Improving the trans-ancestry portability of polygenic risk scores by prioritizing variants in predicted cell-type-specific regulatory elements

Supplement

Significant IMPACT annotation-trait associations

We identified at least one statistically significant IMPACT annotation association with 95 of 111 polygenic traits. These 95 account for 60 of 69 European phenotypes and 35 of 42 East Asian phenotypes. Analogously, across 707 cell type regulatory annotations, we identified at least one significant annotation-trait association for 566 annotations at 5% FDR. For all trait-annotation pairs, the computed τ^* and enrichment estimates, along with their standard errors can be found in **ST4-8**.

Annotations and traits with no observed heritability enrichment

For 16 polygenic traits, we observed no statistically significant annotation association. Of these 16 polygenic traits, 9 were from European GWAS; these are anorexia, cataract, "ever smoked", three pigmentation phenotypes (skin, sunburn, tanning), and three heart disease phenotypes (CHF, IS, AF). The remaining 7 traits with no annotation associations from East Asian GWAS were cataract, COPD, IS, keloid, osteoporosis, pancreatic cancer, and pollinosis. Likewise, for 141 IMPACT annotations, we observed no statistically significant trait association. These annotations included melanoma and heart-labeled annotations (**SF29**). Just over 40% of sarcoma annotations were significantly associated with at least one trait; for all other tissue types, more than 60% of the corresponding annotations were significantly associated with at least one trait. We found that number of training ChIP-seq peaks were significantly correlated with both the size of annotation and the AUPRC of the TF binding model (Pearson r = 0.22, P < 1.5e-9; Pearson r = 0.39, P < 1.5e-24, respectively) (**SF29**). However, the AUPRC and size of annotation are significantly negatively correlated (Pearson r = -0.25, P < 4.8e-11). This perhaps indicates that models with a small number of training peaks and above-average AUPRC (overfitting) will lead to smaller annotations which don't adequately cover the polygenic space, leading to fewer significant heritability enrichments. Moreover, we found that these unassociated annotations have generally significantly smaller annotation sizes (P < 7.0e-10), significantly higher TF binding model AUPRCs (P < 3.2e-18), significantly less training data (P < 0.03), and are biased for particular cell types (**SF29**).

Deep learning comparison across 69 EUR traits

As we performed a more thorough comparison of heritability captured by IMPACT compared to deep learning annotations among the five representative traits by collecting 123 relevant annotations, such an analysis was challenging to perform across all 69 EUR traits. As Basenji and DeepSEA annotations from a previous study[1] accounted for the lead annotation among the five representative traits, we applied these 32 annotations to partition the heritability of the remaining 64 EUR traits. We found that IMPACT annotations captured more heritability (49.5%, sem = 3.3%) than both lead Basenji deep learning annotations (31.9%, sem = 1.9%, one-tailed paired wilcoxon P < 2.0e-11) (**SF8, ST13**) and lead DeepSEA deep learning annotations (27.5%, sem = 1.2%, one-tailed paired wilcoxon P < 1.4e-10) (**SF8, ST13**). Moreover, the τ * of lead IMPACT annotations was almost always greater than that reported for Basenji annotations (by a factor of 2.24x, one-tailed paired wilcoxon P < 3.4e-11) and for DeepSEA annotations (by a factor of 3.55x, one-tailed paired wilcoxon P < 8.8e-12, **SF8, ST13**).

Regulatory concordance of complex traits

Not only did we observe shared regulatory biology between populations, but also among traits. Despite weak genetic correlation among different traits, we observed strong correlations of IMPACT annotation τ^* among traits, revealing large regulatory modules of immunity, white blood cell regulation, red blood cell (RBC) regulation, and body height (**SF30**). These results suggest that while causal effects and variants may differ among biologically related traits, the regulatory elements in which these variants reside may be shared. Moreover, while genetic correlation approaches consider all genetic signals genome-wide which comprise true biological signal and artefact, we believe that IMPACT is more likely to identify true biological effects, which are shared between related traits, unlike artifactual signals.

Conditional S-LDSC analysis to identify independent annotation-trait associations

Before performing serial conditional analyses, for 9 polygenic traits, we observed a single associated cell type: EUR autism (breast), EAS breast cancer (breast), EAS cervical cancer (stem cell), EAS congestive heart failure (colon), EAS diastolic blood pressure (mesendoderm), EAS gastric cancer (stomach), EAS glaucoma (adipocytes), EAS systolic blood pressure (mesendoderm), EAS uterine fibroids (hematopoietic progenitors). However, for 86 traits, we observed that regulatory elements of multiple IMPACT annotations, mostly implicating diverse cell types, significantly capture heritability (**SF9**). After performing serial conditional analyses to resolve dependent and independent associations, there remained a total of 142 independent cell type-trait associations (**SF9**): 1 trait with 4 associations, 7 traits with 3, 30 traits with 2, 57 traits with 1, and 16 traits with none. Four annotations independently explained significant proportions of heritability in EUR prostate cancer: prostate (TFAP4), prostate (RUNX2), mesendoderm (PDX1), and cervix (NFYB). For seven European traits, three IMPACT annotations independently captured polygenic heritability: height (adipocytes, fibroblasts, lung), neutrophil count (monocytes, adipocytes, B cells), osteoporosis (myoblasts, mesenchymal stem cells, cervix), IBD (T cells and two B cell annotations), platelet count (PBMCs, hematopoietic progenitors, muscle), systolic blood pressure (endothelial, mesenchymal stem cells, fibroblasts), and white blood cell count (B cells, adipocytes, hematopoietic progenitors). For each of 22 European traits and 8 East Asian traits, we observed exactly two independent IMPACT annotation associations. Finally, for each of 30 European traits and 27 East Asian traits, we observed exactly one independent IMPACT association. For Crohn's (EUR), Th1s and naive CD4+ T cells independently captured heritability, suggesting two different biological mechanisms one via naive T cells and the other via memory effector cells. Although previous studies suggested an important role of T cells in UC[2], our study identified not only T cells but also B cells as contributors to disease pathogenesis. For UC (EUR), T cells and B cells contribute independently to explain heritability. In summary, we have elucidated the biology of some polygenic traits through resolving not only the most significantly associated cell type, but also secondary, tertiary, and quaternary independent mechanisms. These results also shed light on shared regulatory programs between cell types: in cases where prior to conditioning, we observed many diverse cell type associations, yet upon conditioning revealed a single independent signal. For example, in EUR RA, B cells were most strongly associated, while CD4+ memory T cell annotations also captured significant proportions of heritability. However, these T cell

annotations were not associated independently of B cells, suggesting that RA heritability resides in shared regulatory elements between T and B cells. In summary, we have elucidated the biology of some polygenic traits through resolving not only the most significantly associated cell type, but also secondary, tertiary, and quaternary independent mechanisms.

To investigate the concordance of independent IMPACT signals across related traits, we considered clusters of functionally correlated traits from **SF30**. Among the autoimmune disease and hematological trait cluster, encompassing eosinophil count, asthma, RA, and lymphocyte count, the CD4 T cell:BCL6 and Th1:TBX21 annotations were each three times listed as independent contributors. For the greater hematological trait cluster consisting of monocyte, neutrophil, white blood cell, basophil, platelet, lymphocyte, red blood cell counts as well as MCV, MCH, and MCHC, the PBMC:GATA1 annotation was eight times listed as an independent contributor. Lastly, for the endocrine cluster consisting of BMI, T2D, SBP, Hb, and Ht, the mesendoderm:PDX1 annotation was six times listed as an independent contributor. These observations reveal that there is indeed some degree of persistence of independent genetic contributors and may add a biological basis for the observed genetic correlations among these traits.

We note that our cell type interpretations above rely on the fidelity of the IMPACT model to accurately predict TF binding in the desired cell type; a poor model may learn an epigenetic signature that does not represent the desired cell type. The mean TF binding model AUPRC of independently associated IMPACT annotations was significantly less (mean AUPRC = 0.41, se = 0.04) than than of all IMPACT annotations (mean AUPRC = 0.54, se = 0.01, difference of means $P < 8.1e\text{-}4$). This is consistent with our observation that IMPACT annotations with very high AUPRCs are less likely to capture polygenic heritability (**SF29**).

Cell type composite annotations targeting multiple independent mechanisms of polygenic traits

In light of observing 38 phenotypes for which multiple cell type regulatory element annotations independently captured significant proportions of heritability, we created composite cell type annotations in hopes of improving heritability enrichments. For example, we observed that genetic variation governing neutrophil count (EUR) is independently accounted for by monocytes, adipocytes, and B cell regulatory elements. Then, we annotated SNPs genome-wide using a probabilistic OR gate as follows:

$$score_j = 1 - \prod_i^a (1 - IMPACT_{i,j}),$$

where *j* is the SNP index, *i* is the i^{th} annotation, *a* is the number of independently associated annotations for the trait of interest and $IMPACT_{i,j}$ is the IMPACT score of variant *j* in annotation *i*.

We created 38 composite cell type annotations and observed that these annotations captured significantly more overall enrichment (one-tailed paired wilcoxon $P < 4.9e\text{-}10$), significantly more per-SNP heritability in terms of τ^* (one-tailed paired wilcoxon $P < 3.2e\text{-}8$), and significantly more heritability in the top 5% of SNPs (one-tailed paired wilcoxon $P < 0.004$) (**SF31**).

Trends of multi-ethnic marginal effect size correlation at various *P* value thresholds

We observed that at lenient *P* value thresholds, the difference in correlation between EUR and EAS effect sizes is more pronounced using IMPACT annotations, suggesting that they may be more effective for prioritizing causal variation particularly when statistical evidence is weak. For example, at the most lenient *P* value thresholds between $P < 1$ and $P > 3e\text{-}4$, we observed more dramatic improvements in correlation using IMPACT while on the other hand, at more stringent *P* value thresholds, IMPACT annotations offer less of an improvement in multi-ethnic effect size correlation (**SF32**).

Robustness of PRS analysis to scale on which effect sizes are estimated

For case/control diseases, we estimated marginal effect sizes on the logistic scale. To ensure that our results were consistent if effect sizes were to be estimated on the liability scale, for each of 5 case/control diseases considered in PRS analyses, we converted effect sizes from logistic scale to liability scale (**Online Methods**). The conversion had negligible effects on our findings: 1) effect size estimates were nearly perfectly correlated (**SF25**), 2) PRS values were also nearly perfectly correlated (**SF26**), and 3) the predictive power of PRS models were highly consistent (for EUR PRS resulting in an average change in magnitude of pseudo-R^2 equivalent to 1.8e-5 or a 0.16% average increase in pseudo-R^2 values relative to logistic-based PRS; and for EAS PRS resulting in an average change in magnitude of pseudo-R^2 equivalent to 1.3e-4 or a 0.81% average increase in pseudo-R^2 values relative to logistic-based PRS, **SF26**). These results demonstrate that the way in which effect sizes are defined has negligible effects on our findings.

Supplementary Figures

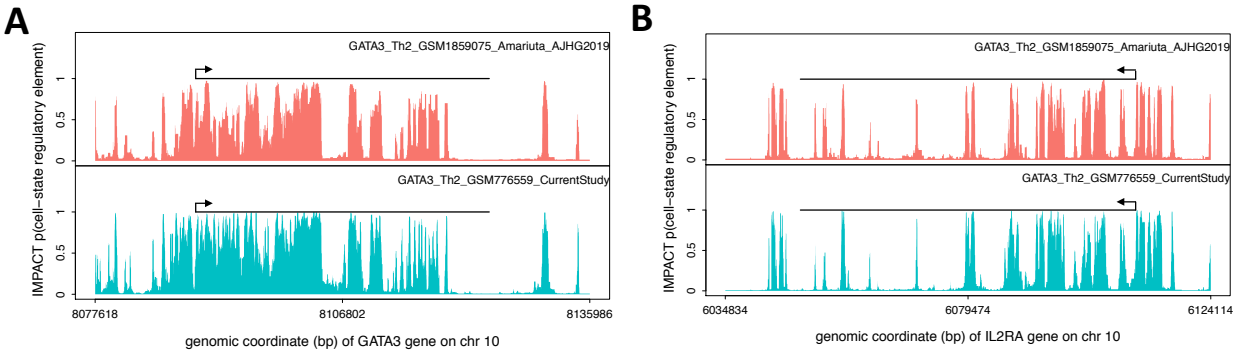


Figure S1 legend. Consistency of IMPACT predictions for the same TF/cell type pair (GATA2/Th2) using different experiments and different feature sets: GSM1859075 used in Amariuta et al AJHG 2019 with 515 epigenetic features and GSM776559 used in the current study with 5,345 total epigenetic features. A) GATA3 gene locus on chr10. B) IL2RA gene locus on chr10.

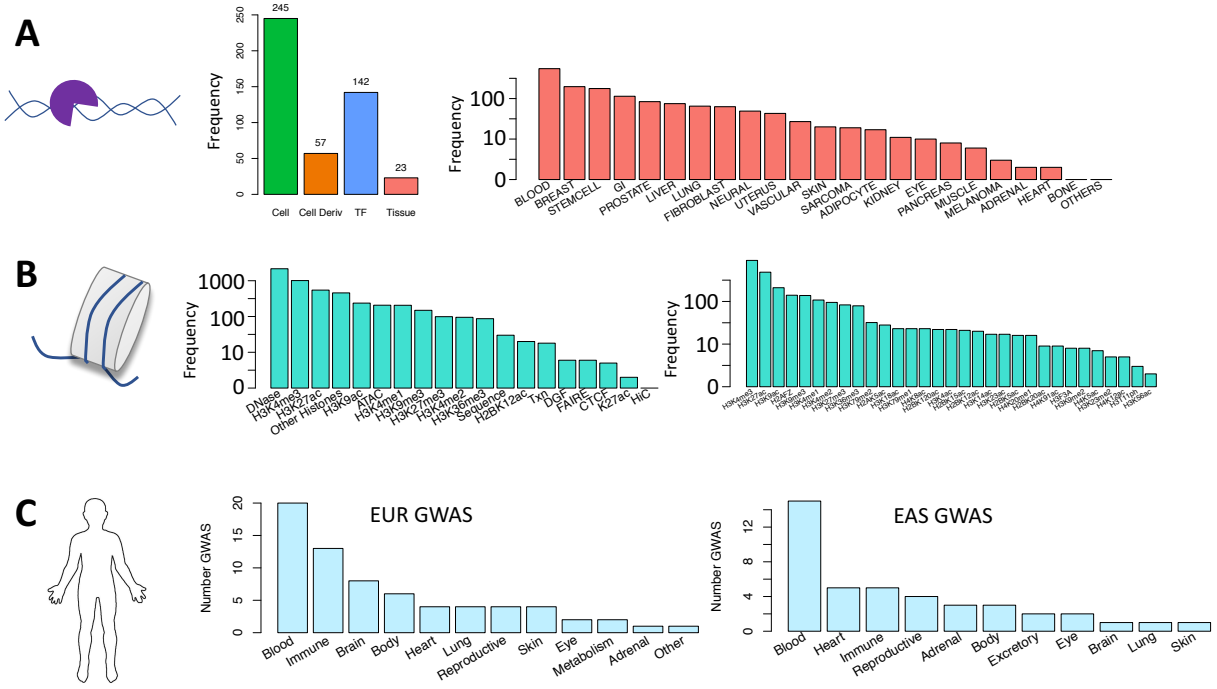
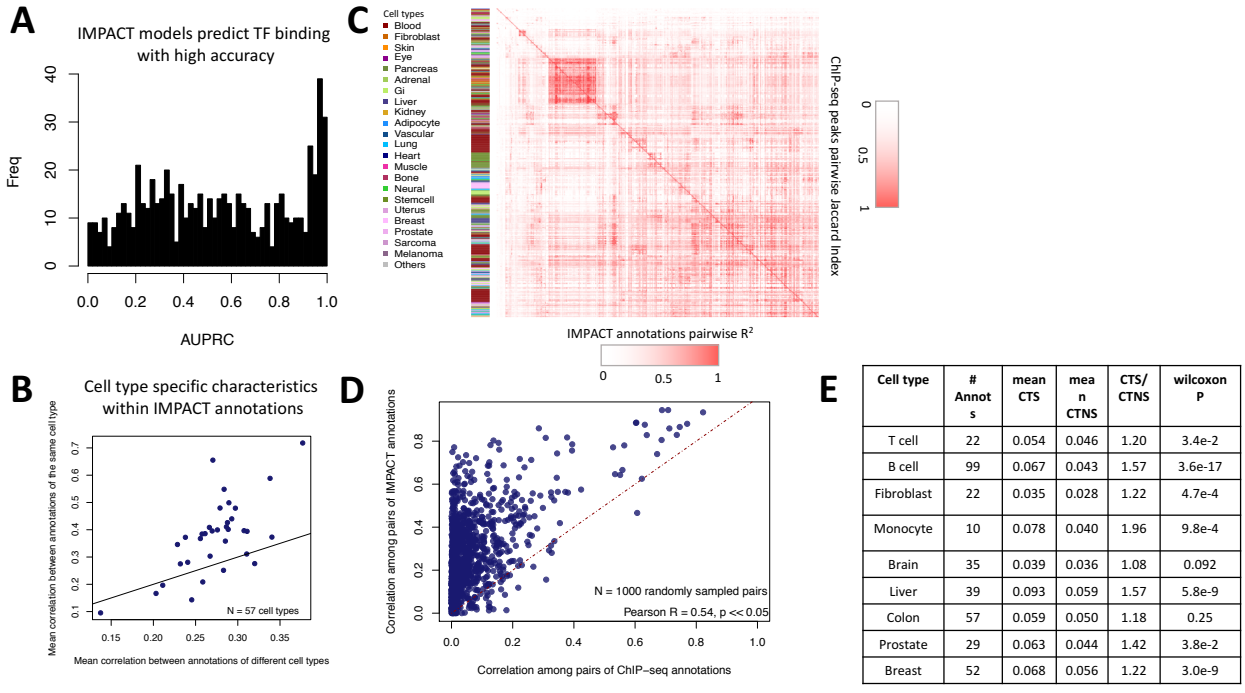


Figure S2 legend. A) TF ChIP-seq collection from NCBI: (left) cell type and TF diversity where "Cell Deriv" indicates number of unique parental cell types, for example GM12878 and GM10847 are both B cell lines, (right) diversity of tissue types. B) (left) Epigenomic and sequence features to be used in IMPACT models, (right) diversity of histone modification ChIP-seq in features. C) Diversity of European (EUR) and East Asian (EAS) GWAS summary statistics across phenotypic categories.



E

Cell type	# Annots	mean CTS	mean CTNS	CTS/ CTNS	wilcoxon P
T cell	22	0.054	0.046	1.20	3.4e-2
B cell	99	0.067	0.043	1.57	3.6e-17
Fibroblast	22	0.035	0.028	1.22	4.7e-4
Monocyte	10	0.078	0.040	1.96	9.8e-4
Brain	35	0.039	0.036	1.08	0.092
Liver	39	0.093	0.059	1.57	5.8e-9
Colon	57	0.059	0.050	1.18	0.25
Prostate	29	0.063	0.044	1.42	3.8e-2
Breast	52	0.068	0.056	1.22	3.0e-9

Figure S3 legend. A) Histogram of prediction performance of 707 IMPACT models (metric = AUPRC). B) IMPACT annotations of the same cell type are more similar to one another than annotations of different cell types. C) Pairwise correlation of IMPACT regulatory element annotations (lower triangle of matrix) relative to pairwise correlation of corresponding TF ChIP-seq annotations (upper triangle of matrix). Pearson r was calculated using probabilities assigned to 779,355 SNPs on chr1 from phase 3 of 1000G (EUR), Jaccard indices were calculated for binary ChIP-seq tracks genome-wide, in which the size of the intersection of base pairs between two datasets was divided by the size of the union of base pairs. D) Pairwise correlations between 1000 randomly selected datasets between TF ChIP-seq and their corresponding IMPACT annotations; values sampled from C). E) IMPACT assigns larger cell-type-specific regulatory elements probabilities at cell-type-specifically expressed genes across nine cell types.

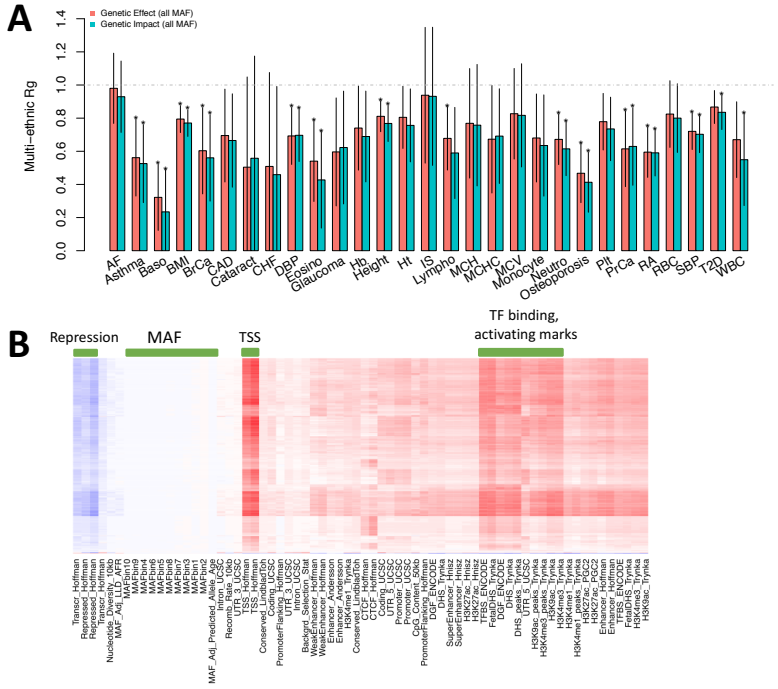


Figure S4 legend. A) For the 29 traits for which we collected both EUR and EAS GWAS summary statistics, we computed the multi-ethnic genetic correlation with Popcorn. According to genetic effect, for 13 traits, the genetic correlation is significantly less than 1, indicated by an asterisk ($P < 0.05 / 29$ traits). We plot both the genetic correlation computed separately using genetic effect (effect size estimates unnormalized to allele frequency) and genetic impact (allele variance normalized effect sizes). B) IMPACT annotations correlate most with TSS, TFBS, and activation histone mark annotations, while no correlation is present with European ancestry MAF bins.

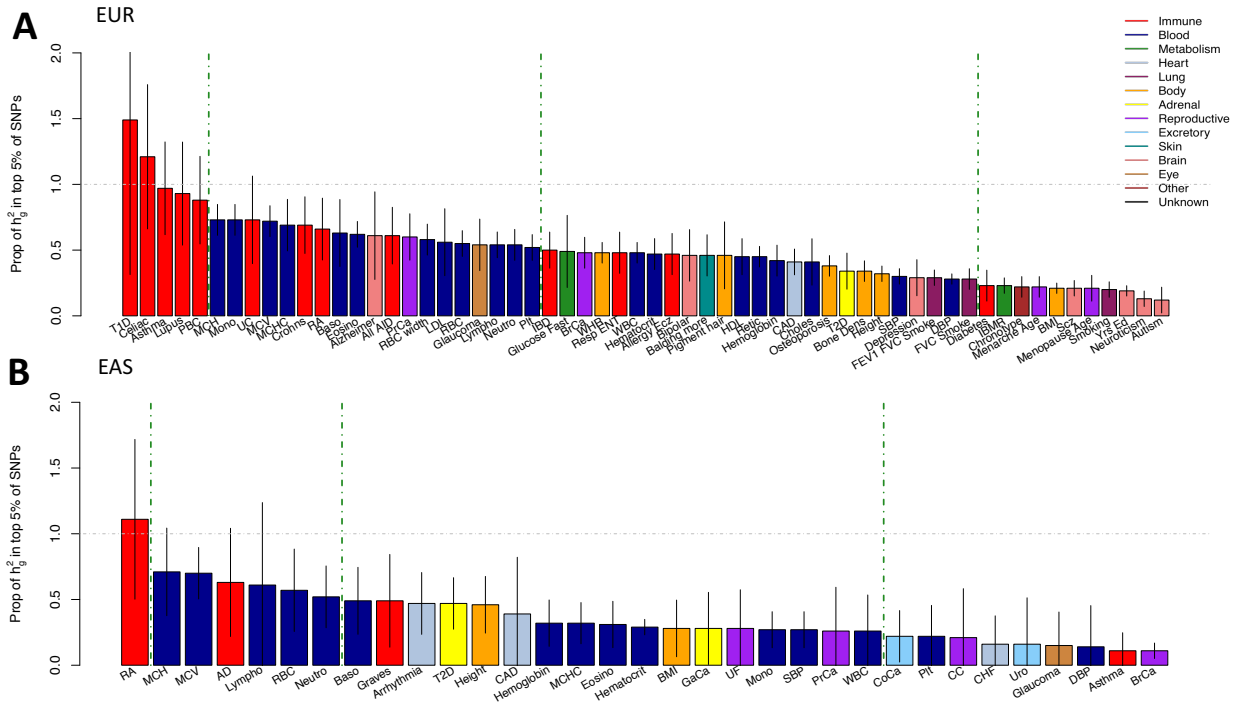


Figure S5 legend. A) Common SNP heritability captured by the top 5% of SNPs according to the lead cell type association for each EUR GWAS. Lead association determined by largest τ^* estimate that is significantly positive. B) Similar for each EAS GWAS. Gray bars indicate the standard error of the heritability estimate. Color represents the category of the complex trait or disease.

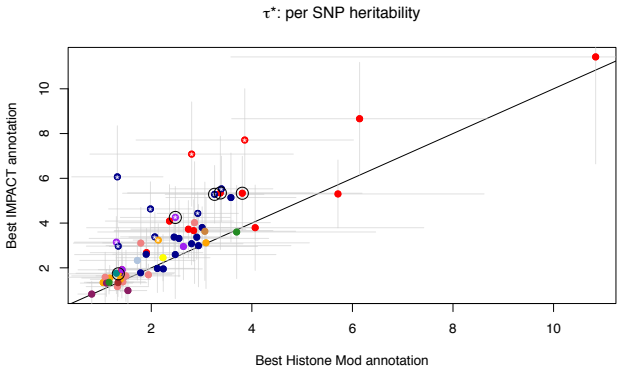


Figure S6 legend. Comparison of two different functional annotations, IMPACT and cell type specific histone marks, to capture polygenic heritability assessed by quantifying τ * per-SNP heritability value. Circled are five representative traits used throughout the study: asthma, RA, PrCa, MCV, and height.

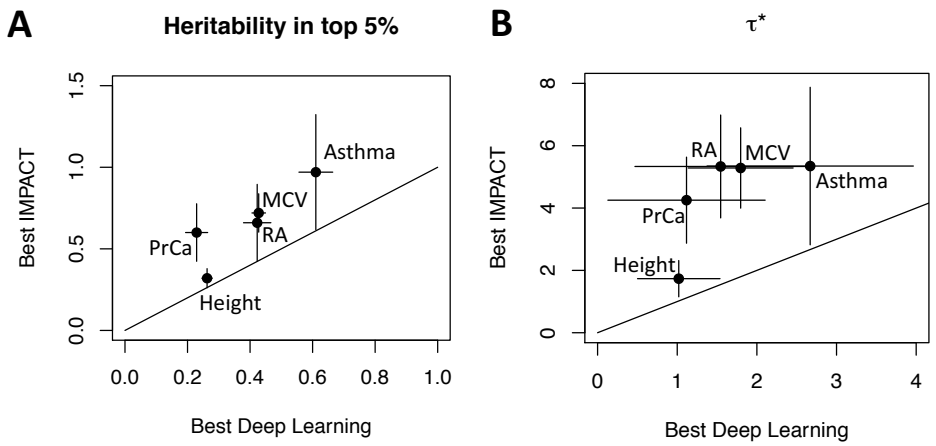


Figure S7 legend. A) Among five representative traits, proportion of total SNP heritability captured by the lead IMPACT annotation compared to the lead deep learning annotation, from a set of 123 annotations. B) Among five representative traits, τ^* of the lead IMPACT annotation compared to the lead deep learning annotation, from a set of 123 annotations.

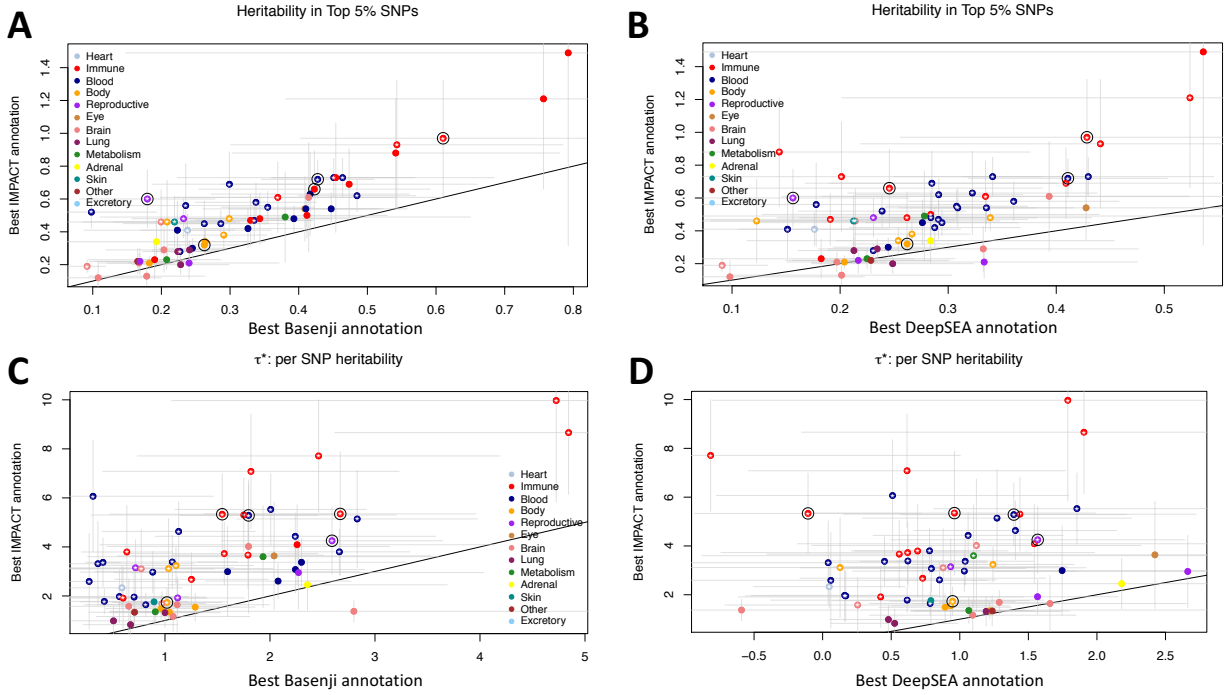


Figure S8 legend. Proportion of total SNP heritability captured by top 5% of SNPs according to lead IMPACT annotation (y axis) and lead Basenji annotation (x axis) in panel A or lead DeepSEA annotation in panel B. Standardized annotation effect size τ^* according to lead IMPACT annotation (y axis) and lead Basenji annotation (x axis) in panel C or lead DeepSEA annotation in panel D.

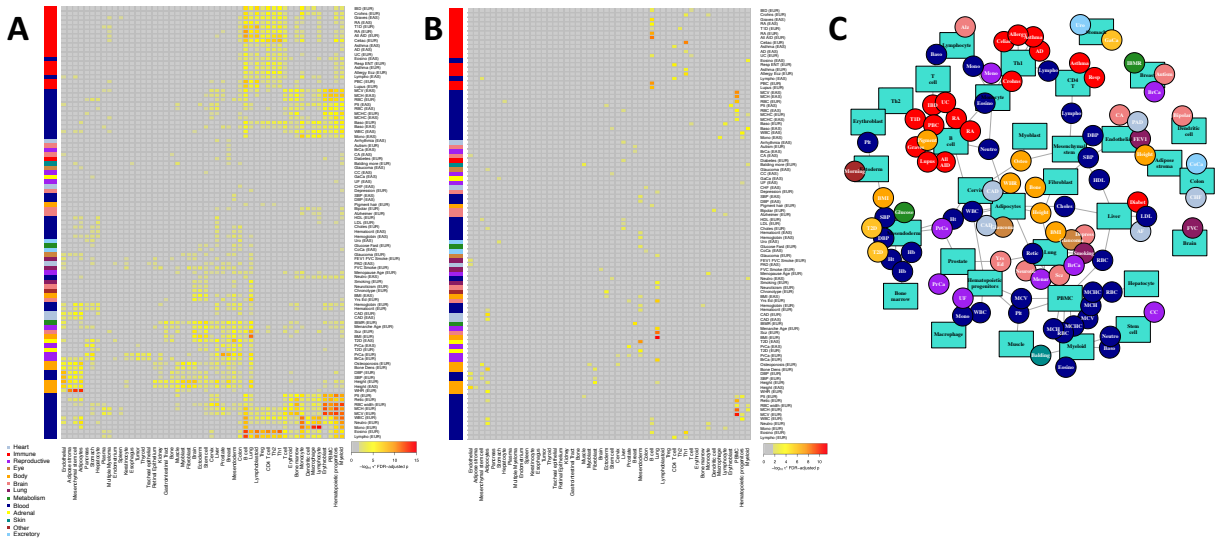


Figure S9 legend. A) Stratification of IMPACT annotation associations by 50 cell types across the 95 polygenic traits and diseases of 111 with at least one association. For each cell type, the strongest annotation association is represented ($-log_{10}\,\tau^*$ *P* value, FDR 5% adjusted). B) After four rounds of conditional analysis, non-independent associations were removed. Shown are the remaining independent annotation associations of the same 50 cell types and 95 traits. Color indicates $-log_{10}\,\tau^*$ *P* value adjusted for FDR 5%; if more than one independent cell type association, $-log_{10}\,\tau^*$ conditional *P* value adjusted for FDR 5% is indicated. C) Network of remaining independent associations, same information as in B), reveals clusters of regulatory modules that recapitulate known biology.

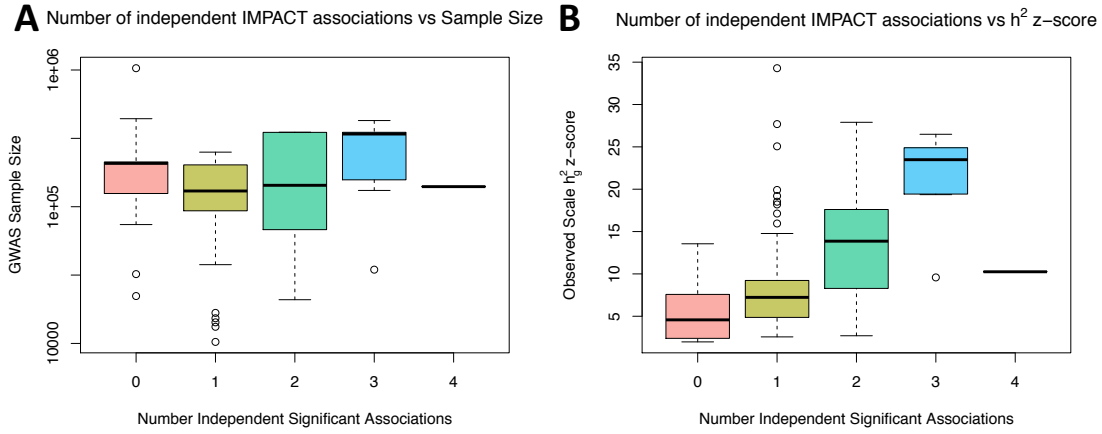


Figure S10 legend. A) Number of independent IMPACT cell type associations is not significantly correlated with the sample size of the GWAS (*P* = 0.19). B) Number of independent associations is significantly positively correlated with the observed scale heritability z-score of the trait (*P* < 5.4e-9).

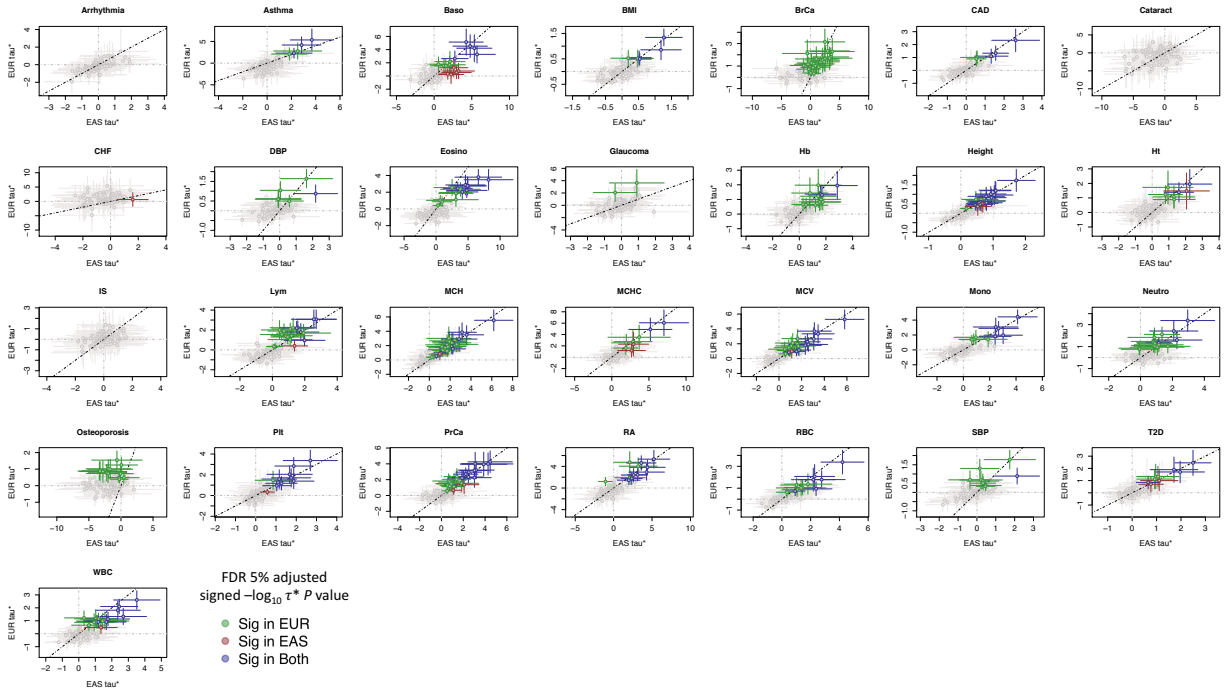


Figure S11 legend. Common per-SNP heritability (τ*) estimate for sets of independent IMPACT cell type annotations across 29 traits. Dotted line is the identity line, y = x. τ* values with their standard errors are colored green if significantly positive in EUR and not EAS, red if significantly positive in EAS but not in EUR, green if significantly positive in both EUR and EAS, and gray if not significantly positive in either population.

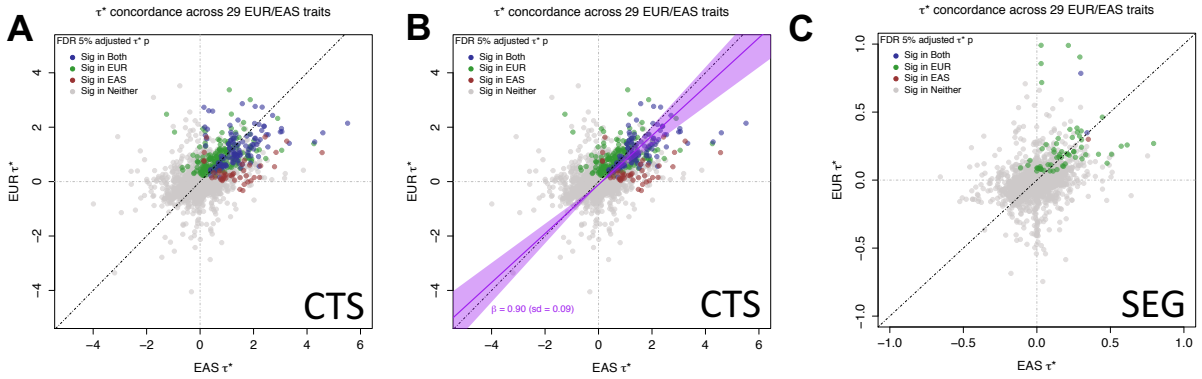


Figure S12 legend. A) Common per-SNP heritability (τ*) estimate for sets of independent cell-type-specific histone mark annotations from Finucane et al Nature Genetics 2015 (EUR annotations) and Kanai et al Nature Genetics 2018 (EAS annotations) across 29 traits. B) As in A) after removing eight outlier annotations from "Sig in Both" category with noticeably larger EUR τ* and small EAS τ*, revealing a cross-ancestry relationship that is not dissimilar from identity. Line of best fit through annotations significant in both populations (dark purple line, 95% CI in light purple). C) As in A) for sets of independent cell-type-specifically expressed gene sets from Finucane et al Nature Genetics 2018 (EUR annotations) and Kanai et al Nature Genetics 2018 (EAS annotations). For all panels, the dotted line is the identity line, y = x. τ* values with their standard errors are colored

green if significantly positive in EUR and not EAS, red if significantly positive in EAS but not in EUR, green if significantly positive in both EUR and EAS, and gray if not significantly positive in either population.

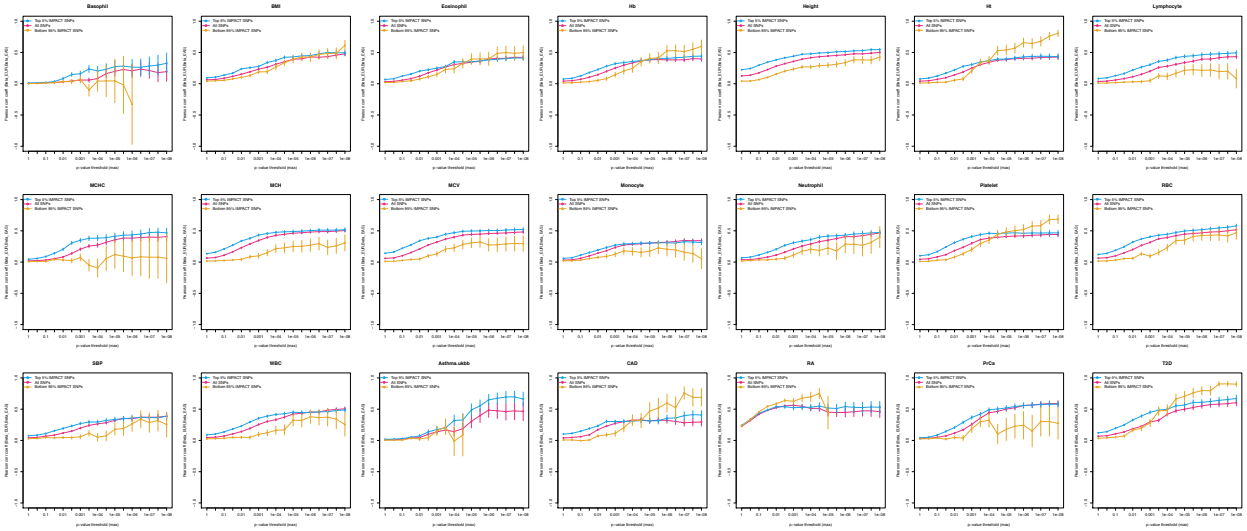


Figure S13 legend. For 21 traits shared between EUR and EAS, effect size correlation (Pearson correlation coefficient) across 17 *P* value thresholds for three partitions of SNPs genome-wide: 1) lead SNPs with no IMPACT inference (red), 2) top 5% of SNPs according to the largest τ* effect size IMPACT annotation (blue), and 3) the bottom 95% of SNPs according to the same IMPACT annotation (yellow). Vertical lines indicate one standard deviation of the correlation coefficient estimate.

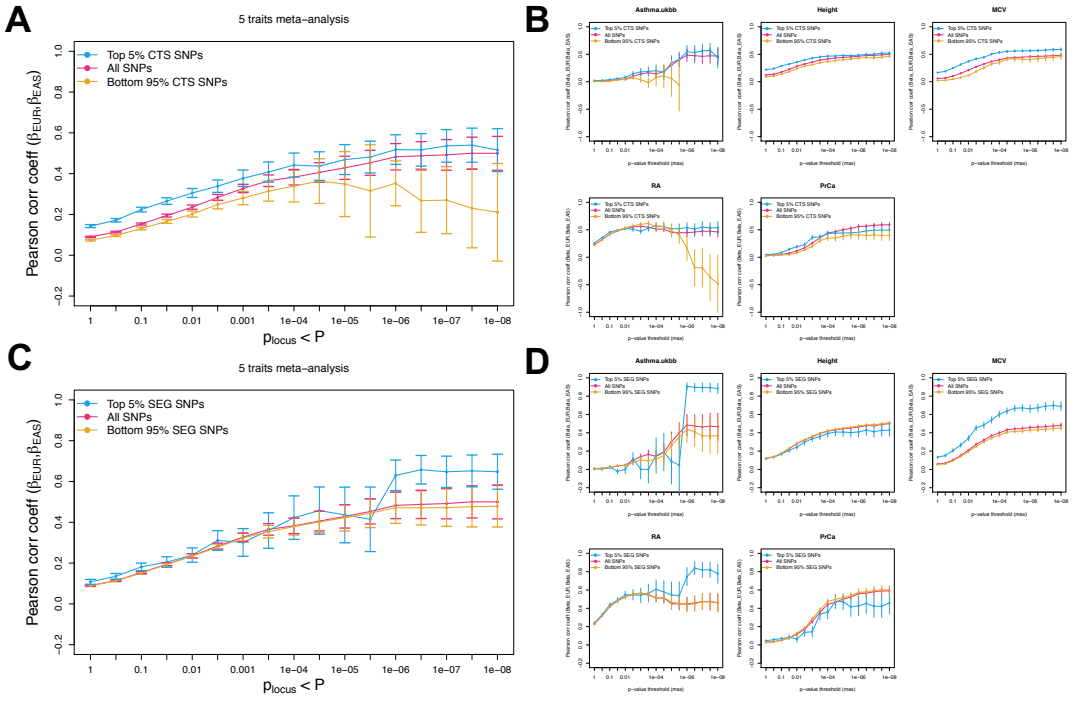


Figure S14 legend. For 5 traits representing different biological underpinnings shared between EUR and EAS (subset of 21 investigated in our study), we report the effect size correlation (Pearson correlation coefficient) across 17 *P* value thresholds for three partitions of SNPs genome-wide: 1) lead SNPs with no functional inference (red), 2) top 5% of SNPs according to the largest τ^* annotation effect size (blue), and 3) the bottom 95% of SNPs according to the same functional annotations (yellow). Here, we select the top annotation in two categories of previously published functional annotations: first, from LDSC-CTS annotations (meta-analysis in A, individual traits in B) and second, from LDSC-SEG annotations (meta-analysis in C, individual traits in D). Vertical lines indicate one standard deviation of the correlation coefficient estimate.

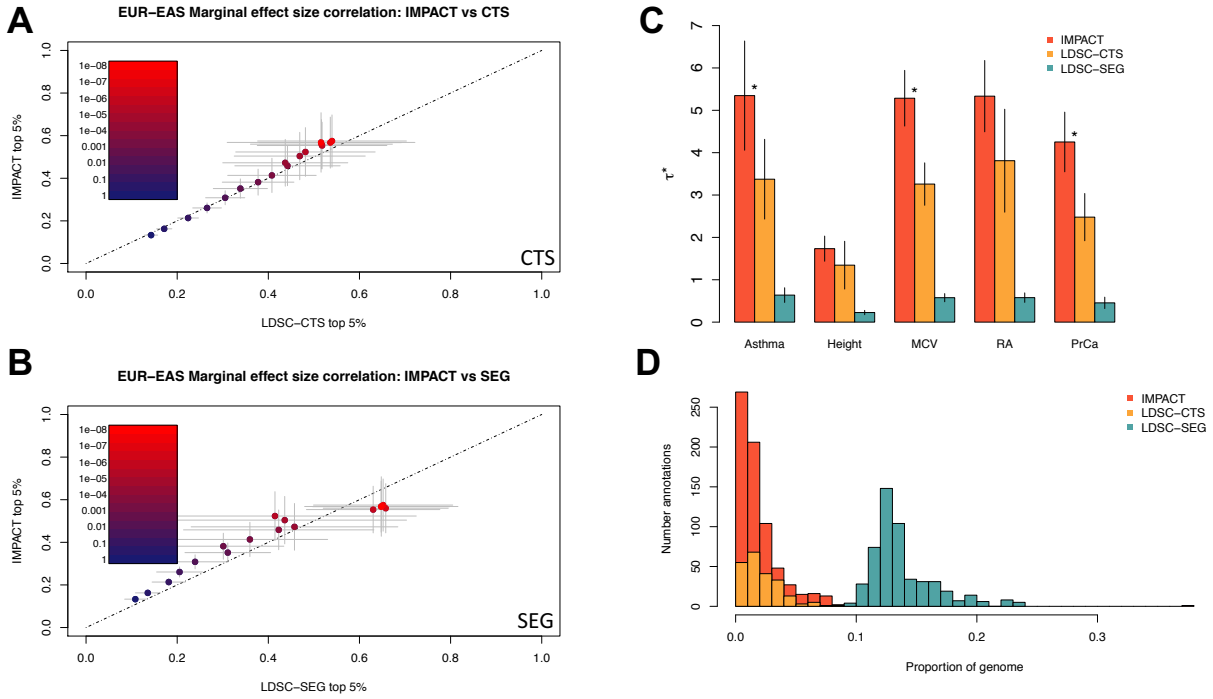


Figure S15 legend. A) Comparison of top LDSC-CTS annotations in multi-ethnic effect size correlation analysis with top IMPACT annotations meta-analyzed over 5 traits. B) Similar to A) but for LDSC-SEG annotations. C) τ^* across the 5 selected traits reveals that IMPACT annotations are more strongly enriched for trait heritability than LDSC-CTS annotations (indicated by asterisk, difference of means $P < 0.05$) and consistently more than LDSC-SEG annotations. D) Distribution of annotation sizes for three different functional regimes: IMPACT (red), LDSC-CTS (yellow), LDSC-SEG (teal).

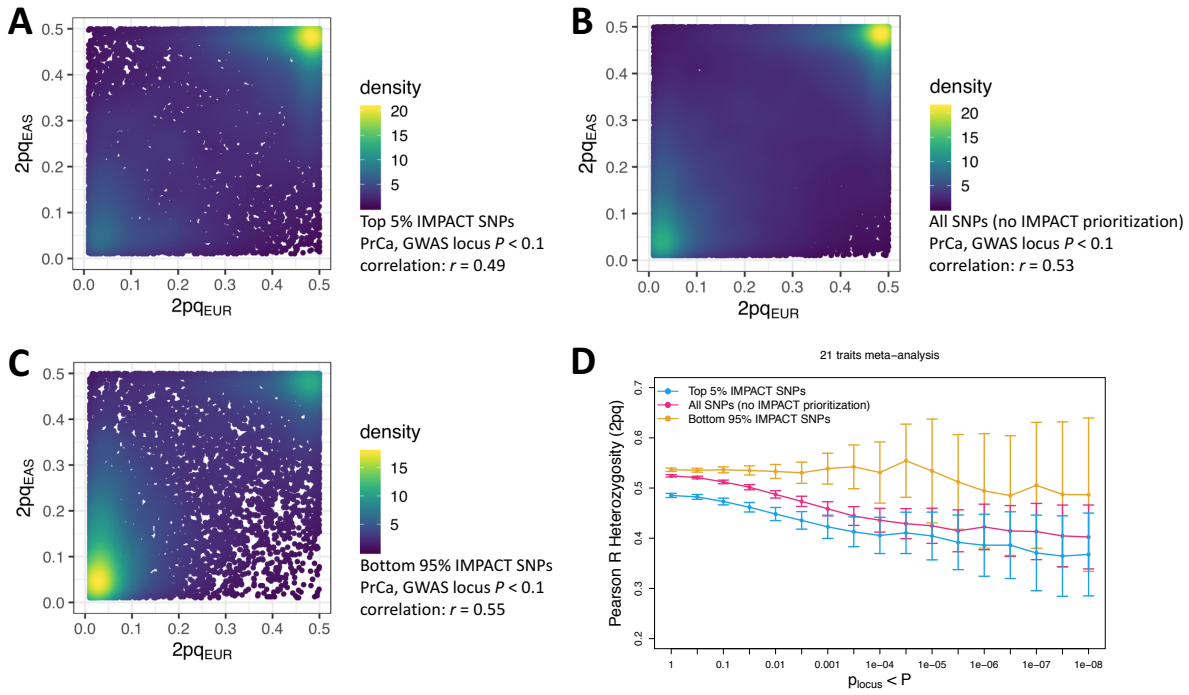


Figure S16 legend. Population concordance of heterozygosity (2pq) among variants prioritized by IMPACT compared to standard P+T. A) Heterozygosity of variants from genome-wide EUR and EAS PrCa summary statistics in the top 5% of the lead IMPACT annotation for EUR PrCa. B) Heterozygosity of variants from genome-wide EUR and EAS PrCa summary statistics using standard P+T. C) Heterozygosity of variants from genome-wide EUR and EAS PrCa summary statistics in the bottom 95% of the lead IMPACT annotation for PrCa; mutually exclusive with SNPs in A). D) Meta-analysis of heterozygosity correlations between populations across 21 traits shared between EUR and EAS cohorts over 17 GWAS *P* value thresholds (with reference to the EUR GWAS).

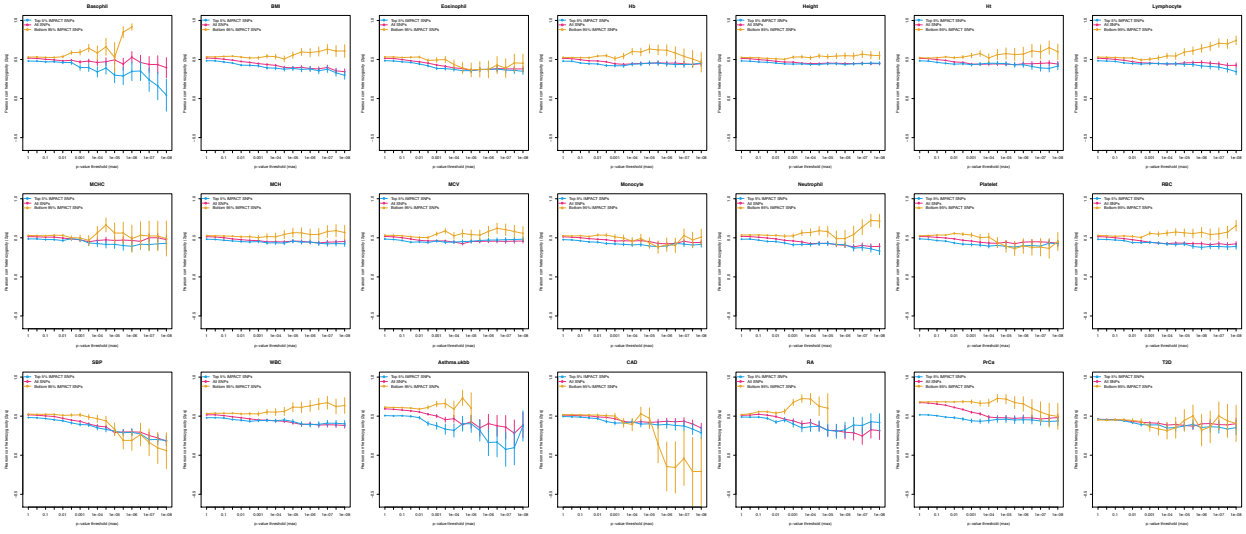


Figure S17 legend. For 21 traits shared between EUR and EAS, heterozygosity (2pq) correlation (Pearson correlation coefficient) across 17 *P* value thresholds for three

partitions of SNPs genome-wide: 1) lead SNPs with no IMPACT inference (red), 2) top 5% of SNPs according to the largest τ^* effect size IMPACT annotation (blue), and 3) the bottom 95% of SNPs according to the same IMPACT annotation (yellow). Vertical lines indicate one standard deviation of the correlation coefficient estimate.

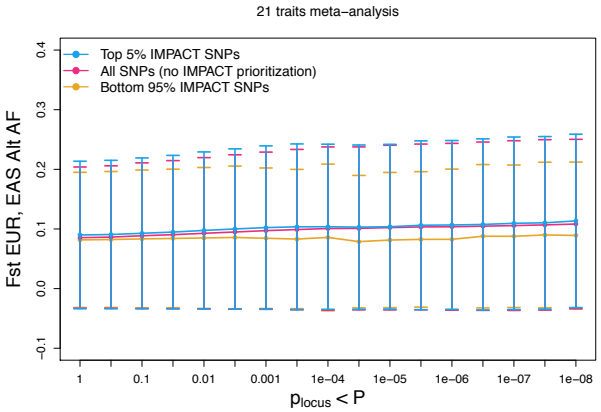


Figure S18 legend. Population divergence, measured by F_{st}, where larger values indicate a reduction in heterozygosity, among variants prioritized by IMPACT compared to standard P+T. Meta-analysis of F_{st} between EUR and EAS populations across 21 traits shared between EUR and EAS cohorts over 17 GWAS *P* value thresholds (with reference to the EUR GWAS).

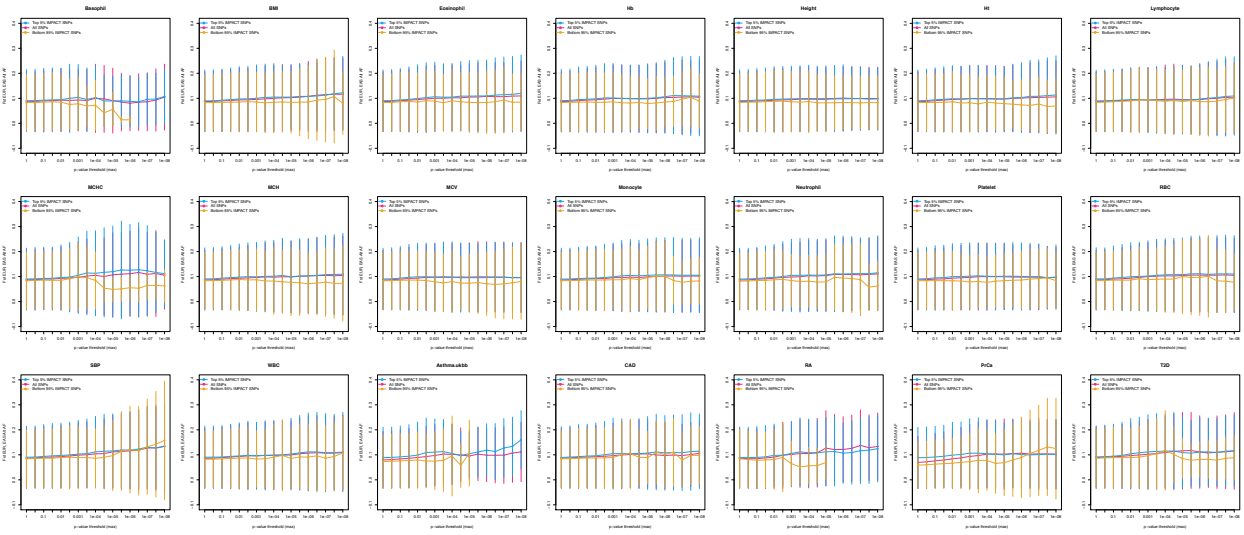


Figure S19 legend. For 21 traits shared between EUR and EAS, we computed the average F_{st}, where large values indicate a reduction in heterozygosity, of sets of variants across 17 *P* value thresholds for three partitions of SNPs genome-wide: 1) lead SNPs with no IMPACT inference (red), 2) top 5% of SNPs according to the largest τ^* effect size IMPACT annotation (blue), and 3) the bottom 95% of SNPs according to the same IMPACT annotation (yellow). Vertical lines indicate one standard deviation of the mean F_{st} estimate.

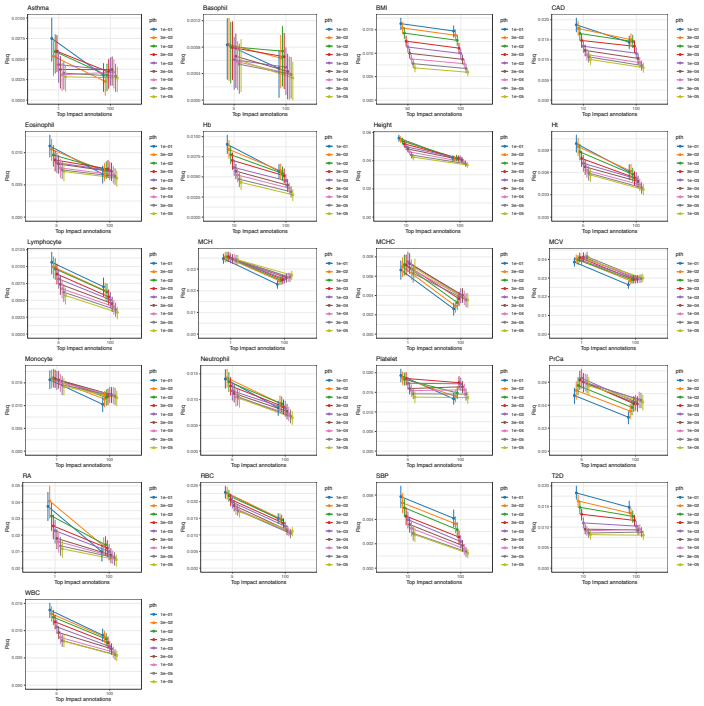


Figure S20 legend. EUR PRS model evaluated on EAS individuals from BBJ. For each trait, we evaluate the predictive value of standard PRS models (top 100% of IMPACT SNPs) and functionally-informed PRS models (using a subset of SNPs prioritized by IMPACT). The top 100% of SNPs according to IMPACT represents the PRS model with no functional annotation information. Intervals represent the 95% confidence interval around the R^2 estimate. For quantitative traits, R^2 represents the proportion of variance captured by the linear PRS model. For case control traits, R^2 represents the liability scale R^2 from the logistic regression PRS model.

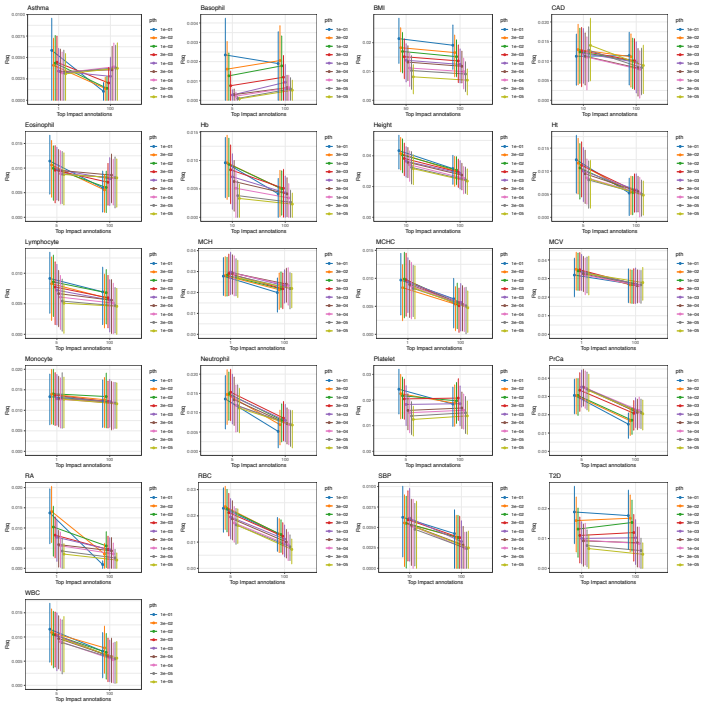


Figure S21 legend. EUR PRS model evaluated on 5,000 randomly selected EAS individuals from BBJ. For each trait, we evaluate the predictive value of standard PRS models (top 100% of IMPACT SNPs) and functionally-informed PRS models (using a subset of SNPs prioritized by IMPACT). Intervals represent the 95% confidence interval around the R^2 estimate. For quantitative traits, R^2 represents the proportion of variance captured by the linear PRS model. For case control traits, R^2 represents the liability scale R^2 from the logistic regression PRS model.

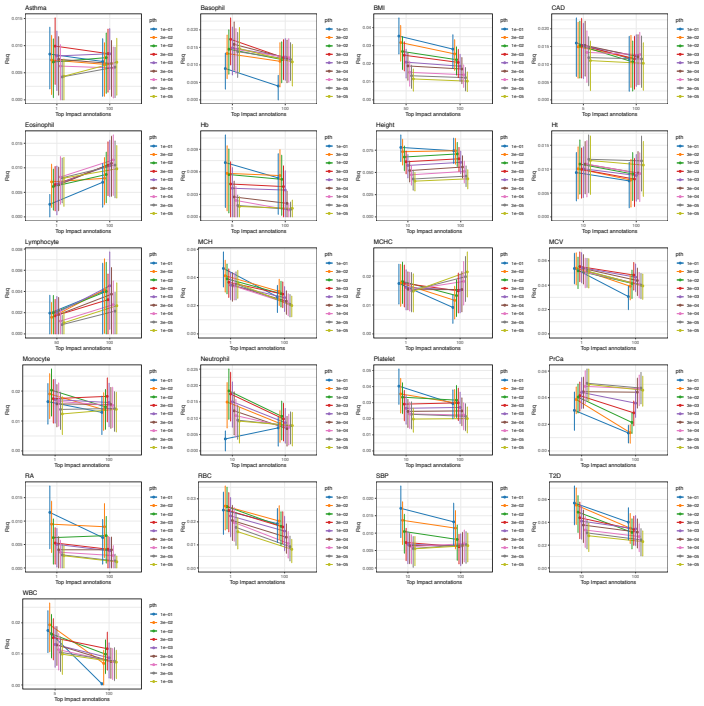


Figure S22 legend. EAS PRS model evaluated on 5,000 non-overlapping EAS individuals from BBJ; these 5,000 individuals are the same as EAS test individuals in SF15. For each trait, we evaluate the predictive value of standard PRS models (top 100% of IMPACT SNPs) and functionally-informed PRS models (using a subset of SNPs prioritized by IMPACT). Intervals represent the 95% confidence interval around the R^2 estimate. For quantitative traits, R^2 represents the proportion of variance captured by the linear PRS model. For case control traits, R^2 represents the liability scale R^2 from the logistic regression PRS model.

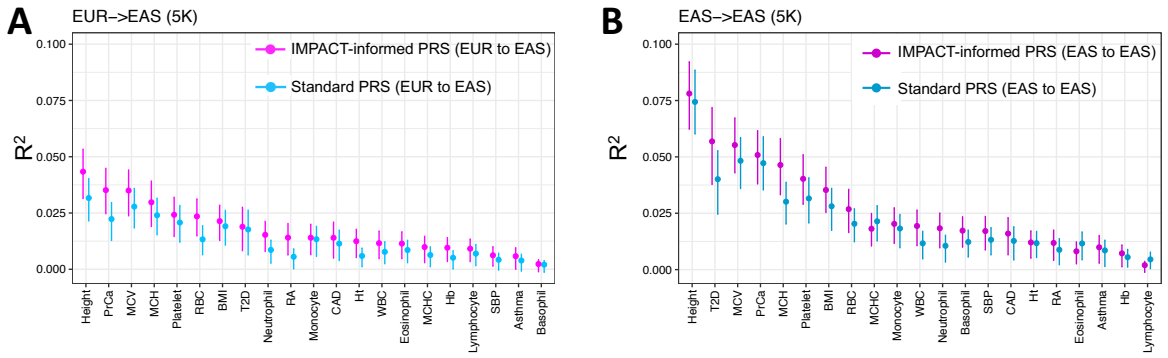


Figure S23 legend. A) Phenotypic variance (R^2) in 5,000 BBJ individuals explained by IMPACT-informed PRS-EUR (dark pink) and standard PRS-EUR (dark blue). B) Phenotypic variance (R^2) in 5,000 BBJ individuals explained by IMPACT-informed PRS-EAS (light pink) and standard PRS-EAS (light blue). Error bars indicate 95% CI calculated via 1,000 bootstraps.

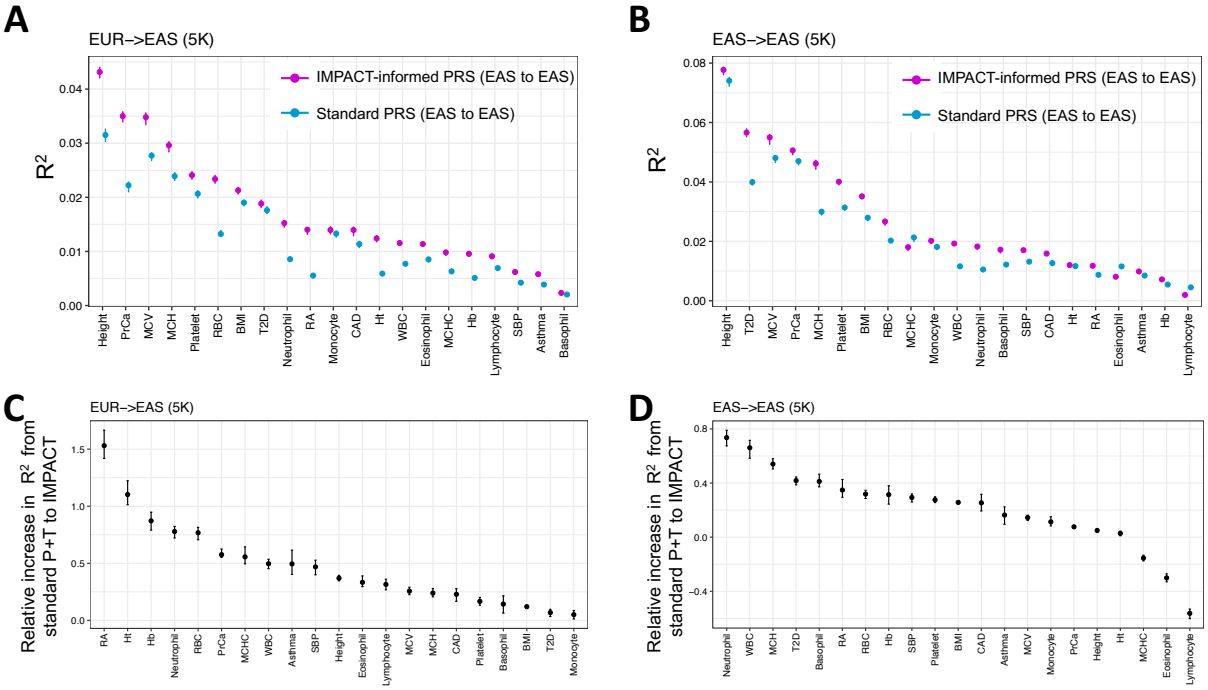


Figure S24 legend. We recomputed confidence intervals around the R^2 estimates (panels A and B) and around the relative improvements in R^2 estimates of IMPACT PRS over standard P+T PRS (panels C and D) via block jackknife across the genome, using 200

adjacent equally-sized bins and iteratively removing variants within each bin and computing the R^2. A) Trans-ethnic analysis of EUR PRS to BBJ individuals. B) Within-population analysis of EAS PRS to BBJ individuals. Error bars indicate 95% CI around the R^2 estimates. C) Trans-ethnic analysis of EUR PRS to BBJ individuals, relative improvement in R^2 estimates defined as (IMPACT R^2 - standard P+T R^2) / standard P+T R^2. D) Within-population analysis of EAS PRS to BBJ individuals, relative improvement in R^2 estimates defined as (IMPACT R^2 - standard P+T R^2) / standard P+T R^2 .

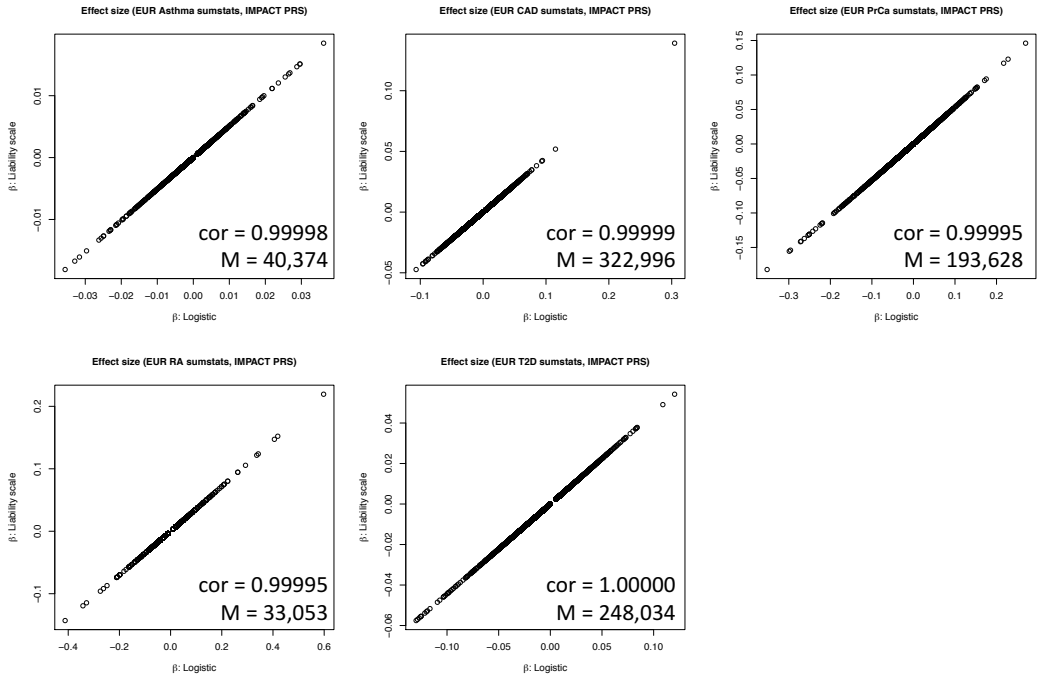


Figure S25 legend. For each of five case/control diseases considered in PRS analyses, we computed the correlation of effect size estimates on the logistic scale versus the liability scale. The set of variants selected for each disease corresponds to the IMPACT-informed PRS model with the highest R^2.

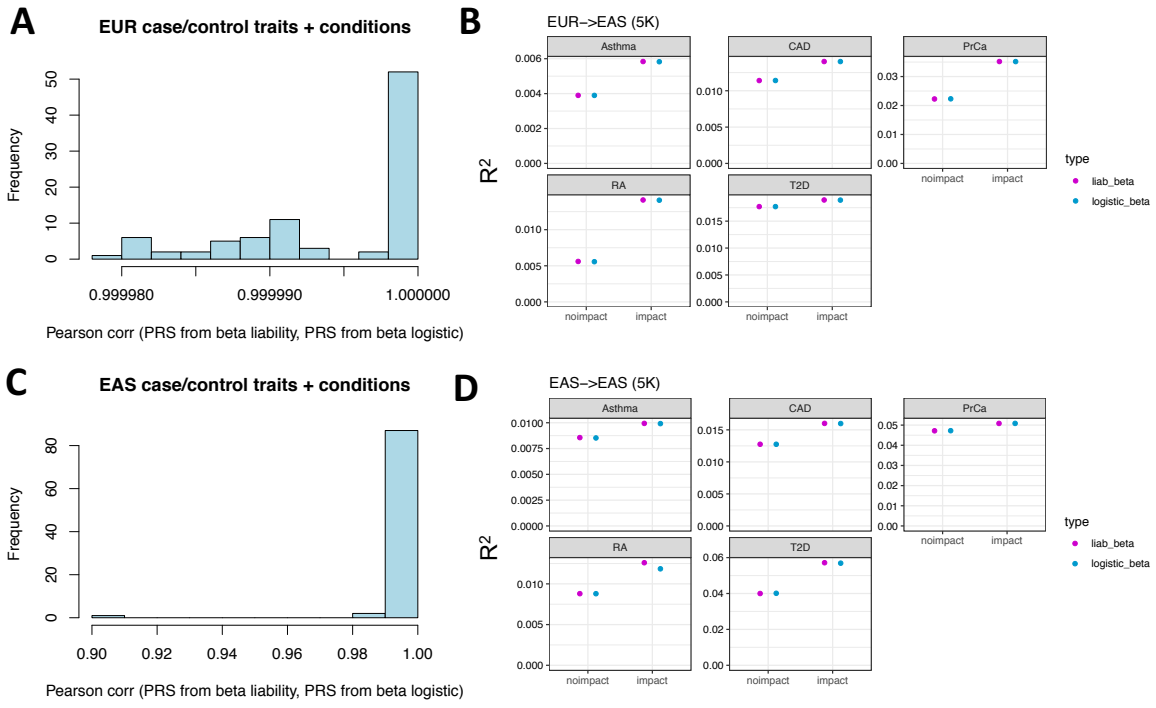


Figure S26 legend. For each of five case/control diseases considered in PRS analyses, we computed the correlation of PRS values based on EUR effect size estimates calculated on the logistic scale versus the liability scale (panel A for PRS-EUR and panel C for PRS-EAS). All sets of variants were considered for this analysis, e.g. 9 *P* value thresholds x 2 model types (IMPACT/standard PRS) x 5 case/control diseases = 90. We also compare logistic and liability scale PRS R^2 between IMPACT-informed and standard P+T models (panel B for PRS-EUR and panel D for PRS-EAS). For this analysis, we only considered the *P* value threshold that achieved the highest R^2 for IMPACT and standard P+T models.

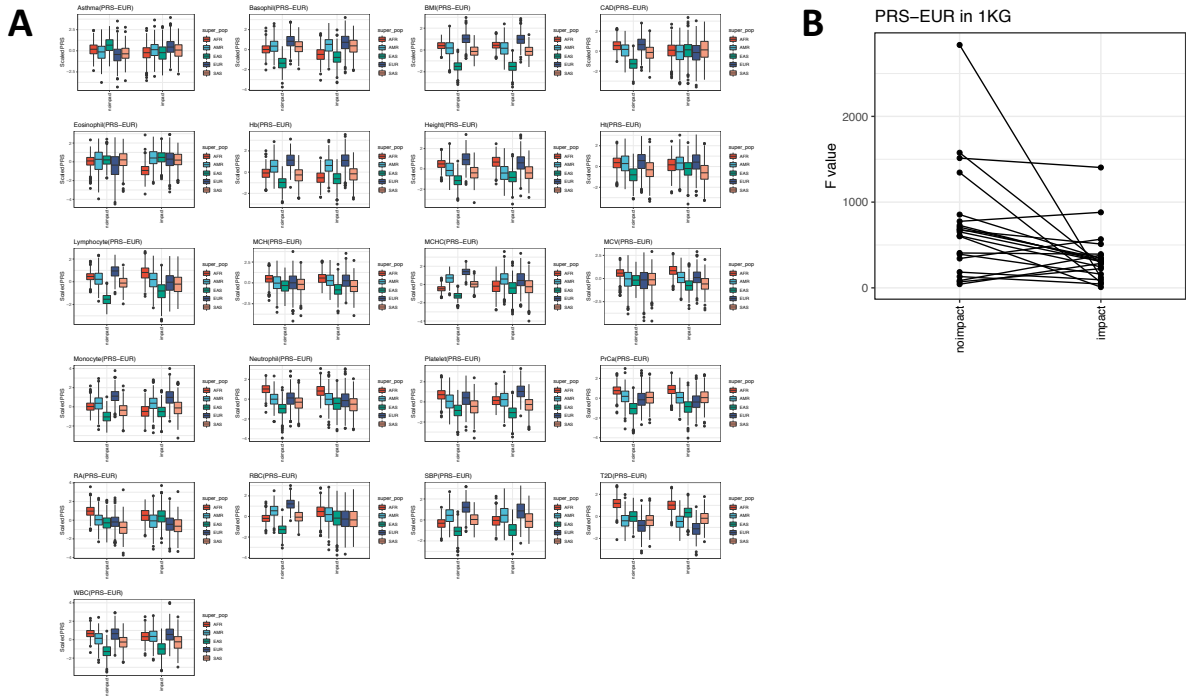


Figure S27 legend. A) For each of 21 traits considered in the EUR PRS analyses, we compare the variance in the polygenic risk scores based on standard P+T and IMPACT-informed P+T using the model that achieved the highest R^2. B) We used anova to compare the observed variance of PRS distributions across the five different 1000G populations, for each trait between standard P+T PRS and IMPACT-informed PRS, by computing F-statistics.

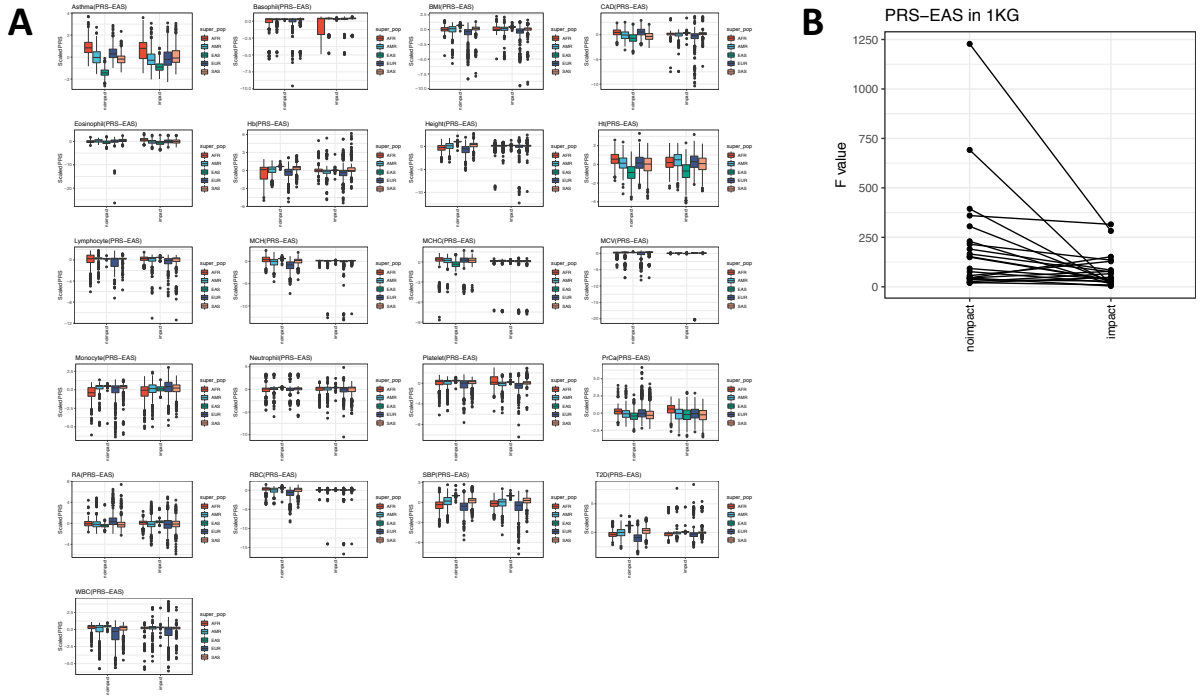
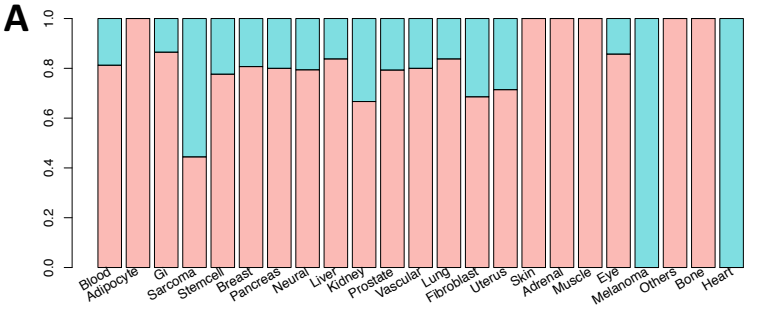


Figure S28 legend. A) For each of 21 traits considered in the EAS PRS analyses, we compare the variance in the polygenic risk scores based on standard P+T and IMPACT-informed P+T using the model that achieved the highest R^2. B) We used anova to compare the observed variance of PRS distributions across the five different 1000G populations, for each trait between standard P+T PRS and IMPACT-informed PRS, by computing F-statistics.



B

A	B	Correlation
No. ChIP-seq peaks	Size of Annotation	r = 0.22, p < 1.5e-9
No. ChIP-seq peaks	AUPRC	r = 0.39, p < 1.5e-24
AUPRC	Size of Annotation	r = -0.25, p < 4.8e-11

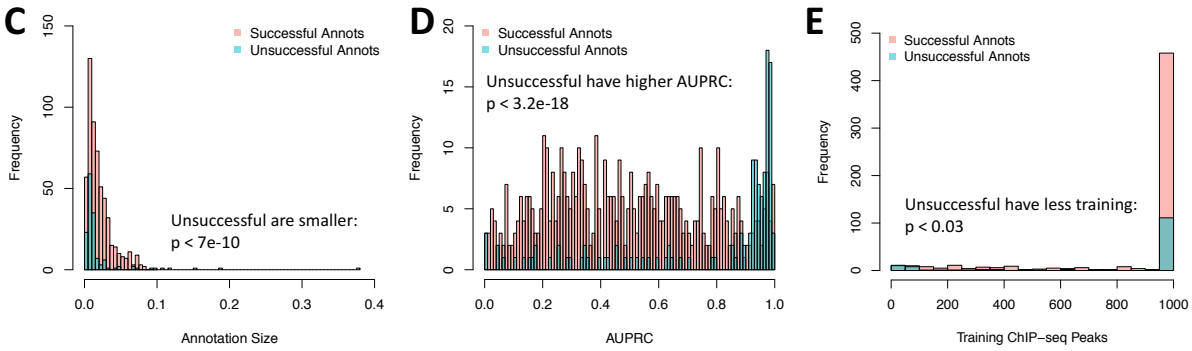


Figure S29 legend. A) Distribution of annotation size (average IMPACT score over annotated SNPs) for "successful" and "unsuccessful" annotations. B) Distribution of TF binding model AUPRC for "successful" and "unsuccessful" annotations. C) Distribution of training set size (number of TF ChIP-seq peaks) for "successful" and "unsuccessful" annotations. D) Correlation of metadata factors of IMPACT annotations: number of ChIP-seq peaks available to training data, AUPRC of TF binding prediction model, and annotation size. E) For each tissue type category of IMPACT annotation, the proportion of annotations that were significantly associated with at least one polygenic trait or disease ("successful") is indicated by the height of the pink bar. "Unsuccessful" annotations were not found to be significantly associated with any phenotype and are indicated by the green bar. For example, heart-labeled annotations had no significant associations.

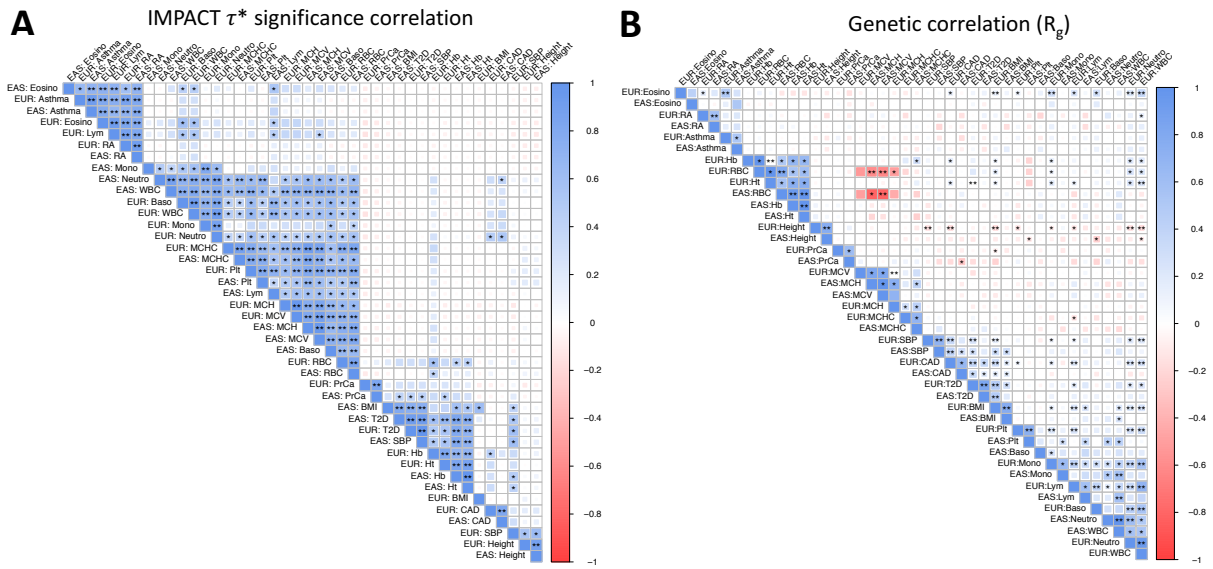


Figure S30 legend. A) Pairwise correlation of IMPACT functional annotations' τ^* significance across 42 traits, accounting for 21 unique phenotypes (those with at least one significant IMPACT association in both EUR and EAS) and two populations. * indicates FDR-adjusted $P < 0.05$, ** indicates FDR-adjusted $P < 1e\text{-}10$. B) Pairwise genetic correlation across the same 42 traits as in (A). * indicates nominal $P < 0.05$, ** indicates nominal $P < 1e\text{-}10$.

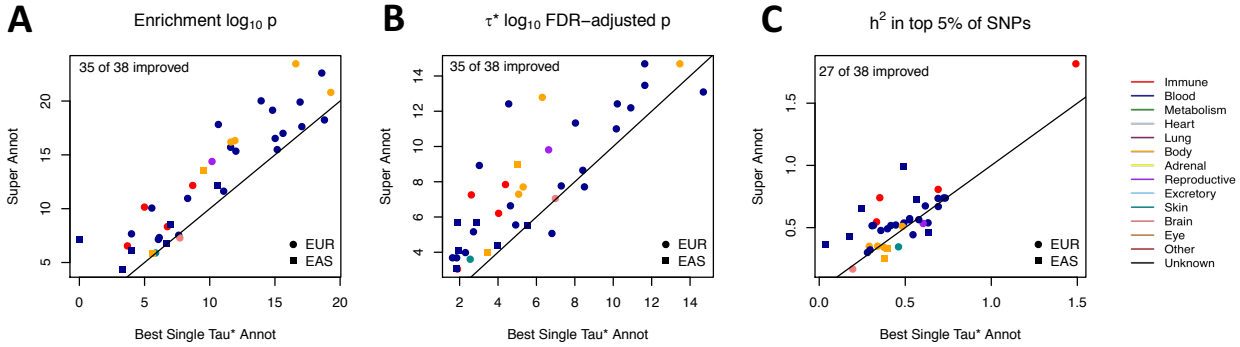


Figure S31 legend. Comparison of heritability metrics between the lead annotation and the composite annotation, created from independently associated IMPACT annotations. A) Statistical significance of the enrichment estimate. B) Statistical significance of the τ^* S-LDSC regression coefficient estimate. C) Proportion of observed scaled heritability in the top 5% SNPs scored by IMPACT.

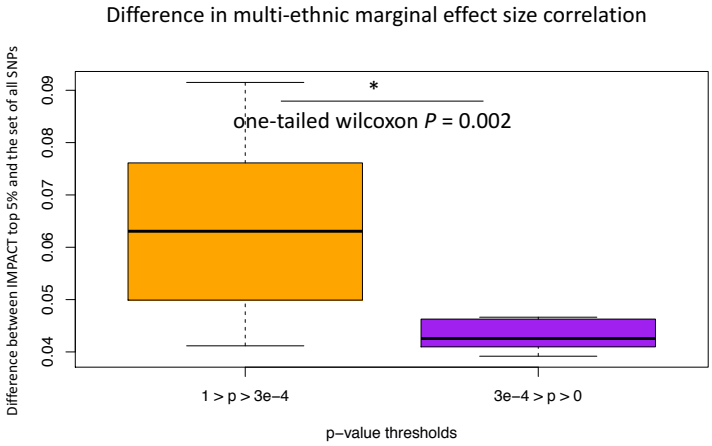
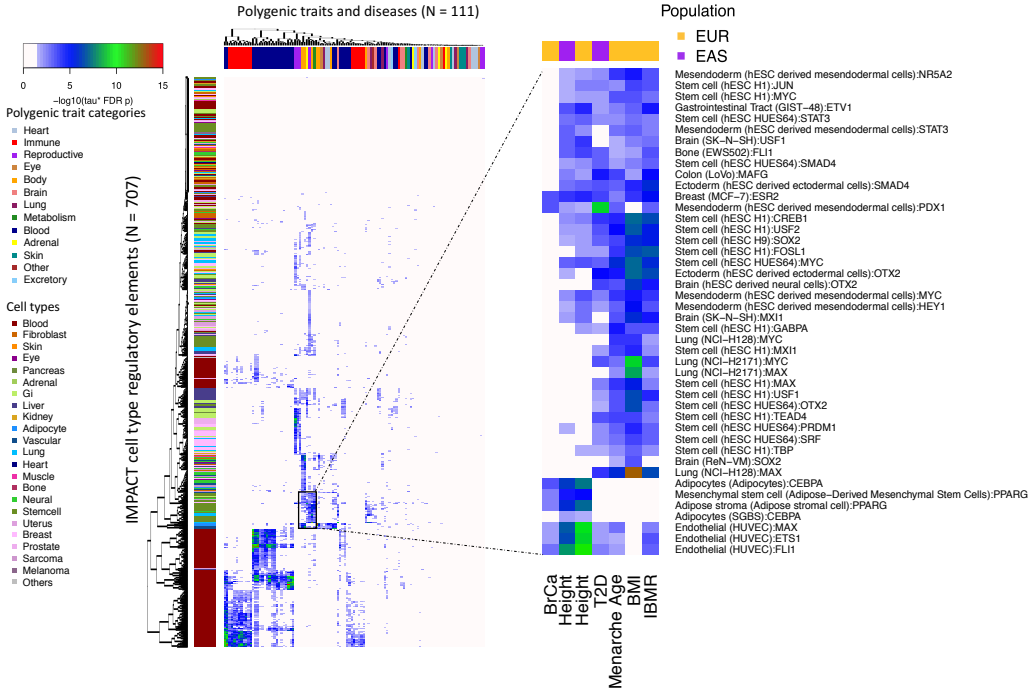


Figure S32 legend. Improvement by functional data (IMPACT top 5% SNP selection) varies by *P* value threshold. Improvement is greatest when p-values are lenient (orange). Improvement is minimized when the EUR GWAS *P* value is near or past the genome-wide significant threshold (purple).

Extended Data Figures



Ext. Data 1 legend. Significant cell type-phenotype associations across 707 IMPACT regulatory annotations and 111 complex traits and diseases at τ^* 5% FDR, color indicates -log10 FDR 5% adjusted *P* value of τ^*. Zooms shows particular cell type categories enriched for polygenic trait associations.

Supplement References

1. Dey, K. K., Van de Geijn, B., Kim, S. S. & Hormozdiari, F. Evaluating the informativeness of

deep learning annotations for human complex diseases. *bioRxiv* (2019).

2. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).