0.59	6.9e-06
PC-HiC (2.79%)	-0.023	0.049	0.63	2.9	0.11	1.7e-10
5kb (3.89%)	-0.011	0.021	0.6	2.4	0.05	2e-09
100kb (9.76%)	-0.045	0.023	0.049	1.9	0.034	8.6e-11

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S21. Top significant features of Imperio-DeepSEA and Imperio-Basenji models: We report the top 10 chromatin marks (if significant) based on their magnitude of effect size. We consider 4 different models - the Imperio model fitted and evaluated on all genes for DeepSEA (Imperio-DeepSEA) and Basenji (Imperio-Basenji) and the Imperio model fitted and evaluated on PPI-enhancer genes for DeepSEA (Imperio-DeepSEA (PPI-enhancer)) and Basenji (Imperio-Basenji (PPI-enhancer)).

Imperio model	Top features
Imperio-DeepSEA	TF::BRF1::HeLa-S3 (TSS), TF::Znf143::HeLa-S3 (Roadmap), TF::ZNF274::HeLa-S3 (TSS), TF::USF2::HepG2 (Roadmap), TF::USF2::HepG2 (TSS), TF::Znf143::GM12878 (Roadmap), TF::Pol2::GM12878 (Roadmap), TF::BRF1::HeLa-S3 (Roadmap), TF::Mxi1::K562 (Roadmap), TF::BHLHE40::K562 (Roadmap)
Imperio-Basenji	CAGE:spleen, adult, pool1 (Roadmap), CAGE:spinal cord, adult, donor10252 (TSS), DNASE:SW480 (TSS), HISTONE:H3K9ac neural progenitor cell derived from H9 (TSS), CAGE:Smooth muscle cells - airway, control, donor1 (Roadmap), CAGE:epithelioid sarcoma cell line:HS-ES-1 (Roadmap), DNASE:LNCaP clone FGC treated with 17B-hydroxy-17-methylestra-4,9,11-trien-3-one (Roadmap), HISTONE:H3K9ac keratinocyte female (TSS), HISTONE:H3K27me3 fibroblast of arm male adult (53 years) (TSS), HISTONE:H3K27me3 PC9 (TSS)
Imperio-DeepSEA (PPI-enhancer)	TF::Znf143::GM12878 (TSS), TF::ZNF274::K562 (TSS), TF::BRF1::HeLa-S3 (TSS), TF::Rad21::A549 (TSS), TF::USF2::HepG2 (TSS), TF::Znf143::HeLa-S3 (TSS), TF::BHLHE40::K562 (TSS), TF::YY1::NT2-D1 (TSS), forskolin::ERRA::HepG2 (Roadmap), TF::BRCA1::HeLa-S3 (TSS)
Imperio-Basenji (PPI-enhancer)	HISTONE:H3K4me3 kidney epithelial cell (TSS), CAGE:spleen, adult, pool1 (Roadmap), CAGE:Macrophage monocyte derived, donor3 (Roadmap), HISTONE:H3K4me2 OCI-LY3 (TSS), CAGE:CD14+ Monocytes, donor2 (Roadmap), CAGE:peripheral neuroectodermal tumor cell line:KU-SN (Roadmap), CAGE:caudate nucleus, adult, donor10252 (Roadmap), CAGE:CD34 cells differentiated to erythrocyte lineage (Roadmap), CAGE:carcinoid cell line:NCI-H1770 (Roadmap), CAGE:immature langerhans cells, donor1 (TSS)

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S22. Comparison of 5 Imperio models utilizing a single S2G strategy with respect to using all 5 S2G strategies: We perform the Imperio prediction model for DeepSEA and Basenji features corresponding to only one of the 5 S2G strategies, and compare the resulting model fit with that of the full Imperio model corresponding to all 5 S2G strategies. We use two measures of model fit - the r^2 metric and the correlation of predicted expression (corr.pred) with original expression on the genes of a holdout chromosome (chr8).

S2G	Imperio-DeepSEA		Imperio-Basenji	
	r^2	corr.pred	r^2	corr.pred
ABC	0.36	0.60	0.39	0.66
Roadmap	0.33	0.63	0.37	0.68
TSS	0.59	0.69	061	0.71
5kb	0.40	0.66	0.52	0.73
100kb	0.29	0.52	0.35	0.61
All	0.66	0.72	0.69	0.76

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S23. Proportion of cis-heritability captured by Imperio and ExPecto predictions of gene expression across individuals: Results are averaged across all 22,020 genes for the 2 Imperio models (Imperio-DeepSEA and Imperio-Basenji) and the ExPecto-DeepSEA model.

Model	Correlation
Imperio-DeepSEA	0.82
Imperio-Basenji	0.79
ExPecto-DeepSEA	0.75

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S24. Correlation of boosted-restricted deep learning allelic-effect annotations restricted using S2G strategies with Imperio annotations: We report the correlation of boosted-restricted deep learning allelic-effect annotations restricted using S2G strategies with Imperio annotations for DeepSEA and Basenji deep learning models.

restricted S2G	Imperio-DeepSEA	Imperio-Basenji
BasenjiΔ-boosted $\times$ ABC	0.12	0.10
BasenjiΔ-boosted $\times$ TSS	0.15	0.16
BasenjiΔ-boosted $\times$ Coding	0.03	0.04
BasenjiΔ-boosted $\times$ ATAC	0.05	0.07
BasenjiΔ-boosted $\times$ eQTL	0.07	0.08
BasenjiΔ-boosted $\times$ Roadmap	0.11	0.13
BasenjiΔ-boosted $\times$ Promoter	0.12	0.15
BasenjiΔ-boosted $\times$ PC-HiC	0.11	0.12
BasenjiΔ-boosted $\times$ 5kb	0.12	0.14
BasenjiΔ-boosted $\times$ 100kb	0.11	0.13
DeepSEAΔ-boosted $\times$ ABC	0.13	0.10
DeepSEAΔ-boosted $\times$ TSS	0.17	0.17
DeepSEAΔ-boosted $\times$ Coding	0.03	0.05
DeepSEAΔ-boosted $\times$ ATAC	0.05	0.06
DeepSEAΔ-boosted $\times$ eQTL	0.07	0.07
DeepSEAΔ-boosted $\times$ Roadmap	0.13	0.13
DeepSEAΔ-boosted $\times$ Promoter	0.12	0.15
DeepSEAΔ-boosted $\times$ PC-HiC	0.12	0.11
DeepSEAΔ-boosted $\times$ 5kb	0.12	0.12
DeepSEAΔ-boosted $\times$ 100kb	0.11	0.10

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S25. S-LDSC results for Imperio and ExPecto annotations, conditional on the baseline-LD-deep-S2G-geneset model annotations: Standardized Effect sizes ($\tau^\star$) and Enrichment (E) of Imperio annotations for DeepSEA (Imperio-DeepSEA) and Basenji (Imperio-Basenji) along with a similarly defined ExPecto (ExPecto-DeepSEA) annotation. Results were conditional on 115 baseline-LD-deep-S2G-geneset annotations. Reports are meta-analyzed across 11 blood-related traits.

	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
Imperio-DeepSEA (0.11%)	0.29	0.086	0.00075	6.5	0.47	4.9e-06
Imperio-Basenji (0.10%)	0.48	0.096	4.9e-07	8.2	0.61	6.1e-07
ExPecto-DeepSEA (0.07%)	0.13	0.063	0.045	5.5	0.53	2.1e-05

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S26. Standardized enrichment results for Imperio and ExPecto annotations conditional on the baseline-LD-deep model annotations: Standardized Enrichment of Imperio annotations for DeepSEA (Imperio-DeepSEA) and Basenji (Imperio-Basenji) along with a similarly defined ExPecto (ExPecto-DeepSEA) annotations. Results were conditional on 115 baseline-LD-deep-S2G-geneset annotations. Reports are meta-analyzed across 11 blood-related traits.

	StdE	se(StdE)	p(StdE)
Imperio-DeepSEA (0.11%)	0.045	0.0033	6.7e-06
Imperio-Basenji (0.10%)	0.057	0.0043	7.3e-07
ExPecto-DeepSEA (0.07%)	0.03	0.003	5.5e-05

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S27. S-LDSC results for gene-set specific Imperio annotations, conditional on the baseline-LD-deep-S2G-geneset model annotations: Standardized Effect sizes ($\tau^\star$) and Enrichment (E) of gene set-specific Imperio annotations corresponding to 2 deep learning models (DeepSEA and Basenji) and 2 gene sets (pLI and PP-enhancer). Results were conditional on 115 baseline-LD-deep-S2G-geneset annotations. Reports are meta-analyzed across 11 blood-related traits.

	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
Imperio-DeepSEA (PPI-enhancer) (0.04%)	0.78	0.16	1.5e-06	20	2.3	3.5e-08
Imperio-DeepSEA (pLI) (0.04%)	0.21	0.098	0.035	8.7	1.1	1.5e-05
Imperio-Basenji (PPI-enhancer) (0.04%)	0.53	0.14	0.0003	18	2.4	4.7e-07
Imperio-Basenji (pLI) (0.05%)	0.24	0.093	0.011	9	1.1	6.9e-07

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S28. Standardized enrichment results for gene-set specific Imperio annotations conditional on the baseline-LD-deep-S2G-geneset model: Standardized Enrichment of gene set-specific Imperio annotations corresponding to 2 deep learning models (DeepSEA and Basenji) and 2 gene sets (pLI and PP-enhancer) and. Results were conditional on 115 baseline-LD-deep-S2G-geneset annotations. Reports are meta-analyzed across 11 blood-related traits.

	StdE	se(StdE)	p(StdE)
Imperio-DeepSEA (PPI-enhancer) (0.04%)	0.11	0.013	3.5e-08
Imperio-DeepSEA (pLI) (0.04%)	0.044	0.0055	1.5e-05
Imperio-Basenji (PPI-enhancer) (0.04%)	0.081	0.011	4.7e-07
Imperio-Basenji (pLI) (0.05%)	0.051	0.0061	6.9e-07

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S29. S-LDSC results for joint model of gene set-specific Imperio annotations conditional on the baseline-LD-deep-S2G-geneset model annotations: Joint Standardized Effect sizes ($\tau^\star$) and Enrichment (E) of the 2 marginally significant gene-set specific Imperio annotations, Imperio-DeepSEA (PPI-enhancer) and Imperio-Basenji (PPI-enhancer). Results were conditional on 115 baseline-LD-deep-S2G-geneset annotations. Reports are meta-analyzed across 11 blood-related traits.

	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
Imperio-DeepSEA (PPI-enhancer) (0.04%)	0.57	0.14	3.9e-05	19	2.1	5.1e-08
Imperio-Basenji (PPI-enhancer) (0.04%)	0.16	0.14	0.26	16	2.2	3.5e-06

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S30. S-LDSC results for gene-set specific Imperio annotations, conditional on baseline-LD-deep-S2G-geneset model annotations plus 1 Imperio-Basenji annotation from Figure 3B: Standardized Effect sizes ($\tau^\star$) and Enrichment (E) of gene-set specific Imperio annotations corresponding to 2 deep learning models (DeepSEA and Basenji) and 2 gene-sets (pLI and PP-enhancer). Results were conditional on 115 baseline-LD-deep-S2G-geneset model annotations and 1 Imperio-Basenji annotation from Figure 3B. Reports are meta-analyzed across 11 blood-related traits.

	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
Imperio-DeepSEA (PPI-enhancer) (0.04%)	0.65	0.14	7.9e-06	19	2.3	4.2e-08
Imperio-DeepSEA (pLI) (0.04%)	0.02	0.11	0.89	7.8	1.2	1e-04
Imperio-Basenji (PPI-enhancer) (0.04%)	0.42	0.14	0.002	18	2	5.5e-07
Imperio-Basenji (pLI) (0.05%)	0.04	0.10	0.66	8.9	1.16	2.4e-06

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S31. S-LDSC results for Imperio+ExPecto annotations, conditional on the baseline-LD-deep-S2G-geneset model annotations: Standardized Effect sizes ($\tau^\star$) and Enrichment (E) of the Δ_s SNP level annotations computed using a combination both ExPecto[4] and Imperio features for DeepSEA and Basenji models. Results were conditional either on 115 baseline-LD-deep-S2G-geneset annotations or baseline-LD-deep-S2G-geneset plus 1 Imperio-Basenji annotation from Figure 3B. Reports are meta-analyzed across 11 blood-related traits.

conditional on the baseline-LD-deep-S2G-geneset model

	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
ExPecto-Imperio-DeepSEA (0.08%)	0.26	0.087	0.0023	6.1	0.42	2.5e-06
ExPecto-Imperio-Basenji (0.08%)	0.41	0.093	9.7e-06	7.1	0.5	3.6e-07

conditional on the baseline-LD+geneset+Imperio-Basenji-BLD model

	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
ExPecto-Imperio-DeepSEA (0.08%)	-0.042	0.12	0.73	5.9	0.42	3.6e-06
ExPecto-Imperio-Basenji (0.08%)	-0.17	0.13	0.2	7.1	0.49	3.6e-07

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S32. S-LDSC results for partially restricted gene set-specific Imperio annotations defined by restricting only the fitting of feature weights, conditional on the baseline-LD-deep-S2G-geneset model annotations: Standardized Effect sizes ($\tau^\star$) and Enrichment (E) of the intermediate Imperio (Imperio-int1) annotations computed by using all genes for fitting the model and gene sets for computing the expression allelic effects for 2 deep learning models (DeepSEA and Basenji) and 2 gene sets (pLI and PPI-enhancer). Results were conditional either on 115 baseline-LD-deep-S2G-geneset annotations or baseline-LD-deep-S2G-geneset plus 2 significant Imperio annotations from Figure 3. Reports are meta-analyzed across 11 blood-related traits.

conditional on the baseline-LD-deep-S2G-geneset model

	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
Imperio-int1-DeepSEA (PPI-enhancer) (0.04%)	0.62	0.13	3.9e-06	16	1.7	2.5e-08
Imperio-int1-Basenji (PPI-enhancer) (0.04%)	0.53	0.18	0.0025	19	2.8	1.4e-06
Imperio-int1-DeepSEA (pLI) (0.04%)	0.091	0.092	0.32	6.6	0.74	6.7e-06
Imperio-int1-Basenji (pLI) (0.04%)	0.065	0.1	0.53	7.3	1.1	3.6e-05

conditional on baseline-LD-deep-S2G-geneset plus 2 significant Imperio annotations from Figure 3

	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
Imperio-int1-DeepSEA (PPI-enhancer) (0.04%)	0.21	0.23	0.36	16	1.7	3.3e-08
Imperio-int1-Basenji (PPI-enhancer) (0.04%)	0.077	0.14	0.63	17	2.6	1.2e-05
Imperio-int1-DeepSEA (pLI) (0.04%)	0.028	0.088	0.75	7.9	0.79	8.7e-07
Imperio-int1-Basenji (pLI) (0.04%)	-0.034	0.51	0.91	8.5	1.2	6.3e-06

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S33. S-LDSC results for partially restricted gene set-specific Imperio annotations defined by restricting only the gene expression predictions, conditional on the baseline-LD-deep-S2G-geneset model annotations: Standardized Effect sizes ($\tau^\star$) and Enrichment (E) of the intermediate Imperio (Imperio-int-2) annotations computed by using genes in a geneset for fitting the model and all genes for computing the expression allelic effects for 2 deep learning models (DeepSEA and Basenji) and 2 gene sets (pLI and PPI-enhancer). Results were conditional either on 115 baseline-LD-deep-S2G-geneset annotations or baseline-LD-deep-S2G-geneset plus 2 significant Imperio annotations from Figure 3. Reports are meta-analyzed across 11 blood-related traits.

conditional on the baseline-LD-deep-S2G-geneset model

	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
Imperio-int2-DeepSEA (PPI-enhancer) (0.08%)	0.46	0.14	0.00099	8.6	0.85	3.7e-06
Imperio-int2-Basenji (PPI-enhancer) (0.10%)	0.59	0.1	3.3e-09	9.4	0.66	2.2e-07
Imperio-int2-DeepSEA (pLI) (0.09%)	0.51	0.11	2.9e-06	8.8	0.75	5.7e-07
Imperio-int2-Basenji (pLI) (0.10%)	0.48	0.11	4.3e-06	8.6	0.61	4.3e-07

conditional on baseline-LD-deep-S2G-geneset plus 2 significant Imperio annotations from Figure 3

	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
Imperio-int2-DeepSEA (PPI-enhancer) (0.08%)	-0.12	0.20	0.54	8.2	0.62	8.5e-07
Imperio-int2-Basenji (PPI-enhancer) (0.10%)	0.44	0.26	0.09	9.6	0.67	1.5e-07
Imperio-int2-DeepSEA (pLI) (0.09%)	0.023	0.24	0.89	8.5	0.62	5.7e-07
Imperio-int2-Basenji (pLI) (0.10%)	0.021	0.18	0.91	8.4	0.78	9.9e-07

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S34. S-LDSC results for annotations defined by the total number of genes linked to each SNP by each S2G strategy, conditional on the baseline-LD-deep-S2G-geneset model annotations: Standardized Effect sizes ($\tau^\star$) and Enrichment (E) of the number of genes (N_{sd}) linked to each SNP s by the S2G strategy d. The number of genes was thresholded at 5 and annotations were standardized to probabilistic scale. Results were conditional either on 115 baseline-LD-deep-S2G-geneset annotations. Reports are meta-analyzed across 11 blood-related traits.

	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
ABC (1.93%)	0.23	0.19	0.24	9.2	0.81	8.9e-10
Roadmap (0.87%)	-0.16	0.27	0.54	5.6	0.49	1.2e-07
TSS (2.31%)	-0.067	0.043	0.11	2.8	0.18	3.5e-05
5kb (13%)	-0.025	0.033	0.46	1.5	0.034	4.1e-08
100kb (38%)	0.0047	0.023	0.84	1.6	0.027	3.4e-10

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S35. S-LDSC results for annotations defined by the number of PPI-enhancer genes linked to each SNP by each S2G strategy, conditional on the baseline-LD-deep-S2G-geneset model annotations: Standardized Effect sizes ($\tau^\star$) and Enrichment (E) of the number of PPI-enhancer genes linked to each SNP s by the S2G strategy d. The number of genes was thresholded at 5 and annotations were standardized to probabilistic scale. Results were conditional either on 115 baseline-LD-deep-S2G-geneset annotations. Reports are meta-analyzed across 11 blood-related traits.

	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
ABC (0.44%)	0.40	0.15	0.009	17	1.8	4.8e-09
Roadmap (0.38%)	0.32	0.14	0.019	11	1.5	3.4e-10
TSS (0.39%)	0.03	0.076	0.69	8.9	0.84	1.7e-07
5kb (3.1%)	0.0032	0.023	0.89	3.4	0.13	2.9e-10
100kb (5.3%)	0.082	0.026	0.0013	3	0.12	2.8e-11

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S36. S-LDSC results for Imperio annotations defined using the maximum across genes proximal to the annotated SNPs (instead of the sum), conditional on the baseline-LD-deep-S2G-geneset model annotations plus the two significant annotations from Figure 3B,D: Standardized Effect sizes ($\tau^\star$) and Enrichment (E) of the SNP level annotations computed using the maximum across genes proximal to the annotated SNPs (instead of the sum), conditional on the baseline-LD-deep-S2G-geneset model annotations plus the two significant annotations from Figure 3B,D. Reports are meta-analyzed across 11 blood-related traits.

	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
Imperio-DeepSEA (0.58%)	-0.21	0.07	0.0035	1.4	0.15	0.058
Imperio-Basenji (0.44%)	0.13	0.056	0.023	2.6	0.14	2.3e-06
Imperio-DeepSEA (PPI-enhancer) (0.34%)	-0.11	0.055	0.045	1.6	0.17	0.0095
Imperio-Basenji (PPI-enhancer) (0.34%)	0.17	0.082	0.036	2.9	0.16	5.5e-07

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S37. S-LDSC results for Whole blood MaxCPP annotations conditional on different baseline models: Standardized Effect sizes ($\tau^\star$) and Enrichment (E) of Whole blood MaxCPP (MaxCPP) annotations. Results were conditional on either 115 baseline-LD-deep-geneset, or 107 baseline-LD-deep-S2G, or 100 baseline-LD-deep annotations. Reports are meta-analyzed across 11 blood-related traits.

	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
conditional on the baseline-LD-deep-S2G-geneset model						
MaxCPP (0.07%)	0.27	0.12	0.027	10	2.1	0.0015
conditional on the baseline-LD-deep-S2G model						
MaxCPP (0.07%)	0.29	0.12	0.018	10.5	2.1	0.0001
conditional on the baseline-LD-deep model						
MaxCPP (0.07%)	0.36	0.15	0.014	12	2.5	0.0009

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S38. S-LDSC results for combined joint model: Standardized Effect sizes ($\tau^\star$) and Enrichment (E) in a joint model comprising of significant SNP annotations from Figure 1, Figure 2 and Figure 3. Only results for the 3 jointly Bonferroni significant annotations are reported. The results were conditioned on 115 baseline-LD-deep-S2G-geneset annotations. Reports are meta-analyzed across 11 blood-related traits.

	$\tau^\star$	se($\tau^\star$)	p($\tau^\star$)	E	se(E)	p(E)
BasenjiΔ-boosted × TSS (0.9%)	1.1	0.29	0.0001	16	1.8	8.3e-07
Imperio-Basenji (0.10%)	0.33	0.09	0.0001	8.2	0.61	6.4e-07
Imperio-DeepSEA (PPI-enhancer) (0.04%)	0.67	0.15	5e-06	20	2.3	3.9e-08

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S39. $\Delta\log l_{SS}$ results for the the combined joint model and other heritability models. We report $\Delta\log l_{SS}$ derived from the $\log l_{\mathrm{SS}}$ metric, proposed in ref.[41], for the different heritability models studied in this paper: baseline-LD, baseline-LD-deep, baseline-LD-deep-S2G, baseline-LD-S2G-geneset and combined joint model in Figure 4 (Table S2). We compute $\Delta\log l_{SS}$ as the difference in $\log l_{\mathrm{SS}}$ for each model with respect to s baselineLD-no-funct model with 17 annotations that include no functional annotations[41,75]. We also report the percentage increase in $\Delta\log l_{SS}$ for each model over the baseline-LD model. We do not report AIC as the number of annotations are not too different to alter conclusions based on just the $\log l_{\mathrm{SS}}$. We report three summary $\Delta\log l_{\mathrm{SS}}$ results - one averaged across 30 UK Biobank traits[75] (All), one averaged across 6 blood-related traits from UK Biobank (Blood) and one averaged across the other 24 non blood related traits from UK Biobank (Non-blood) (Table S40).

Model	$\Delta\log l_{\mathrm{SS}}$ (All)	% incr. over baseline-LD (All)	pval (All)	$\Delta\log l_{\mathrm{SS}}$ (Blood)	% incr. over baseline-LD (Blood)	pval (Blood)	$\Delta\log l_{\mathrm{SS}}$ (Non-blood)	% incr. over baseline-LD (Non-blood)	pval (Non-blood)
baseline-LD (n=86)	1379	0	-	2668	0	-	1121	0	-
baseline-LD-deep (n=100)	1501	8.8%	5e-44	2997	12%	2e-131	1179	5.1%	4e-18
baseline-LD-deep-S2G (n=107)	1542	12%	1e-56	3102	16%	3e-170	1218	8.6%	5e-30
baseline-LD-deep-S2G -geneset (n=115)	1582	14.7%	2e-66	3142	17.8%	8e-181	1259	12%	1e-40
Combined joint model (Figure 4) n=118)	1603	16.2%	7e-75	3209	20.3%	8e-207	1280	14.2%	8e-49

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S40. List of UKBiobank traits used for $\log l_{SS}$ calculations. The list consists of 6 blood-related traits and 24 non blood-related traits.

Trait	Category	N
disease AID Sure	Blood	459324
blood Eosinophil count	Blood	439938
Blood Platelet count	Blood	444382
Blood Red count	Blood	445174
Blood RBC distr. width	Blood	442700
blood White count	Blood	444502
reproduction Menarche Age	Non-blood	242278
reproduction Menopause Age	Non-blood	143025
Body balding	Non-blood	208336
Body BMIz	Non-blood	457824
cov EDU Years	Non-blood	454813
disease Dermatology	Non-blood	459324
disease Allergy Eczema	Non-blood	458699
lung FVCzSmoke	Non-blood	371949
lung FEV1FVCzSmoke	Non-blood	371949
pigment Hair	Non-blood	452720
bmd Heel Tscorez	Non-blood	445921
body Heightz	Non-blood	458303
disease Hi-chol-self-rep	Non-blood	459324
disease Hypothyroidism self rep	Non-blood	459324
Other Morning-person	Non-blood	41050
Mental Neuroticism	Non-blood	372066
disease Respiratory ENT	Non-blood	459324
pigment Skin	Non-blood	453609
cov Smoking Status	Non-blood	457683
pigment Sunburn	Non-blood	344229
bp SystolicadjMedz	Non-blood	422771
pigment Tanning	Non-blood	449984
disease T2D	Non-blood	459324
body WHRadjBMIz	Non-blood	458417

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Figures

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

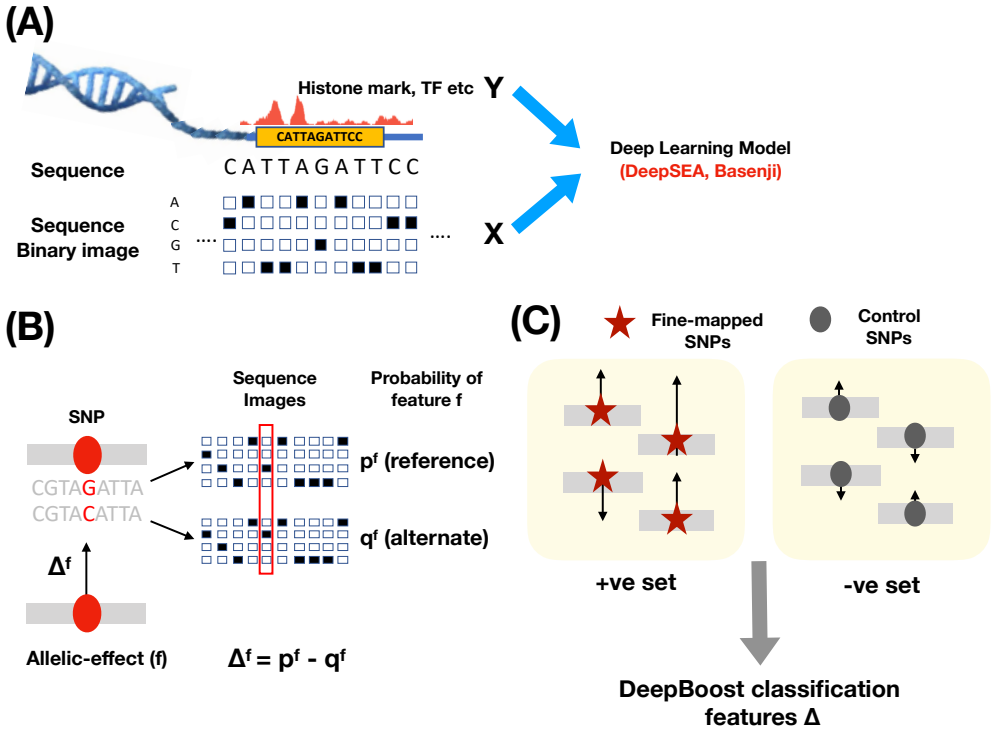

Figure S1. **Illustration of the DeepBoost model**: (A) An overview of a sequence based genomic deep learning model like DeepSEA and Basenji, that trains on sequence images for a region and the chromatin features around that region using a deep Convolutional Neural Net (CNN) model. (B) Illustration of how the alelic effect annotation for a particular feature f is computed at a SNP s. The number of features f is 2,002 for the DeepSEA model and 4,229 for the Basenji model. The length of the vertical arrow at the SNP site denotes the magnitude of the allelic effect and its direction represents the sign (up and down for positive and negative allelic effect respectively). (C) Illustration of the DeepBoost classification model where we classify positive set of fine-mapped SNPs from the negative set of matched controls using the allelic effect features.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

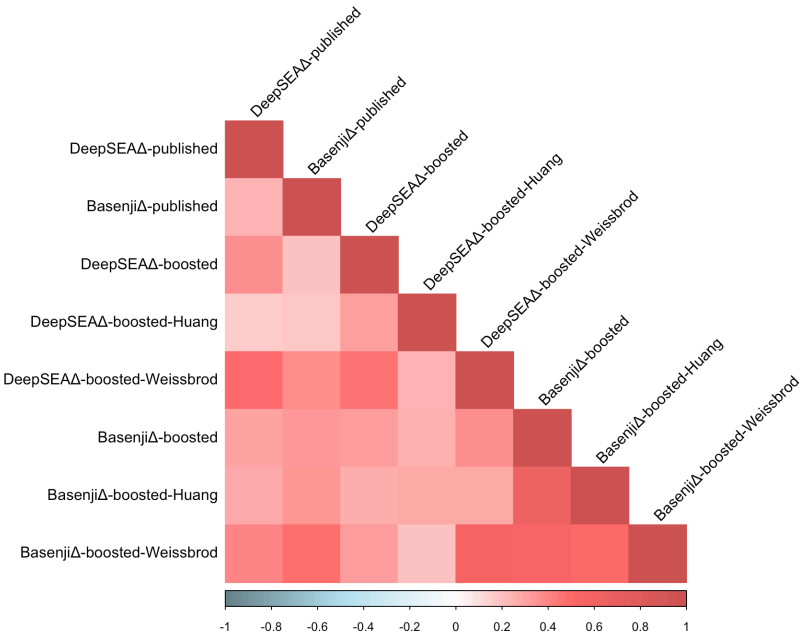

Figure S2. **Correlation between DeepSEA, Basenji and DeepBoost annotations for different fine-mapped sets of SNPs**. Correlation matrix of DeepSEA and Basenji annotations from Figure S1 and 6 DeepBoost annotations corresponding to 2 deep learning models (DeepSEA and Basenji) and 3 sets of fine-mapped SNPs in autoimmune diseases[18–20]. The correlations range from slightly positive to medium high positive values.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

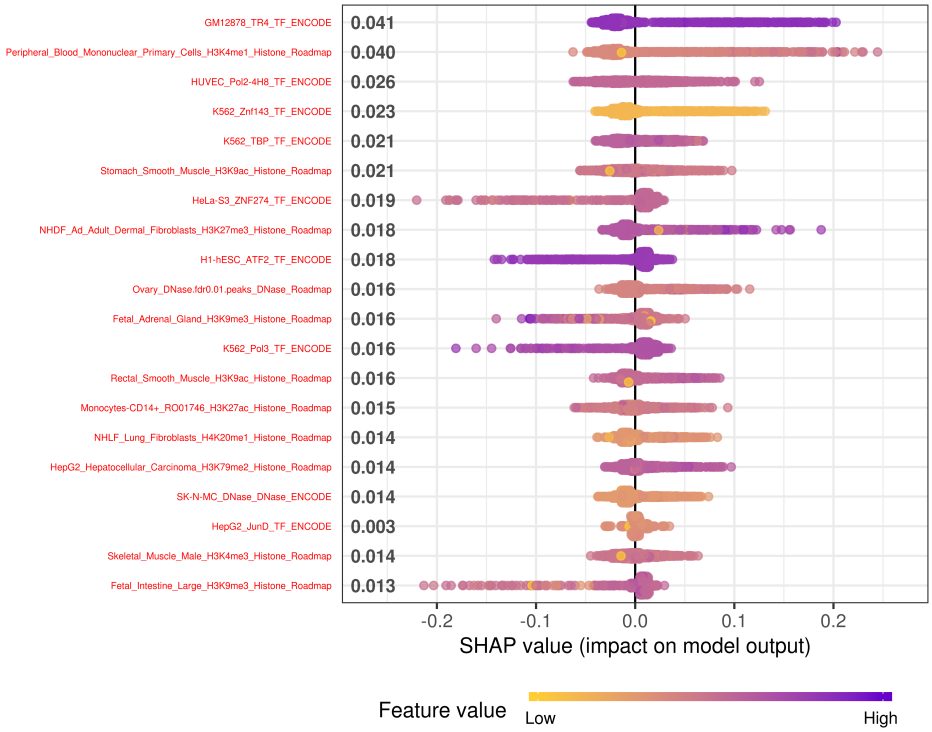

Figure S3. **Feature importance of boosted annotations for DeepBoost using the DeepSEA model**. We applied SHAP[36] to assess which deep learning features were most important for the prediction of boosted annotations using the DeepSEA (Methods). We report the top 20 features with signed SHAP scored ordered from top to bottom based on importance.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

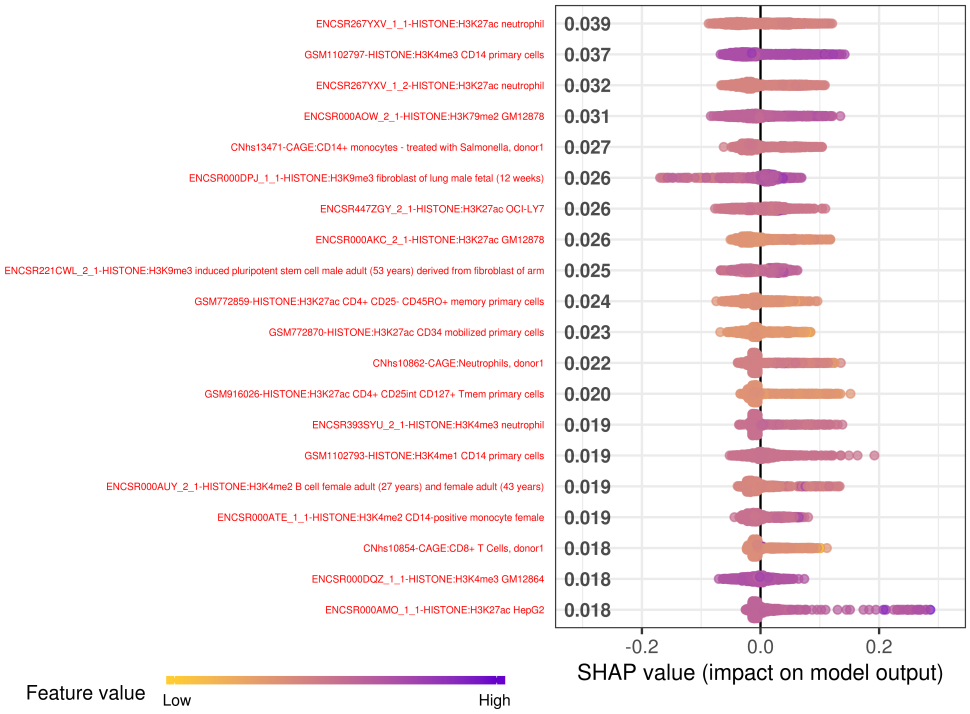

Figure S4. **Feature importance of boosted annotations for DeepBoost using the Basenji model**. We applied SHAP[36] to assess which deep learning features were most important for the prediction of boosted annotations using the Basenji (Methods). We report the top 20 features with signed SHAP scored ordered from top to bottom based on importance.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

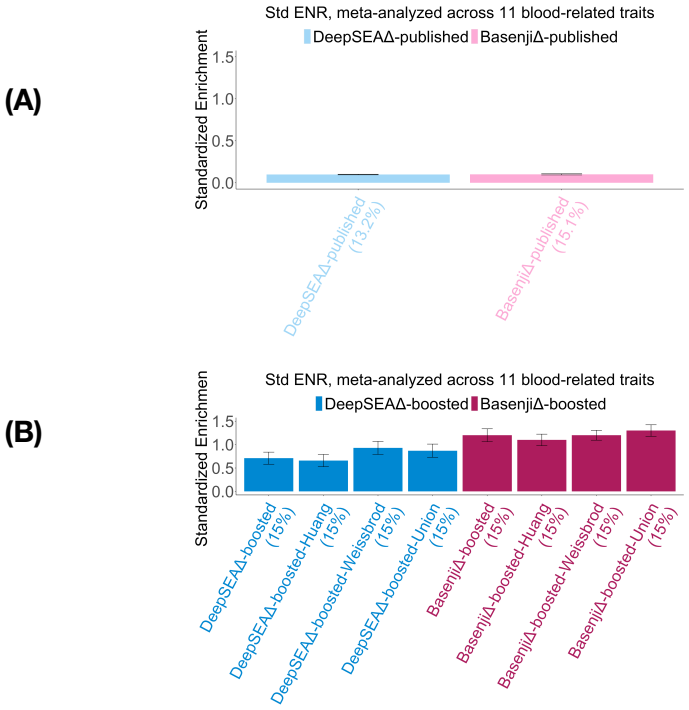

Figure S5. **Standardized enrichment of SNP annotations for published and boosted deep learning allelic-effect annotations**. Barplot representing standardized enrichment metric, as proposed in ref.[76], for SNP annotations corresponding to (A) 2 published annotations for 2 deep learning models DeepSEA and Basenji and (B) 8 boosted annotations corresponding to 4 sets of fine-mapped SNPs and the above 2 deep learning models, conditional on the baseline-LD-deep model annotations.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

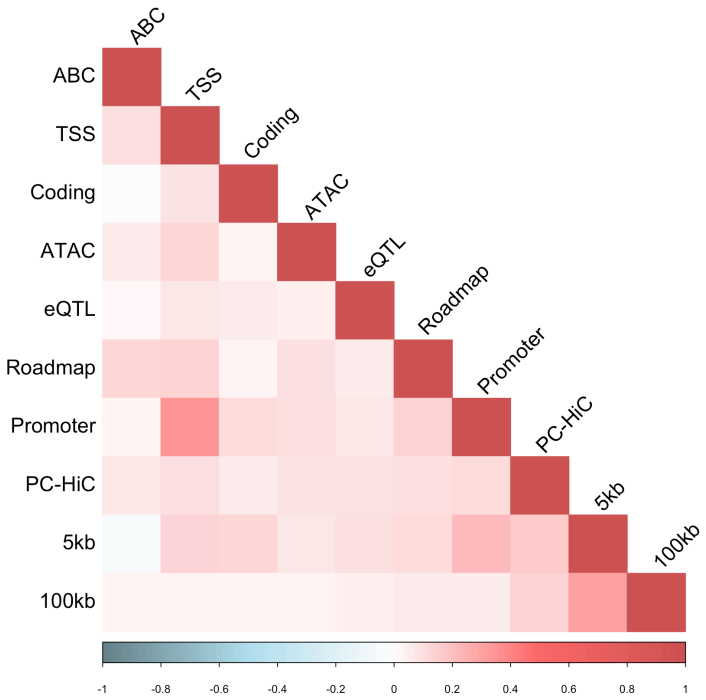

Figure S6. Correlation between S2G annotations. Correlation matrix of S2G annotations derived from all 10 SNP-to-gene (S2G) linking strategies (Table 1), as defined by the sets of SNPs linked to all genes.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

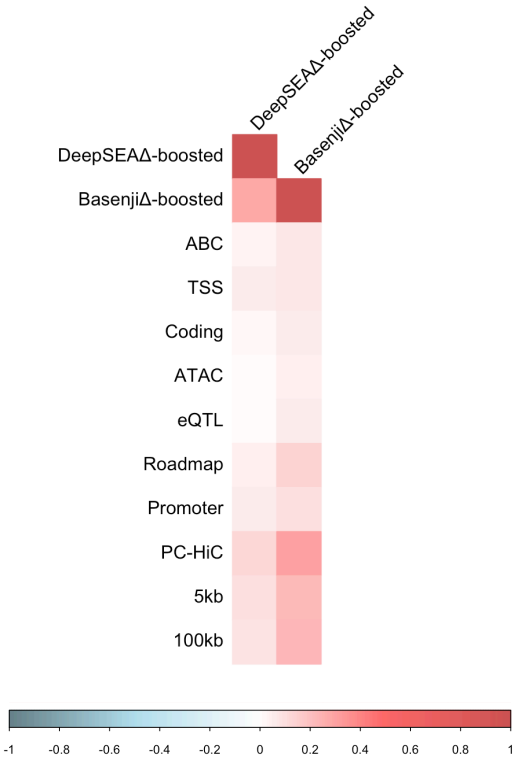

Figure S7. Correlation between DeepSEA-boosted and Basenji-boosted annotations and S2G annotations. Correlation matrix of DeepSEA-boosted and Basenji-boosted annotations with the 10 S2G annotations. The correlations range from weakly positive to moderately positive.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

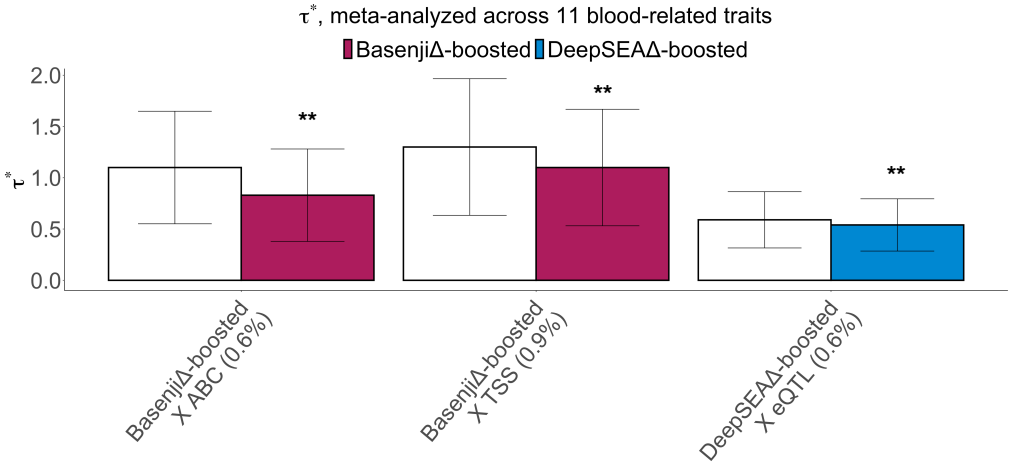

Figure S8. **S-LDSC results for joint model of published-restricted and boosted-restricted deep learning allelic-effect annotations restricted using S2G strategies, conditional on the baseline-LD-deep-S2G model annotations**. Standardized effect size ($\tau^\star$) conditional on baseline-LD-deep-S2G and other significant restricted S2G annotations (right column, shading) compared to the effect size from Figure 1 Panel B right panel (left column, white). Results are meta-analyzed across 11 blood-related traits. ** denotes $P < 0.05/90$. Error bars denote 95% confidence intervals. Numerical results are reported in Table S13.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

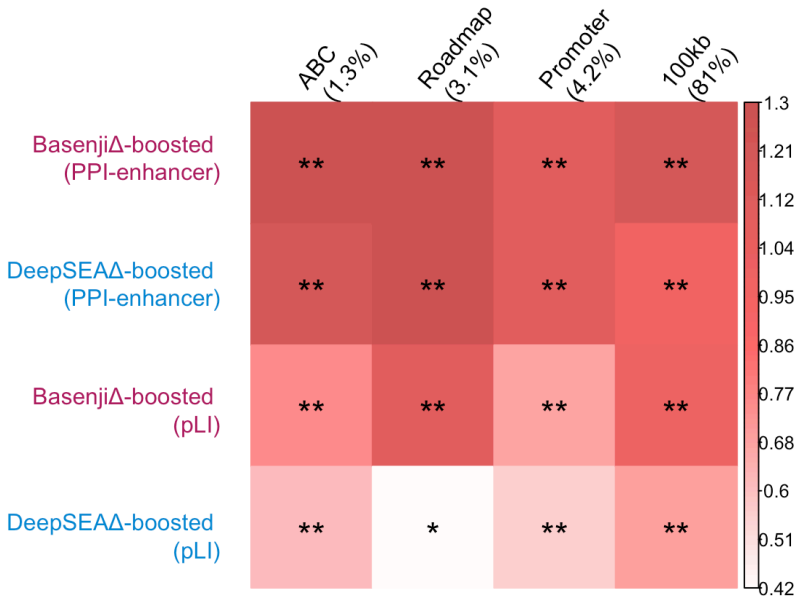

Figure S9. **Standardized enrichment of gene set-specific boosted-restricted annotations, conditional on baseline-LD-deep-S2G-geneset model annotations**. Standardized enrichment metric, as proposed in ref.[76], for 40 SNP annotations corresponding to 2 gene scores (PPI-enhancer[22], pLI[31]) with 10 S2G annotations prioritized by 2 boosted deep-learning allelic-effect annotations (DeepSEA-boost and Basenji-boost. ** denotes $P < 0.05/90$. Error bars denote 95% confidence intervals. Numerical results are reported in Table S16.

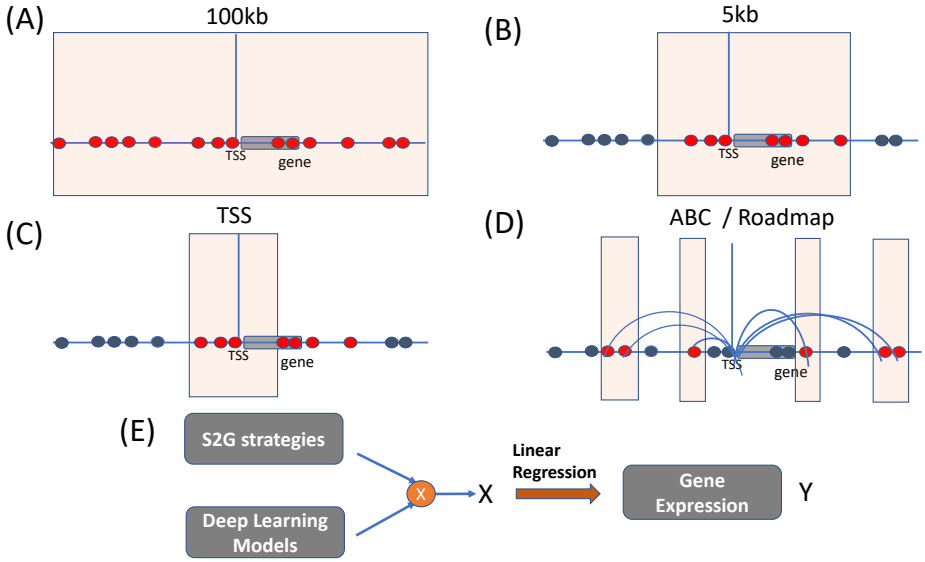

Figure S10. **Illustration of the Imperio model**: A schematic representation of the different S2G straategie used in the Imperio model : (A) 100kb, (B) 5kb, (C) TSS and (D) ABC or Roadmap. (E) Illustration of how the deep learning variant level or allelic effect annotations are combined with these S2G strategies to generate the featues which are used as predictors in a regression model with GTEx Whole blood expression (log CPM) used as response.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

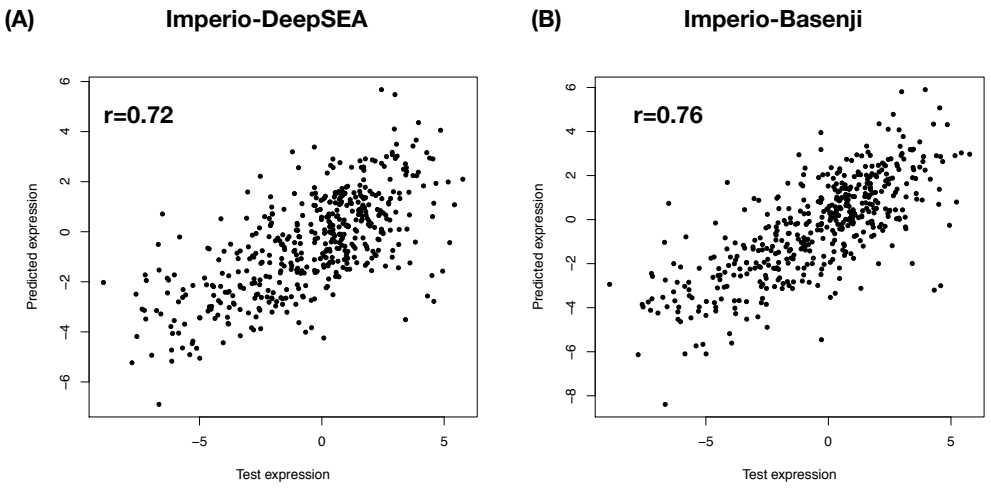

Figure S11. **Accuracy of Imperio in predicting gene expression across genes on chromosome 8**. For both Imperio-DeepSEA and Imperio-Basenji, we plot predicted expression vs. observed log RPKM expression, for 990 genes on chromosome 8.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

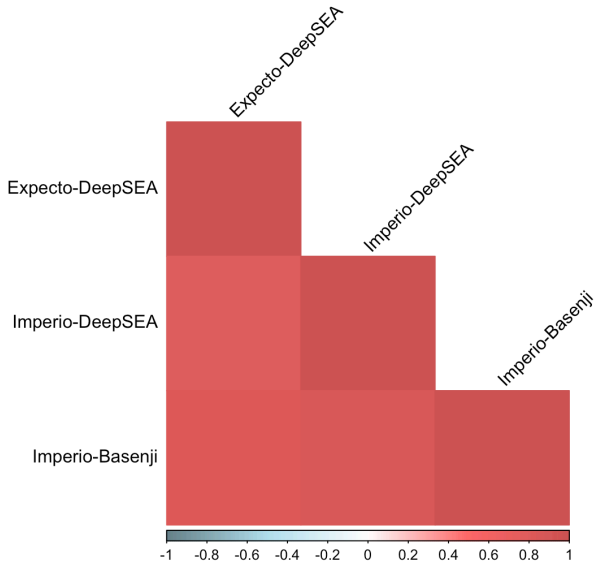

Figure S12. **Correlation in predicted expression between Imperio and ExPecto models**. Correlation in predicted expression for 990 chr8 genes used as held-out test set for the ExPecto method[4] and the two Imperio models corresponding to DeepSEA and Basenji deep learning models.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

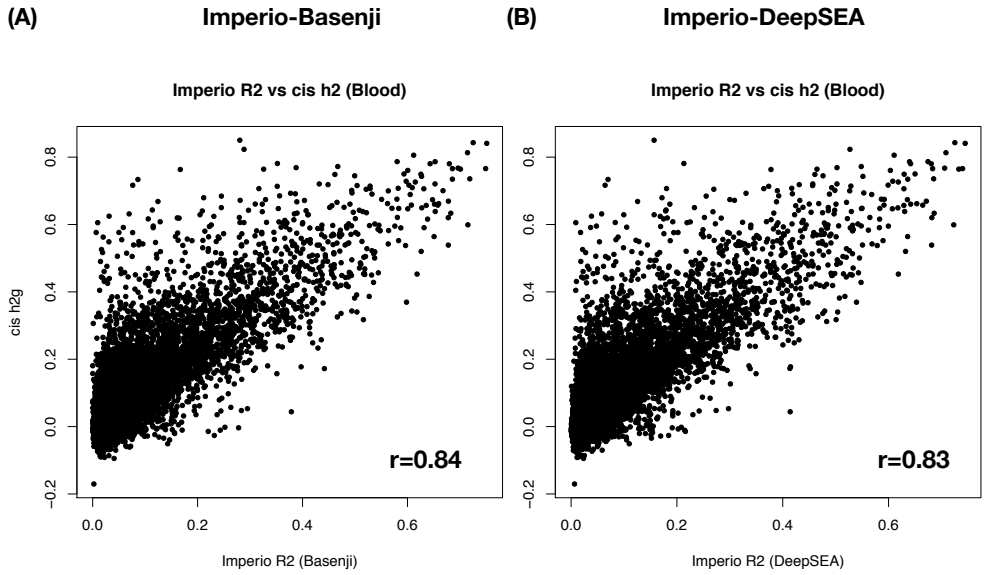

Figure S13. **Comparison of Imperio prediction r^2 for predictions of gene expression across individuals vs. cis-heritability**. We plot the prediction r^2 for the inter-individual model comprising of the Imperio predicted expression effects (see Methods) and the per-gene cis-heritability in Whole blood as estimated from trancriptiome wide association studies[43] for two deep learning models - DeepSEA (Panel A) and Basenji (Panel B).

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

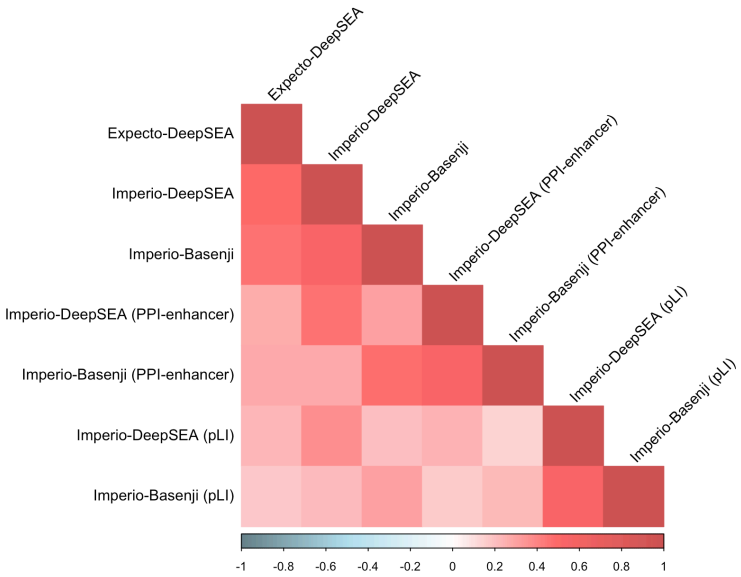

Figure S14. Correlation between all-genes ExPecto, all-genes Imperio, and gene set-specific Imperio annotations. Correlation matrix of Whole blood MaxCPP[25], Whole blood ExPecto[25], and 6 Imperio annotations corresponding to 2 deep learning models (DeepSEA and Basenji) and three sets of genes (all genes, pLI genes and PPI-enhancer genes). The correlations range from slightly positive to medium high positive values.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. B. Alipanahi, A. Delong, M.T. Weirauch, and B.J. Frey. Predicting the sequence specificities of DNA-and RNA-binding proteins by deep learning. *Nature biotechnology*, 33:831–838, 2015.

2. J. Zhou and O.G. Troyanskaya. Predicting effects of noncoding variants with deep learning–based sequence model. *Nature methods*, 12:931–934, 2015.

3. D.R. Kelley, J. Snoek, and J.L. Rinn. Basset: learning the regulatory code of the accessible genome with deep convolutional neural networks. *Genome research*, 26:990–999, 2016.

4. J. Zhou, C.L. Theesfeld, K. Yao, K.M. Chen, A.K. Wong, and O.G. Troyanskaya. Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. *Nature genetics*, 50:1171–1179, 2018.

5. D.R. Kelley, Y.A. Reshef, M. Bileschi, D. Belanger, C.Y. McLean, and J. Snoek. Sequential regulatory activity prediction across chromosomes with convolutional neural networks. *Genome research*, 28:739–750, 2018.

6. J. Zou et al. A primer on deep learning in genomics. *Nature genetics*, 51:12–18, 2019.

7. G. Eraslan et al. Deep learning: new computational modelling techniques for genomics. *Nature Reviews Genetics*, 20:389–403, 2019.

8. J. Zhou et al. Whole-genome deep-learning analysis identifies contribution of noncoding mutations to autism risk. *Nature genetics*, 51(6):973–980, 2019.

9. L.A. Hindorff et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences*, 106(23):9362–9367, 2009.

10. M.T. Maurano et al. Systematic localization of common disease-associated variation in regulatory dna. *Science*, 337(6099):1190–1195, 2012.

11. G. Trynka et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nature genetics*, 45(2):124–130, 2013.

12. J.K. Pickrell. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *The American Journal of Human Genetics*, 94(4):559–573, 2014.

13. H.K. Finucane, B. Bulik-Sullivan, A. Gusev, G. Trynka, Y. Reshef, P.R. Loh, V. Anttila, H. Xu, C. Zang, K. Farh, and S. Ripke. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature genetics*, 47:1228–1235, 2015.

14. A.L. Price, C.C. Spencer, and P. Donnelly. Progress and promise in understanding the genetic basis of common diseases. *Proceedings of the Royal Society B: Biological Sciences*, 282(1821):20151684, 2015.

15. P.M. Visscher et al. 10 years of GWAS discovery: biology, function, and translation. *The American Journal of Human Genetics*, 101(1):5–22, 2017.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

16. K.K. Dey et al. Evaluating the informativeness of deep learning annotations for human complex diseases. *bioRxiv (accepted in principle, Nat Commun)*, page p.784439, 2019.

17. T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. *In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, ACM:785–794, 2016.

18. K.K.H. Farh et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*, 518:337–343, 2015.

19. H. Huang et al. Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature*, 547:173–178, 2017.

20. O. Weissbrod et al. Functionally-informed fine-mapping and polygenic localization of complex trait heritability. *bioRxiv*, page p.807792, 2020.

21. J. Nasser et al. Genome-wide maps of enhancer regulation connect risk variants to disease genes. *bioRxiv*, 2020.

22. K.K. Dey et al. Unique contribution of enhancer-driven and master-regulator genes to autoimmune disease revealed using functionally informed SNP-to-gene strategies. *bioRxiv*, 2020.

23. C.P. Fulco et al. Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. *Nature Genetics*, 51:1664–1669, 2019.

24. H. Yoshida et al. The cis-Regulatory Atlas of the Mouse Immune System. *Cell*, 176:897–912, 2019.

25. F. Hormozdiari et al. Leveraging molecular quantitative trait loci to understand the genetic architecture of diseases and complex traits. *Nature genetics*, 50(7):1041–1047, 2018.

26. Y. Liu et al. Evidence of a recombination rate valley in human regulatory domains. *Genome Biology*, 18:p.193, 2017.

27. B.M. Javierre et al. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell*, 167:1369–1384, 2016.

28. M.M. Hoffman, J. Ernst, S.P. Wilder, A. Kundaje, R.S. Harris, and M. Libbrecht. A method to predict the impact of regulatory variants from DNA sequence. *Nucleic acids research*, 41:827–841, 2012.

29. M.M. Hoffman, O.J. Buske, J. Wang, Z. Weng, J.A. Bilmes, and W.S. Noble. Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nature methods*, 9:473–476, 2012.

30. J. Ernst et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, 473(7345):43–49, 2011.

31. M. Lek, K.J. Karczewski, E.V. Minikel, K.E. Samocha, E. Banks, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, 536:285–291, 2016.

32. S. Gazal et al. Linkage disequilibrium–dependent architecture of human complex traits shows action of negative selection. *Nature genetics*, 49 (10):1421–1427, 2017.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

33. S. Gazal, C. Marquez-Luna, H.K. Finucane, and A.L. Price. Reconciling s-ldsc and ldak models and functional enrichment estimates. *Nature genetics*, 51(8):1202–1204, 2019.

34. 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*, 526(7571):68–74, 2015.

35. A. Kundaje et al. Integrative analysis of 111 reference human epigenomes. *Nature*, 518:317–330, 2015.

36. S.M. Lundberg and S.I. Lee. A unified approach to interpreting model predictions. *In Advances in neural information processing systems*, pages 4765–4774, 2017.

37. C. Bycroft et al. The uk biobank resource with deep phenotyping and genomic data. *Nature*, 562(7726):203–209, 2018.

38. F.I. Hormozdiari et al. Functional disease architectures reveal unique biological role of transposable elements. *Nature Communications*, 10(1):4054, 2019.

39. GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature*, 550(7675):204–213, 2017.

40. B. van de Geijn, H. Finucane, S. Gazal, F. Hormozdiari, T. Amariuta, and X Liu. Annotations capturing cell-type-specific TF binding explain a large fraction of disease heritability. *Human Molecular Genetics*, 29:1057–1067, 2020.

41. D. Speed, J. Holmes, and D.J Balding. Evaluating and improving heritability models using summary statistics. *Nature Genetics*, 52:458–462, 2020.

42. E.R. Gamazon et al. A gene-based association method for mapping traits using reference transcriptome data. *Nature genetics*, 47:1091–1098, 2015.

43. A. Gusev et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nature genetics*, 48:245–252, 2016.

44. D.W. Yao et al. Quantifying genetic effects on disease mediated by assayed gene expression levels. *Nature Genetics*, 52:626–633, 2020.

45. G. Kichaev et al. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS genetics*, 10(10):e1004722, 2014.

46. W. Chen et al. Incorporating functional annotations for fine-mapping causal variants in a Bayesian framework using summary statistics. *Genetics*, 204(3):933–958, 2016.

47. G. Kichaev et al. Improved methods for multi-trait fine mapping of pleiotropic risk loci. *Bioinformatics*, 33(2):248–255, 2017.

48. J.P. Ray et al. Prioritizing disease and trait causal variants at the TNFAIP3 locus using functional and genomic features. *Nature communications*, 11(1):1–13, 2020.

49. Y. Hu et al. Leveraging functional annotations in genetic risk prediction for human complex diseases. *PLoS computational biology*, 13(6):e1005589, 2017.

50. C. Márquez-Luna et al. Multiethnic polygenic risk scores improve risk prediction in diverse populations. *Genetic epidemiology*, 41(8):811–823, 2017.

51. K. Jaganathan et al. Predicting splicing from primary sequence with deep learning. *Cell*, 176(3):535–548, 2019.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

52. L. Liu et al. Biological relevance of computationally predicted pathogenicity of noncoding variants. *Nature Communications*, 10:330, 2019.

53. D.R. Kelley. Cross-species regulatory sequence activity prediction. *PLOS Computational Biology*, 16(7):e1008050, 2020.

54. M. Lizio et al. Gateways to the fantom5 promoter level mammalian expression atlas. *Genome biology*, 16(1):22, 2015.

55. M. Lizio et al. Update of the fantom web resource: high resolution transcriptome of diverse cell types in mammals. *Nucleic acids research*, 45(D1):D737, 2017.

56. J. Ernst and M. Kellis. ChromHMM: automating chromatin-state discovery and characterization. *Nature methods*, 9:215–216, 2012.

57. J.H. Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

58. B. Caron, Y. Luo, and A. Rausell. NCBoost classifies pathogenic non-coding variants in Mendelian diseases through supervised learning on purifying selection signals in humans. *Genome biology*, 20:32, 2019.

59. W.J. Kent et al. The human genome browser at ucsc. *Genome research*, 12(6):996–1006, 2002.

60. D. Karolchik, A.S. Hinrichs, T.S. Furey, K.M. Roskin, C.W. Sugnet, et al. The UCSC Table Browser data retrieval tool. *Nucleic acids research*, 32:D493–D496, 2004.

61. H. Tong, C. Faloutsos, and J.Y. Pan. Random walk with restart: fast solutions and applications. *Knowledge and Information Systems*, 14:327–346, 2008.

62. D. Szklarczyk et al. STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic acids research*, 43:D447–D452, 2014.

63. D. Szklarczyk et al. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic acids research*, 45(Database issue):D362–D368, 2017.

64. X. Wang and D.B. Goldstein. Enhancer Domains Predict Gene Pathogenicity and Inform Gene Discovery in Complex Disease. *The American Journal of Human Genetics*, 106:215–233, 2020.

65. H.K. Finucane, Y.A. Reshef, V. Anttila, K. Slowikowski, A. Gusev, A. Byrnes, et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nature genetics*, 50:621–629, 2018.

66. A.P. Schoech et al. Quantification of frequency-dependent genetic architectures in 25 UK Biobank traits reveals action of negative selection. *Nature communications*, 10(1):1–10, 2019.

67. H.M. Amemiya, A. Kundaje, and A.P. Boyle. The ENCODE blacklist: identification of problematic regions of the genome. *Scientific reports*, 9(1):1–5, 2019.

68. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489:57–74, 2012.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.08.288563; this version posted September 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

69. Jan-Renier AJ Moonen et al. KLF4 Recruits SWI/SNF to Increase Chromatin Accessibility and Reprogram the Endothelial Enhancer Landscape under Laminar Shear Stress. *bioRxiv*, 2020.

70. S. McCarthy, S. Das, W. Kretzschmar, O. Delaneau, A.R. Wood, A. Teumer, H.M. Kang, C. Fuchsberger, P. Danecek, K. Sharp, and Y. Luo. A reference panel of 64,976 haplotypes for genotype imputation. Nature genetics. *Nature genetics*, 48:1279–1283, 2016.

71. L. Jostins et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*, 491:119–124, 2012.

72. Y. Okada et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*, 506:376–381, 2014.

73. J. Bentham et al. Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nature genetics*, 47(12):1457–1464, 2015.

74. P.C. Dubois et al. Multiple common variants for celiac disease influencing immune gene expression. *Nature genetics*, 42(4):295–302, 2010.

75. S.S. Kim et al. Improving the informativeness of Mendelian disease pathogenicity scores for common disease. *bioRxiv*, 2020.

76. S.S. Kim et al. Genes with high network connectivity are enriched for disease heritability. *The American Journal of Human Genetics*, 104:896–913, 2019.