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

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

Tiffany Amariuta*[1-5], Kazuyoshi Ishigaki*[1-3,6], Hiroki Sugishita[7], Tazro Ohta[8,9], Masaru Koido[6,10], Kushal Dey[11], Koichi Matsuda[12,13], Yoshinori Murakami[10], Alkes L. Price[3,11,14], Eiryo Kawakami[8,15], Chikashi Terao[6,16,17], Soumya Raychaudhuri[1-5,18]

[1]Center for Data Sciences, Harvard Medical School, Boston, Massachusetts, 02115, USA.
[2]Divisions of Genetics and Rheumatology, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, 02115, USA.
[3]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, 02142, USA.
[4]Department of Biomedical Informatics, Harvard Medical School, Boston, Massachusetts, 02115, USA.
[5]Graduate School of Arts and Sciences, Harvard University, Cambridge, Massachusetts, 02138, USA.
[6]Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences, Kanagawa, 230-0045 Japan.
[7]Laboratory for Developmental Genetics, RIKEN Center for Integrative Medical Sciences (IMS), Kanagawa, Japan.
[8]Medical Sciences Innovation Hub Program, RIKEN, Kanagawa, Japan.
[9]Database Center for Life Science, Joint Support-Center for Data Science Research, Research Organization of Information and Systems, Shizuoka, Japan.
[10]Division of Molecular Pathology, Institute of Medical Science, The University of Tokyo, Tokyo, 108-8639 Japan.
[11]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA.
[12]Laboratory of Genome Technology, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, 108-8639 Japan.
[13]Laboratory of Clinical Genome Sequencing, Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, 108-8639 Japan.
[14]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA.
[15]Artificial Intelligence Medicine, Graduate School of Medicine, Chiba University, Chiba, Japan.
[16]Clinical Research Center, Shizuoka General Hospital, Shizuoka, 420-8527 Japan.
[17]The Department of Applied Genetics, The School of Pharmaceutical Sciences, University of Shizuoka, Shizuoka, 422-8526 Japan.
[18]Centre for Genetics and Genomics Versus Arthritis, Centre for Musculoskeletal Research, Manchester Academic Health Science Centre, The University of Manchester, Manchester, UK.

*indicates equal contributions
**Correspondence to Soumya Raychaudhuri, soumya@broadinstitute.org

Abstract

Poor trans-ethnic portability of polygenic risk score (PRS) models is an important issue caused in part by Eurocentric genetic studies and in part by limited knowledge of causal variants shared among populations. Hence, leveraging noncoding regulatory annotations that capture genetic variation across populations has the potential to enhance the trans-ethnic portability of PRS. To this end, we constructed a unique resource of 707 cell-type-specific IMPACT regulatory annotations by aggregating 5,345 public epigenetic datasets to predict binding patterns of 142 cell-type-regulating transcription factors across 245 cell types. With this resource, we partitioned the common SNP heritability of diverse polygenic traits and diseases from 111 GWAS summary statistics of European (EUR, average N=180K) and East Asian (EAS, average N=157K) origin. For 95 traits, we were able to identify a single IMPACT annotation most strongly enriched for trait heritability. Across traits, these annotations captured an average of 43.3% of heritability (sem = 2.8%) with the top 5% of SNPs. Strikingly, we observed highly concordant polygenic trait regulation between populations: the same regulatory annotations captured statistically indistinguishable SNP heritability (fitted slope = 0.98, sem = 0.04). Since IMPACT annotations capture both large and consistent proportions of heritability across populations, prioritizing variants in IMPACT regulatory elements may improve the trans-ethnic portability of PRS. Indeed, we observed that EUR PRS models more accurately predicted 21 tested phenotypes of EAS individuals when variants were prioritized by key IMPACT tracks (49.9% mean relative increase in R^2). Notably, the improvement afforded by IMPACT was greater in the trans-ethnic EUR-to-EAS PRS application than in the EAS-to-EAS application (47.3% vs 20.9%, one-tailed paired wilcoxon $P < 0.012$). Overall, our study identifies a crucial

role for functional annotations such as IMPACT to improve the trans-ethnic portability of genetic data.

Introduction

An important challenge for complex trait genetics is that there is no clear framework to transfer population-specific genetic data, such as GWAS results, to individuals of other ancestries[1–3]. The importance of this challenge is accentuated by the fact that approximately 80% of all genetic studies have been performed with individuals of European ancestry, accounting for a minority of the world's population[4]. This is exacerbated by the fact that population-specific linkage disequilibrium (LD) between variants confounds inferences about causal cell types and variants (**Figure 1A**)[5–7]. GWAS have the potential to revolutionize the clinical application and utility of genetic data to the individual, exemplified by current polygenic risk score (PRS) models[5,8–16]. However, while the utility of PRS models relies on accurate estimation of allelic effect sizes from GWAS and benefits from genetic similarity between the target cohort and the training GWAS cohort, recent studies have explicitly observed a lack of trans-ethnic portability[2,3,5,8,17,18]. The Eurocentric GWAS bias has led PRS to be more predictive in European populations, as the largest training data comes from European GWAS[3,5,12,19,20]. As a result, variants used in European PRS tend to be more common among Europeans and less common among non-Europeans. Common variants carry greater disease predictive power which directly contributes to Eurocentric bias in PRS accuracy[3]. The trans-ethnic portability of PRS would not be as critical an issue if large GWAS were performed in all non-EUR populations. Previous studies have extensively shown that functional annotations can improve PRS models when learned and applied to the same population[21,22], by introducing biologically-relevant priors on causal effect sizes and compensating for inflation of association statistics by LD.

Figure 1

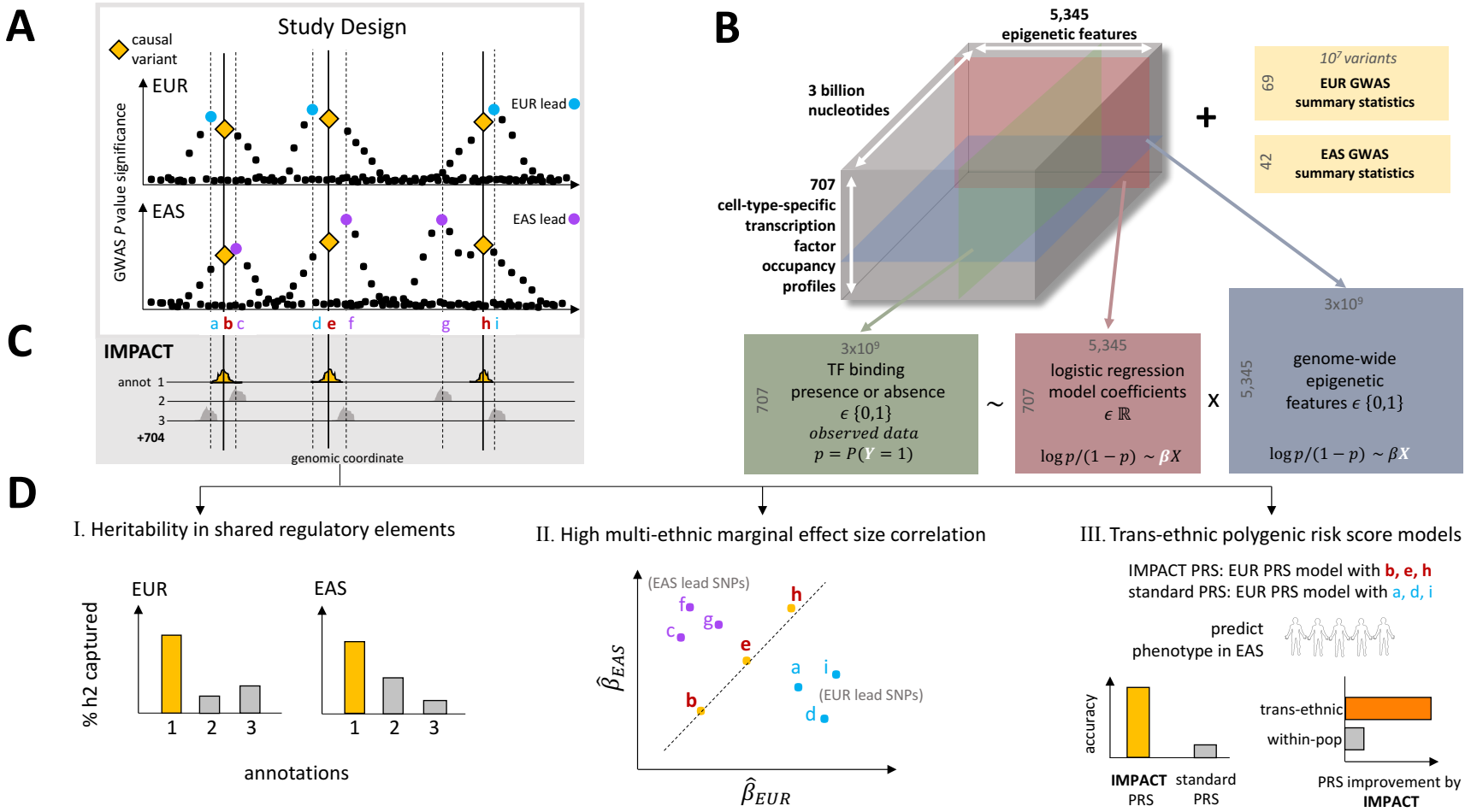


Figure 1 legend. Study design to identify regulatory annotations that prioritize regulatory variants in a multi-ethnic setting. A) Population-specific LD confounding and subsequent inflation of GWAS associations complicate the interpretation of summary statistics and transferability to other populations; functional data may help improve trans-ethnic genetic portability. B) Prism of functional data in IMPACT model: 707 genome-wide TF occupancy profiles (green), 5,345 genome-wide epigenomic feature profiles (blue), and fitted weights for these features (pink) to predict TF binding by logistic regression. Using IMPACT annotations, we investigate 111 GWAS summary datasets (yellow) of EUR and EAS origin. C) Compendium of 707 genome-wide cell-type-specific IMPACT regulatory annotations. D) Annotations that prioritize common regulatory variants must I) capture large proportions of heritability in both populations, II) account for consistent marginal effect size estimations between populations and III) improve the trans-ethnic application of PRS.

However, the potential for functional annotations to improve trans-ethnic PRS frameworks, where the influences of population-specific LD are more profound, has not yet been extensively investigated.

However, designing functional annotations that may improve PRS models is challenging. Functional annotations that best capture polygenic trait genetic variation must identify a large number of functional variants genome-wide without compromising specificity for trait-relevant regulatory programs. Pinpointing these mechanisms is especially difficult despite the fact that genome-wide association studies (GWAS) have identified thousands of genetic associations with complex phenotypes[8,23–25]. It has been estimated that about 90% of these associations reside in protein noncoding regions of the genome, making their mechanisms difficult to interpret[26,27]. Defining the etiology of complex traits and diseases requires knowledge of phenotyping-driving cell types in which these associated variants act. Transcription factors (TFs) are poised to orchestrate large polygenic regulatory programs as genetic variation in their target regions can modulate gene expression, often in cell-type-specific contexts[28,29]. Genomic annotations marking the precise location of TF-mediated cell type regulation can be exploited to elucidate the genetic basis of polygenic traits.

To overcome these challenges, we previously developed IMPACT, a genome-wide cell-type-specific regulatory annotation strategy that models the epigenetic pattern around TF binding using linear combinations of functional annotations[30]. In rheumatoid arthritis (RA), IMPACT CD4+ T cell annotations captured substantially more heritability than functional annotations derived from single experiments, including TF and histone modification ChIP-seq[6]. In this study, we expanded this approach by aggregating 5,345 functional annotations with an identical implementation of the IMPACT model framework using the same set of optimized

parameters as previously calibrated. We created a powerful and generalizable resource of 707 cell-type-specific gene regulatory annotations (**Web Resources**) based on binding profiles of 142 TFs across 245 cell types (**Figure 1B,C**). This study builds on our previous work[30] in which we created 13 annotations (13 TF-cell type pairs) based on 515 functional annotations; we observed remarkable consistency of IMPACT predictions for the same TF-cell type pair despite different training data and epigenetic features (**SF1**). Assuming that causal variants are largely shared between populations[2,23], we hypothesized that restricting PRS models to variants within trait-relevant IMPACT annotations, which are more likely to have regulatory roles and less likely to be solely associated via linkage, will especially improve their trans-ethnic portability.

In this study, we identify key IMPACT regulatory annotations that capture genome-wide polygenic mechanisms underlying a diverse set of complex traits, supported by population non-specific enrichments of genetic heritability, multi-ethnic marginal effect size correlation (a possible mechanism of improved PRS), and improved trans-ethnic portability of PRS models (**Figure 1D**). Here, we defined and employed our compendium of 707 IMPACT regulatory annotations to study polygenic traits and diseases from 111 GWAS summary datasets of European (EUR) and East Asian (EAS) origin. Assuming shared causal variants between populations, annotations that prioritize shared regulatory variants must (1) capture disproportionately large amounts of genetic heritability in both populations, (2) be enriched for multi-ethnic marginal effect size correlation, and (3) improve the trans-ethnic applicability of population-specific PRS models. Using our compendium of regulatory annotations, we identified key annotations for each polygenic trait and demonstrated their utility in each of these three applications toward prioritization of shared regulatory variants. Overall, this work

improves the interpretation and trans-ethnic portability of genetic data and provides implications for future clinical implementations of risk prediction models.

Results

Building a compendium of *in silico* gene regulatory annotations

To capture genetic heritability of diverse polygenic diseases and quantitative traits, we constructed a comprehensive compendium of 707 cell type regulatory annotation tracks. To do this, we applied the IMPACT[30] framework to 707 unique TF-cell type pairs obtained from a total of 3,181 TF ChIP-seq datasets from NCBI, representing 245 cell types and 142 TFs with known sequence motifs (**Figure 1B**, **Online Methods**, **Web Resources**, **ST1**, **SF2**)[31]. We provide publicly available open-source software (see **Web Resources**) corresponding to the analyses presented in this manuscript. We caution that the 707 TF/cell type pairs represented in publicly available data is a small fraction of the total possible pairs of 142 TFs and 245 cell types (n = 34,790), although there are several experimental and practical reasons why this theoretical maximum is not reached (**Discussion**). Briefly, IMPACT learns an epigenetic signature of active TF binding evidenced by ChIP-seq, differentiating bound from unbound TF sequence motifs using logistic regression. We derive this signature from 5,345 epigenetic and sequence features, predominantly generated by ENCODE[32] and Roadmap[33] (**Online Methods**, **ST2**); these data were drawn from diverse cell types, representing the biological range of the 707 candidate models. IMPACT then probabilistically annotates the genome, e.g. on a scale from 0 to 1, without using the TF motif, identifying regulatory regions that are similar to those that the TF binds.

To assess the specificity of our IMPACT annotations, we test whether they (1) accurately predict binding of the modeled TF, (2) share cell-type-specific characteristics with other tracks of the same cell type, and (3) score cell-type-specifically expressed genes higher than nonspecific genes. The 707 models that we defined had a high TF binding prediction accuracy with mean AUPRC = 0.54 (sem = 0.01, **Online Methods**, **SF3**) using cross-validation. Annotations segregated by cell type rather than by TF, excluding CTCF, suggesting the same TF may bind to different enhancers in different cell types (**Figure 2A**). On average, we observed that annotations of the same cell types were more strongly correlated genome-wide (Pearson r = 0.56, sem = 0.02) than annotations of different cell types (Pearson r = 0.48, sem = 0.01, one-tailed difference of means $P < 0.001$, **SF3**). Furthermore, the covariance structure between TF ChIP-seq training datasets is similar to that of corresponding IMPACT annotations, indicating that the IMPACT model does not introduce spurious correlations among unrelated ChIP-seq datasets (**SF3**). Lastly, for nine different cell types, we examined cell-type-specifically expressed genes from Finucane et al[34] and corresponding differential expression t-statistics. For each of nine cell types, we observed larger cell-type-specific IMPACT probabilities at SNPs in and near cell-type-specific genes compared to generally expressed genes (mean fold-change across 10 to 99 cell-type-specific IMPACT tracks ranged from 1.08 to 1.96 across nine cell types, one-tailed paired wilcoxon $P < 0.04$ for seven of nine cell types, **Figure 2B**, **SF3**, **Online Methods**), suggesting that IMPACT annotates relevant cell type regulatory elements.

Partitioning common SNP heritability of 111 GWAS summary statistics in EUR and EAS

We obtained summary statistics from 111 publicly available GWAS for diverse polygenic traits and diseases. For narrative purposes throughout the text, we use five genetically uncorrelated (R_g point estimates between traits ranged from -0.08 to 0.20, **ST3**, although no

Figure 2

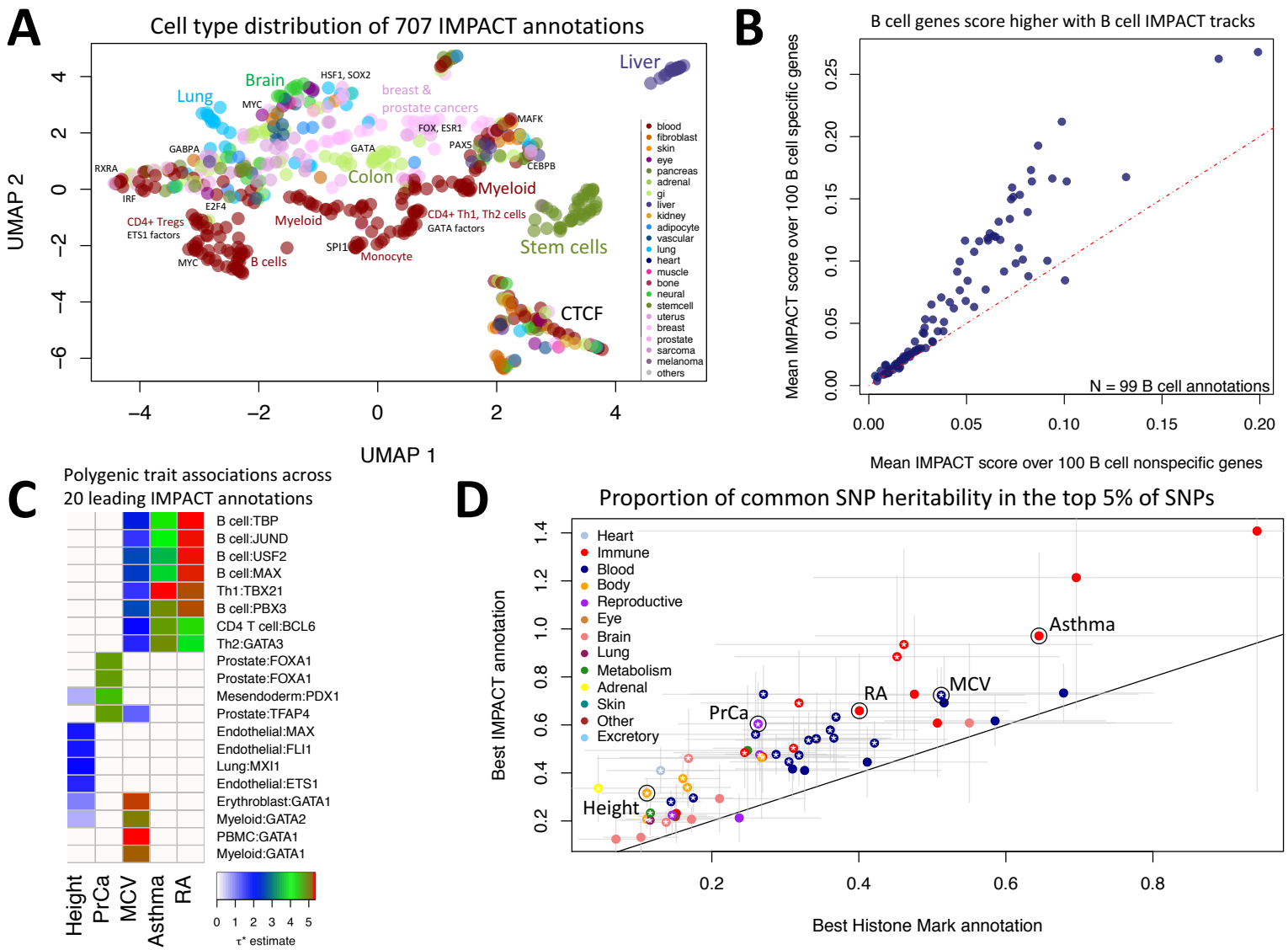


Figure 2 legend. IMPACT annotates relevant cell type regulatory elements. A) Low-dimensional embedding and clustering of 707 IMPACT annotations using uniform manifold approximation projection (UMAP). Annotations colored by cell type category; TF groups indicated where applicable. B) IMPACT annotates cell type specifically expressed genes with higher scores than nonspecific genes. C) Biologically distinct regulatory modules revealed by cell type-trait associations with significantly nonzero τ^*. Shown here are the 5 representative EUR complex traits and the 4 leading IMPACT annotations for each, resulting in 20 IMPACT annotations highlighted from 707 total. Color indicates τ^* value. D) Lead IMPACT annotations capture more heritability than lead cell-type-specific histone modifications across 60 of 69 EUR summary statistics for which a lead IMPACT annotation was identified. * indicates heritability estimate difference of means $P < 0.05$. Gray segments indicate the 95% CI around the heritability estimate.

189 R_g was significantly different from 0, all two-tailed z test $P > 0.40$ after Bonferroni correction
190 for 10 pairs) and biologically diverse traits that capture the spectrum of summary statistics
191 analyzed in order to exemplify our results in addition to reporting metrics averaged over all
192 traits analyzed. These five traits include an allergic phenotype: asthma, an autoimmune disease:
193 RA, a neoplastic type: prostate cancer (PrCa), a hematological quantitative trait: mean
194 corpuscular volume (MCV), and an anthropometric trait: height. These included 69 from EUR
195 participants[30,35] (average N = 180K, average heritability z-score = 12.9, 41/69 from UK
196 BioBank)[6,36] and 42 from EAS participants of BioBank Japan[3,37–39] (average N = 157K, average
197 heritability z-score = 6.6)[24] (**ST3**). We chose to focus our study on EUR and EAS populations, as
198 there is a limited number of large GWAS in populations other than EUR and EAS[4,40,41]. All of the
199 summary statistics used were generated from studies that had a sample size greater than
200 10,000 individuals and also had a significantly non-zero heritability (z-score > 1.97). There are
201 29 phenotypes for which we obtained summary statistics in both EUR and EAS. We were
202 interested to see if any traits had a multi-ethnic genetic correlation that deviated from 1.
203 Therefore, we explicitly tested this and found that 16 traits have multi-ethnic R_g that does not
204 deviate from 1 (one-tailed z test $P > 0.05/29$ tested traits), while 13 traits have multi-ethnic R_g
205 that does deviate from 1 (one-tailed z test $P < 0.05/29$ tested traits). Overall we observed high
206 R_g for most traits, supporting our assumption that causal variants are generally shared across
207 populations (**Online Methods**, **SF4**)[42]. At two extremes, basophil count has a low multi-ethnic
208 R_g of 0.32 (sd = 0.10), while atrial fibrillation has a high multi-ethnic R_g of 0.98 (sd = 0.11),
209 consistent with previous observations made using *Popcorn*, but using different parameter
210 estimation strategies (**Online Methods**)[3].

We then partitioned the common SNP (minor allele frequency (MAF) > 5%) heritability of these 111 datasets using S-LDSC[6] with an adapted baseline-LD model excluding cell-type-specific annotations[30,35] (**SF4, Online Methods**). Here, heritability refers to the inferences made by S-LDSC about the heritability causally explained by common SNPs as defined previously[6], as opposed to genotyping-array-based SNP-heritability[43,44] or other definitions. We caution that the results presented herein are a consequence of the analyzed GWAS populations, polygenic traits and diseases, and available experimental data to create functional annotations. Next, we tested each of the traits against each of the 707 IMPACT annotations, assessing the significance of a non-zero τ^*, which is defined as the proportionate change in per-SNP heritability associated with a one standard deviation increase in the value of the annotation (**Online Methods**)[35]. Of 707 by 111 (n = 78,477) possible associations subjected to 5% FDR, we detected 7,993 associations, 5% of which we expect to be false positives. We observed that 95 phenotypes had at least one significant annotation-trait association ($\tau^* > 0$, two-tailed z test $P <$ 0.05 at 5% FDR, **Ext. Data 1, Online Methods**, **ST4-8**). Here, we highlight the four leading IMPACT annotations associated with EUR summary statistics for each of the five exemplary phenotypes mentioned above: asthma, RA, PrCa, MCV, and height (**Figure 2C**, associations between all traits and annotations in **Ext. Data 1**). Consistent with known biology, B and T cells were strongly associated with asthma[45], RA[46], and MCV[47,48] while other blood cell regulatory annotations predominantly derived from GATA factors were also associated with MCV. Prostate cancer cell lines were associated with PrCa, while many cell types including myoblasts[49], fibroblasts[50], and adipocytes[51,52], lung cells, and endothelial cells were associated with height, perhaps related to musculo-skeletal developmental pathways.

For each trait, we defined the lead IMPACT regulatory annotation as the annotation capturing the greatest per-SNP heritability, e.g. the largest, while significant, τ^* estimate (**ST9**). With the top 5% of SNPs, lead IMPACT annotations captured an average of 43.3% of common SNP heritability (sem = 2.8%) across these 95 polygenic traits (**SF5**, **Online Methods**), with more than 25% of heritability captured for two-thirds of the tested summary statistics (73/111 traits) and more than 50% captured for 28% (31/111). Identifying functional annotations that capture large proportions of heritability is an important step to understanding biological mechanisms of genetic variation. We observed higher heritability enrichments for autoimmune diseases and hematological traits, likely due to the abundance of blood cell types represented by our IMPACT annotations and possibly due to a single or a few related causal cell types. On the other hand, we observed lower heritability enrichment for brain-related, lung-related, and adrenal traits, likely due to the underrepresentation of relevant tissue or cell types in the TF ChIP-seq data and possibly due to multiple different causal cell types. We observed significantly greater τ^* of lead IMPACT annotations among traits with lower estimated polygenicity (linear regression coefficient = -0.11, $P < 3.97e-5$). Traits with higher polygenicity may be driven by more than one causal cell type; therefore a single IMPACT annotation may capture a smaller proportion of total common SNP heritability. Returning to our five exemplary phenotypes, with the top 5% of EUR SNPs, IMPACT captured 97.1% (sd = 17.6%) of asthma heritability with the T-bet Th1 annotation, 65.9% (sd = 12.1%) of RA heritability with the B cell TBP annotation, 60.4% (sd = 8.9%) of PrCa heritability with the prostate cancer cell line (LNCAP) TFAP4 annotation, 72.4% (sd = 6.0%) of MCV heritability with the GATA1 PBMC annotation, and lastly 31.6% (sd = 3.0%) of height heritability with the lung MXI1 annotation (**Figure 2D**). While the observed association

255 between lung and height is not intuitive, within the MXI1 gene lies a genome-wide significant
256 variant associated with height[53].
257 To demonstrate the value of IMPACT tracks, we compared them to annotations derived
258 from single experimental assays and from machine learning models. For example, since each of
259 the IMPACT tracks was trained on TF ChIP-seq data, we compared the per-annotation
260 standardized effect sizes (τ^*) achieved by both annotation types. We observed that on average
261 the τ^* of lead IMPACT annotations (mean τ^* = 3.53, sem = 0.91) was greater than by the
262 analogous TF ChIP-seq used in training (mean τ^* = 1.71, sem = 0.94, across 95 traits one-tailed
263 paired wilcoxon $P < 2.6e-16$). We then compared IMPACT tracks to histone marks, which are
264 commonly used to quantify cell type heritability[6]. From 220 publicly available cell-type-specific
265 histone mark ChIP-seq annotations of EUR SNPs[6], we selected the lead histone mark track for
266 each of 69 EUR summary statistics (**Online Methods**). Restricting to the top 5% of SNPs, we
267 observed that the mean EUR heritability captured by lead IMPACT annotations (49.5%, sem =
268 3.2%) was on average greater than by lead histone mark annotations (29.1%, sem = 2.5%, one-
269 tailed paired wilcoxon $P < 8.8e-12$, **Figure 2D, ST10**). For example, the lead IMPACT annotation
270 for asthma captured 64.2% (sd = 15.5%) of heritability, 1.5x more heritability than the lead
271 histone mark annotation (H3K27ac in CD4+ Th2). Similarly, IMPACT captured 1.7x more RA
272 heritability than H3K4me3 in CD4+ Th17s; IMPACT captured 1.4x more MCV heritability than
273 H3K4me3 in CD34+ cells; IMPACT captured 2.3x more PrCa heritability than H3K4me3 in CD34+
274 cells; and IMPACT captured 3.1x more height heritability than H3K4me3 in lung cells. In terms
275 of τ^*, IMPACT also captured more per-SNP heritability than histone marks (one-tailed paired
276 wilcoxon $P < 9.1e-9$, mean τ^* fold change across traits = 1.38x, **SF6**). We further compared the
277 heritability captured by IMPACT to annotations created from state-of-the-art deep learning

algorithms trained to predict various regulatory element marks, Basenji[54] and DeepSEA[55]. Performing a comprehensive analysis is challenging for two reasons. First, there is a limited set of genome-wide SNP-level deep learning predictions in the public domain with the exception of a few studies[56]. Second, as deep learning models are specific to a particular functional mark, comprehensive genome-wide cataloging is a combinatorially large problem which grows with the number of tested cell types, functional marks, and model types. Therefore, we performed the most comprehensive analysis that was feasible, focusing on the five representative traits. To this end, we collected 123 relevant deep learning annotations to target these traits (**ST11, Online Methods**) and selected the lead deep learning track for each trait (**Online Methods**). We observed that for each of five traits, the lead IMPACT annotation generally captured more heritability in the top 5% of SNPs (mean = 65.4%, sem = 10.9%) and resulted in generally larger τ^* (mean = 4.4, sem = 0.70) than the lead deep learning annotations (heritability mean = 39.1%, sem = 1.9%, τ^* mean = 1.6, sem = 0.30, one-tailed paired wilcoxon $P = 0.031$ for both heritability and τ^*, **SF7**). Although limited by the availability of deep learning annotations, we further compared lead IMPACT annotations to lead deep learning annotations across all 69 EUR traits and in all cases IMPACT trended toward higher heritability and τ^* (Basenji heritability comparison one-tailed paired wilcoxon $P < 2.0e-11$, DeepSEA heritability comparison $P < 1.4e-10$, Basenji τ^* comparison one-tailed paired wilcoxon $P < 3.4e-11$, DeepSEA τ^* comparison $P < 8.8e-12$, **Supplement**, **SF8**, **ST13**).

Since some of our IMPACT annotations are similar to each other (**SF3**), we performed serial conditional analyses in order to identify IMPACT annotations explaining heritability independently from one another (**Online Methods**). This strategy might identify complex traits for which several distinct biological mechanisms are independently regulated by genetic

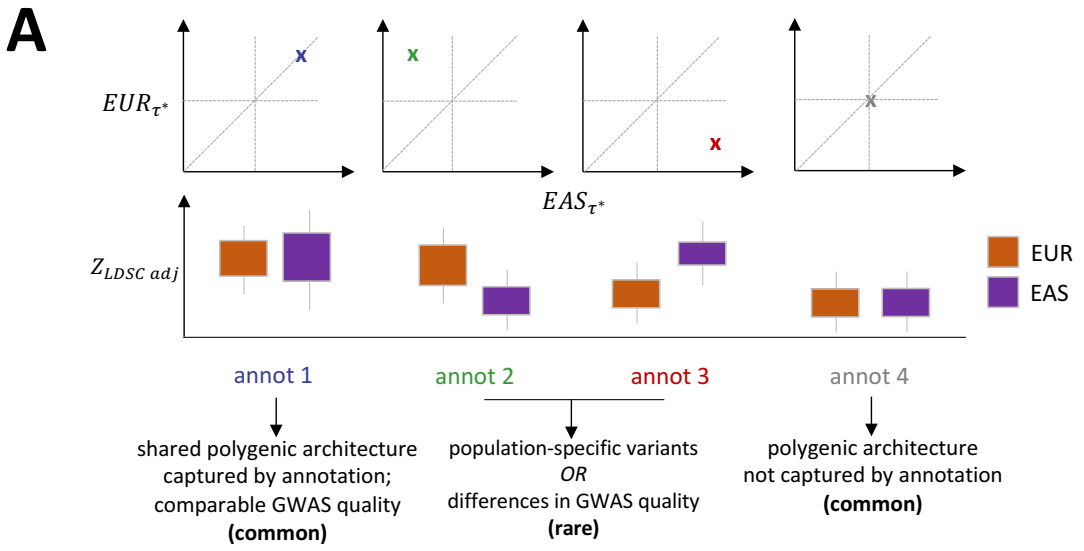
variation. Indeed, we identified 30 EUR phenotypes and 8 EAS phenotypes with multiple independent IMPACT associations (**SF9**, **ST14-15**). For example, four IMPACT annotations were independently associated with EUR PrCa: prostate (TFAP4), prostate (RUNX2), mesendoderm (PDX1), and cervix (NFYB). Moreover, for seven EUR traits, three IMPACT annotations were independently associated: 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). Among functionally correlated traits, we observed consistency in the independently associated IMPACT annotations, proposing a biological basis for genetic correlation (**Supplement**). In general, identifying functional concordance among traits with genetic correlation less than 1 provides a quantitative biological basis for the dissimilarity between traits that is orthogonal to genetic correlation approaches[42,57–60]. We found that the heritability z-score, an index correlated with the power of S-LDSC[6], is strongly predictive of the number of independent regulatory associations (linear regression coefficient = 0.06, $P < 1.2e\text{-}5$), while sample size is not (linear regression $P = 0.59$) (**SF10**). Our findings suggest that multiple independent regulatory programs can contribute to the heritability of complex traits, and we can detect them when phenotypes are sufficiently heritable and the GWAS provide accurate effect size estimation.

Concordance of polygenic regulation between European and East Asian populations

Previous studies have shown concordance of polygenic effects between EUR and EAS individuals in RA[1] and between EUR and African American individuals in PrCa[61]. However, to our knowledge, the extent of these shared effects has not yet been comprehensively investigated

across many functional annotations and in diverse traits. Assuming shared causal variants in EUR and EAS, IMPACT annotations that best prioritize shared genomic regions regulating a phenotype presumably also disproportionately capture similar amounts of heritability in both EUR and EAS (**Figure 1D-I**, **Figure 3A**). Here, we quantified the SNP heritability (τ*) of 29 traits in EUR and EAS captured by a set of approximately 100 independent IMPACT regulatory annotations (**Figure 3B, SF11, Online Methods**). Briefly, we selected independent annotations using an iterative pruning approach: for each trait, we ranked all annotations by τ* and removed any annotation correlated with Pearson r^2 > 0.5 to the lead annotation and then repeated. As IMPACT annotations are independent of population-specific factors including LD and allele frequencies (**SF4**), they are poised to capture the genome-wide distribution of regulatory variation in a population-independent manner. We observed that τ* estimates across annotations for EUR and EAS are strikingly similar, with a regression coefficient that is consistent with identity (slope = 0.98, sem = 0.04). For example, we observed a strong Pearson correlation of τ* between EUR and EAS for asthma (r = 0.98), RA (r = 0.87), MCV (r = 0.96), PrCa (r = 0.90), and height (r = 0.96). Cross-ancestry functional concordance is not specific to IMPACT annotations as we observed a similar relationship among cell-type-specific histone marks using the same strategy (**SF12**)[24]. Additionally considering 513 cell-type-specifically expressed gene sets (SEG)[24,34], we could not observe cross-ancestry concordance due to too few significant associations shared between populations. Furthermore, we found that none of our τ* estimates show evidence of population heterogeneity (all two-tailed difference of means P > 0.56 at 5% FDR). This might be a result of noise around the τ* estimates, such that true heterogeneity is too subtle to detect in this regime. Overall, our results suggest that regulatory variants in EUR and EAS populations are equally enriched within the same classes of regulatory

Figure 3



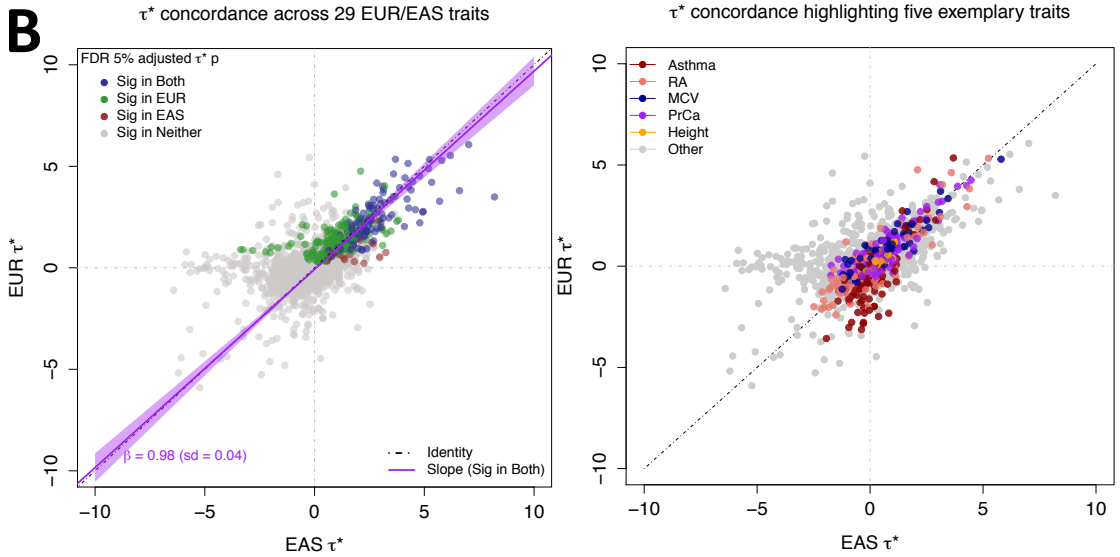


Figure 3 legend. Multi-ethnic concordance of regulatory elements defined by IMPACT. A) Illustrative concept of concordance versus discordance of τ^* between populations. Concordance implies a similar distribution of causal variants and effects captured by the same annotation. The implications of discordant τ^* are not as straightforward. B) Common per-SNP heritability (τ^*) estimate for sets of independent IMPACT annotations across 29 traits shared between EUR and EAS. Left: color indicates τ^* significance (τ^* greater than 0 at 5% FDR) in both populations (blue), significant in only EUR (green), significant in only EAS (red), significant in neither (gray). Line of best fit through annotations significant in both populations (dark purple line, 95% CI in light purple). Black dotted line is the identity line, y = x. Right: color indicates association to one of five exemplary traits.

elements. This does not exclude the possibility of population-specific variants or causal effect sizes, as evidenced by 13 traits with multi-ethnic genetic correlation significantly less than 1 ($P <$ 0.05/29 tested traits). Rather, these results suggest that causal biology, including disease-driving cell types and their regulatory elements, underlying polygenic traits and diseases, is largely shared between these populations.

Assessing variant prioritization with IMPACT toward improving polygenic risk score models

PRS models have great clinical potential: previous studies have shown that individuals with higher PRS have increased risk for disease[8–12]. In the future, polygenic risk assessment may become as common as screening for known mutations of monogenic disease, especially as it has been shown that individuals with severely high PRS may be at similar risk to disease as are carriers of rare monogenic mutations[12]. However, since PRS heavily rely on GWAS with large sample sizes to accurately estimate effect sizes, there is specific demand for the transferability of PRS from populations with larger GWAS to populations underrepresented by GWAS[2,3,5,8,17,18,22]. As we would like to investigate the ability of IMPACT annotations to improve the trans-ethnic application of PRS, we chose pruning and thresholding (P+T) as our model[3,8]. P+T models, as the name suggests, select an independent subset of all SNPs genome-wide by pruning away SNPs correlated by LD and then further thresholding on GWAS *P* value. We elected to use P+T rather than LDpred[5,22] or AnnoPred[21], which compute a posterior effect size estimate for all SNPs genome-wide based on membership to functional categories. With P+T, we can partition the genome by IMPACT-prioritized and deprioritized SNPs, whereas the assumptions of the LDpred and AnnoPred models do not support the removal of variants, making it difficult to directly assess improvement due to IMPACT prioritization. Moreover,

these models have not been explicitly designed or tested for the trans-ethnic application of PRS and thus are beyond the scope of our work. We conventionally define PRS as the product of marginal SNP effect size estimates and imputed allelic dosage (ranging from 0 to 2), summed over M SNPs in the model. Conventional P+T utilizes marginal effect size estimates and therefore is susceptible to selecting a tagging variant over the causal one guided by GWAS *P* values which are inflated by LD. Therefore, we hypothesized that any observed improvement due to incorporation of IMPACT annotations could result from prioritization of variants with higher marginal multi-ethnic effect size correlation (**Figure 1D-II**), suggesting these SNPs are less likely to be solely associated by linkage.

Hence, we tested this hypothesis before assessing PRS performance. We selected 21 of 29 summary statistics shared between EUR and EAS with an identified lead IMPACT association in both populations. Then, using EUR lead IMPACT annotations for each trait (**ST9**), we partitioned the genome in three ways: (1) the SNPs within the top 5% of the IMPACT annotation, (2) the SNPs within the bottom 95% of the IMPACT annotation, and (3) the set of all SNPs genome-wide (with no IMPACT prioritization). We then performed stringent LD pruning (r^2< 0.1 from EUR individuals of phase 3 of 1000 Genomes[62]), guided by the EUR GWAS *P* value, to acquire sets of independent SNPs in order to compute a EUR-EAS marginal effect size estimate correlation (**Online Methods**).

For example, in height, EUR-EAS effect size estimates of SNPs in the top 5% partition are 2.1-fold more similar (Pearson *r* = 0.29, **Figure 4A**) than those in the bottom 95% partition (*r* = 0.14, **Figure 4B**) and 1.6-fold more similar than the set of all SNPs (*r* = 0.18). For each of 17 GWAS *P* value thresholds, the marginal multi-ethnic effect size correlation among the top 5% of IMPACT SNPs tended to be greater than the set of all SNPs genome-wide across 21 traits (all 17

Figure 4

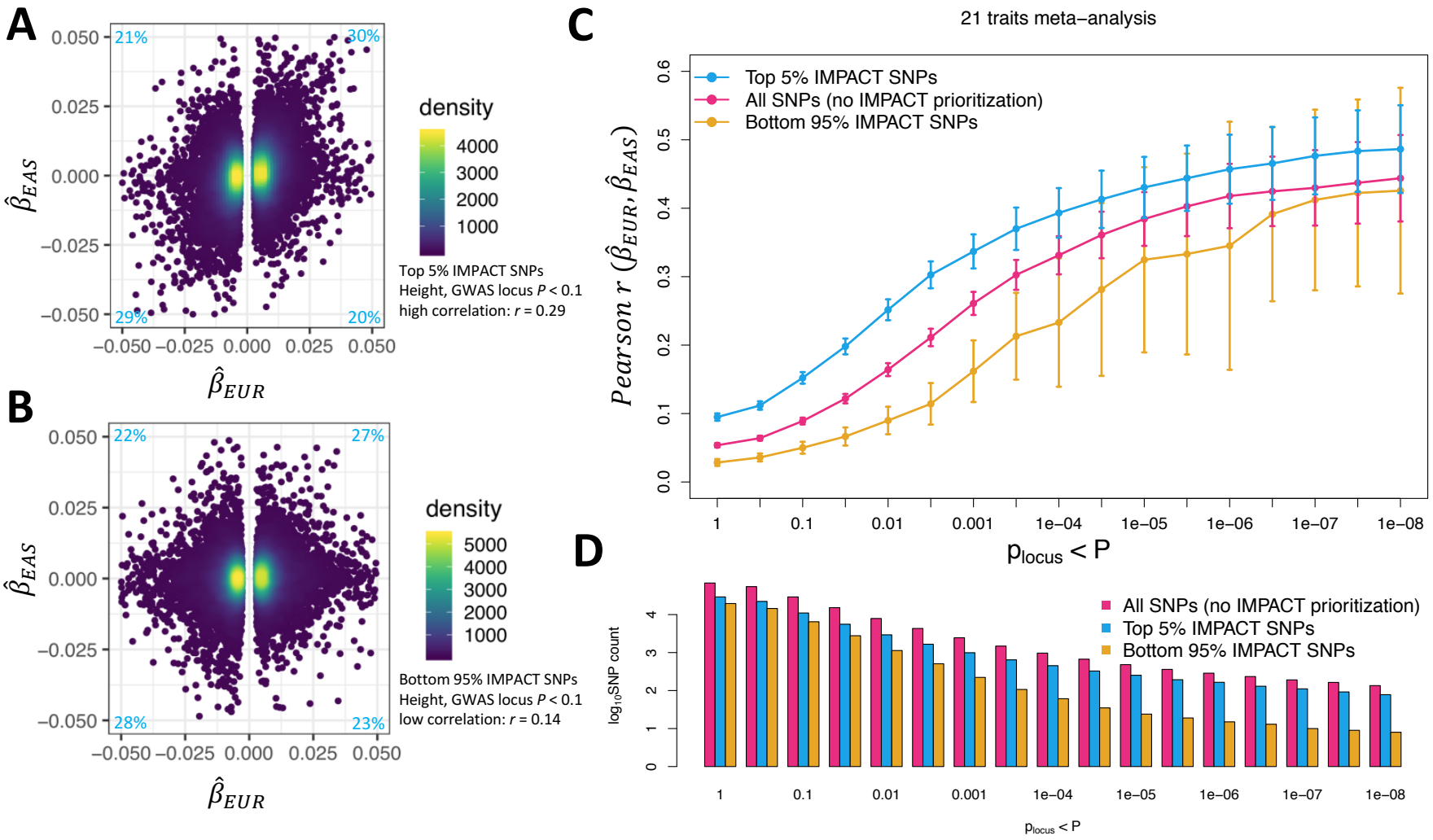


Figure 4 legend. Mechanism by which IMPACT prioritization of shared regulatory variants might improve trans-ethnic PRS performance. A) Estimated effect sizes of variants from genome-wide EUR and EAS height summary statistics in the top 5% of the lead IMPACT annotation for EUR height. Proportions of variants in each quadrant indicated in light blue. B) Estimated effect sizes from genome-wide EUR and EAS height summary statistics of variants in the bottom 95% of the same lead IMPACT annotation for height; mutually exclusive with SNPs in A). C) Meta-analysis of multi-ethnic marginal effect size correlations between populations across 21 traits shared between EUR and EAS cohorts over 17 GWAS *P* value thresholds (with reference to the EUR GWAS). Vertical bars indicate the 95% CI around the Pearson *r* estimate. D) Number of SNPs ($\log_{10}$ scale) at each *P* value threshold for each partition of the genome corresponding to C).

one-tailed paired wilcoxon $P < 6.9e-4$) (**Figure 4C-D**). Furthermore, this observation was consistent across individual traits (**SF13**). For comparison, we performed a similar analysis restricted to the five representative traits using alternative functional annotations: lead annotations from 513 cell-type-specifically expressed gene sets (SEG)[34] and 220 cell-type-specific histone mark annotations (CTS)[6] (**SF14**). Marginal effect size correlation with IMPACT was comparable to CTS when comparing the top 5% of SNPs to the set of all SNPs (at each of 17 GWAS *P* value thresholds, one-tailed paired wilcoxon $P > 0.16$, **SF15**). Similarly assessing marginal effect size correlation, IMPACT prioritization was comparable to SEG prioritization (at each of 17 GWAS *P* value thresholds, one-tailed paired wilcoxon $P > 0.06$, **SF15**). Overall, our results suggest that we might anticipate improved trans-ethnic portability of PRS models by prioritizing SNPs in key functional annotations by decreasing the likelihood of selecting SNPs solely associated by linkage.

While increased concordance of marginal effect size estimates might lead to improved trans-ethnic portability, increased concordance of allelic heterozygosity could also play a role, as allele frequency greatly affects disease predictive power. To this end, we computed the correlation of EUR and EAS heterozygosity (**Online Methods**), defined as 2pq, across the same sets of variants and traits considered in **Figure 4**. We observed IMPACT-selected variants tended to have lower concordance of heterozygosity than conventional P+T selected variants for each of 17 GWAS *P* value thresholds across 21 traits (all one-tailed paired wilcoxon $P < 0.05$, **SF16, SF17**). This is likely due to an enrichment of common variants among IMPACT-prioritized SNPs and a depletion of rare or low frequency variants (**SF16**). We then considered F_{st}, a measure of the reduction of heterozygosity and an indicator of population divergence, among IMPACT-selected SNPs (**Online Methods**). Although F_{st} trended higher among IMPACT-selected

SNPs than among conventional P+T selected variants across 21 traits at each *P* value threshold (all one-tailed paired wilcoxon $P < 0.03$), the large confidence intervals of the meta-analyzed F_{st} across traits suggest that this trend does not indicate substantial differences (across each of 17 *P* value thresholds, all two-tailed difference of means $P > 0.98$, **SF18, SF19**). These results suggest that neither increased concordance of heterozygosity nor substantial difference in F_{st} is a consequence of IMPACT prioritization.

Models incorporating IMPACT functional annotations improve the trans-ethnic portability of polygenic risk scores

Finally, we addressed our hypothesis that IMPACT annotations improve the trans-ethnic portability of PRS (**Figure 1D-III**). For each of the 21 previously analyzed traits, we built a PRS using effect size estimates from EUR summary statistics and applied it to predict phenotypes of EAS individuals from BioBank Japan (BBJ) (**Figure 5A**). Here, we compare two PRS models, both blind to any EAS genetic or functional information and removing SNPs with LD $r^2 > 0.2$, according to European individuals from phase 3 of 1000 Genomes[62]: (i) standard P+T PRS and (ii) functionally-informed P+T PRS using a subset of SNPs prioritized by the lead EUR IMPACT annotation (**Online Methods**). In functionally-informed PRS models, for each trait separately, we *a priori* selected the subset of top-ranked IMPACT SNPs (top 1%, 5%, 10%, or 50%) which explained the closest to 50% of common SNP heritability (**Online Methods**). This ensures that functional prioritization captures approximately the majority of trait-relevant genetic variation and the cumulative genetic signal among functionally-prioritized variants was consistent across traits, allowing for varying degrees of polygenicity. For all PRS models, we report results from the most accurate model across nine EUR GWAS *P* value thresholds.

Figure 5

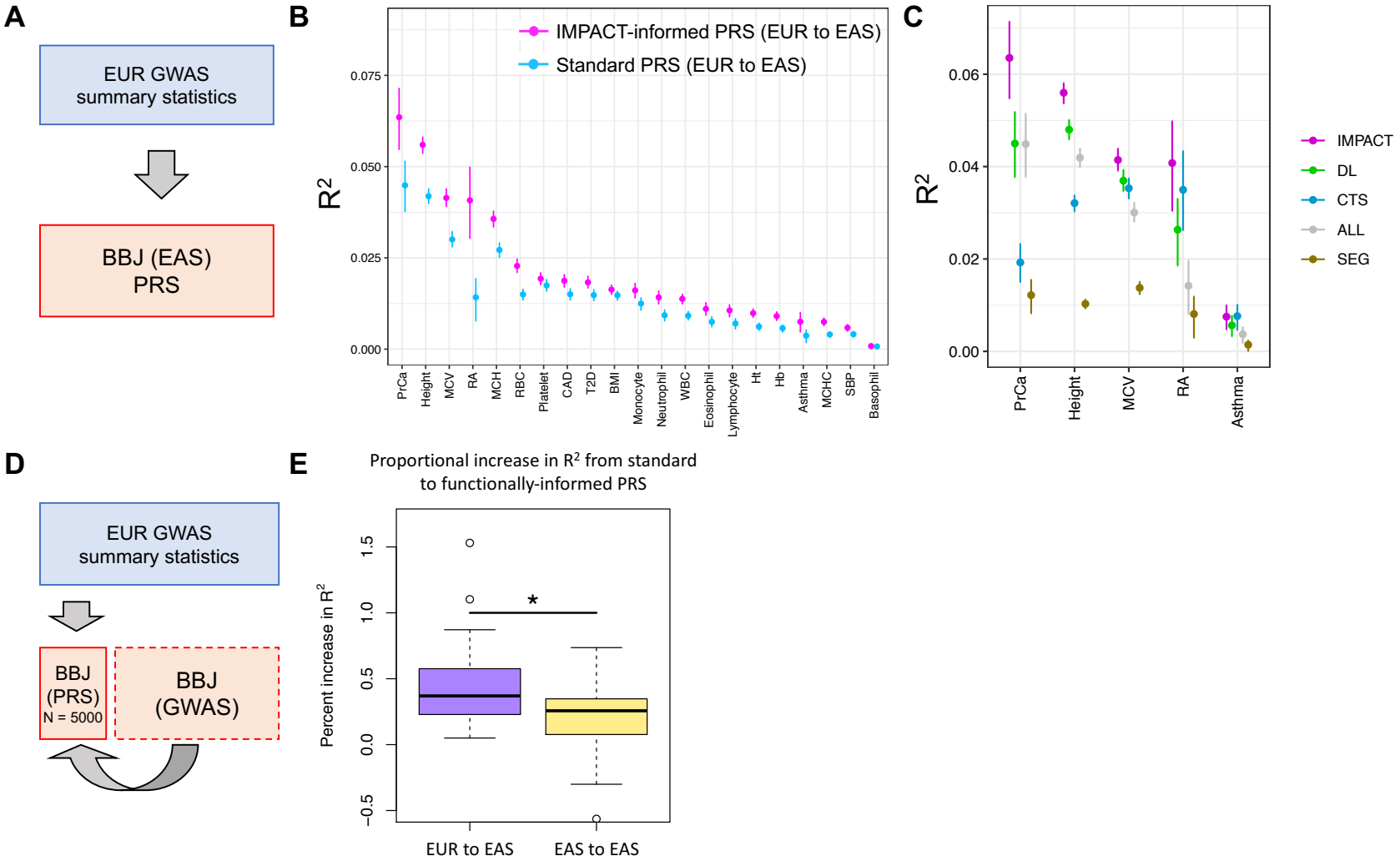


Figure 5 legend. Identifying shared regulatory variants with IMPACT annotations to improve the trans-ethnic portability of PRS. A) Study design applying EUR summary statistics-based PRS models to all individuals in the BBJ cohort. (B) Phenotypic variance (R^2) of BBJ individuals explained by EUR PRS using two methods: functionally-informed PRS with IMPACT (pink) and standard PRS (blue). Error bars indicate 95% CI calculated via 1,000 bootstraps. C) Phenotypic variance (R^2) of BBJ individuals across 5 exemplary traits explained by EUR IMPACT annotations relative to lead deep learning annotations (DL), cell-type-specific histone modification annotations (CTS), and lead cell-type-specifically expressed gene sets (SEG). Error bars indicate 95% CI calculated via 1,000 bootstraps. D) Study design to compare trans-ethnic (EUR to EAS) to within-population (EAS to EAS) improvement afforded by functionally-informed PRS models. For each trait, 5,000 randomly selected individuals from BBJ designated as PRS samples. Remaining BBJ individuals used for GWAS to derive EAS summary statistics-based PRS; no shared individuals between GWAS samples and PRS samples. E) Improvement from standard PRS to functionally-informed PRS compared between trans-ethnic (EUR to EAS) and within-population models (EAS to EAS) using the study design in D). In boxplots, center line indicates the median value; box limits indicate the upper (third) and lower (first) quartiles; the length of whiskers indicate values up to 1.5 times the interquartile range in either direction.

For each of 21 tested traits, we observed that functionally-informed PRS using IMPACT captured more phenotypic variance than standard PRS (49.9% mean relative increase in R^2, **Figure 5B**, **SF20**, **ST16-18**). The mean phenotypic variance explained across traits by functionally-informed PRS (R^2= 2.1%, sem = 0.4%) was greater than by standard PRS (R^2= 1.5%, sem = 0.3%, one-tailed paired wilcoxon $P < 4.8e-7$). For 19 of 21 traits, IMPACT-informed PRS significantly outperformed standard PRS (19 one-tailed difference of means $P < 0.05$); for platelet count $P = 0.052$ and for basophil count $P = 0.40$. Using 10,000 bootstraps of the PRS sample cohort, we found that the IMPACT-informed PRS R^2 estimate was consistently greater than the standard PRS estimate for all traits except basophil count (all bootstrap $P < 0.004$, **ST18**). Intriguingly, we found a strong correlation between the IMPACT-informed PRS R^2 estimate and the EAS heritability captured by the top 5% of SNPs according to the lead EUR IMPACT annotation (Pearson $r = 0.60$, $P = 0.004$, **ST19**). While EAS heritability metrics did not influence the choice of lead IMPACT annotation (EUR-based), this result is unsurprising given the strong multi-ethnic regulatory concordance we observed previously (**Figure 3C**) in which annotations that capture more heritability in EUR tend to capture more in EAS. Even though IMPACT-informed PRS models include between 7.5% and 79.1% of the total number of SNPs included in standard P+T models, the increased prediction R^2 indicates that prioritization of putatively functional variants over tagging variation compensates for the reduction of included loci. We observed the largest improvement for RA from R^2 = 1.4% (sd = 0.33%) in the standard PRS to R^2 = 4.1% (sd = 0.53%, one-tailed difference of means $P < 9.8e-6$) in the functionally-informed PRS using the B cell TBP IMPACT annotation. For asthma, R^2 = 0.37% (sd = 0.10%) in the standard PRS versus R^2 = 0.75% (sd = 0.14%, $P < 0.013$) in the functionally-informed PRS. For MCV, R^2 = 3.0% (sd = 0.10%) in the standard PRS versus R^2 = 4.1% (sd = 0.12%, $P < 1.2e-13$)

in the functionally-informed PRS. For PrCa, R^2 = 4.5% (sd = 0.36%) in the standard PRS versus R^2 = 6.4% (sd = 0.45%, *P* < 6.1e-4) in the functionally-informed PRS. For height, R^2 = 4.2% (sd = 0.10%) in the standard PRS versus R^2 = 5.6% (sd = 0.12%, *P* < 8.7e-20) in the functionally-informed PRS. We observed significantly greater PRS improvement among traits with lower estimated polygenicity (linear regression coefficient = -0.02, *P* < 0.006). As previously stated, more highly polygenic traits may be driven by multiple cell types, of which only one may be captured by the lead IMPACT annotation.

For our five representative traits asthma, RA, MCV, PrCa, and height, we further compared functionally-informed PRS$_{EUR}$ using IMPACT to models using 123 DeepSEA and Basenji deep learning annotations[54–56,63], 220 cell-type-specifically expressed genes (SEG)[34] and 513 cell-type-specific histone modification tracks (CTS)[6] (**Figure 5C**, **ST11, ST20, Online Methods**). To our knowledge, deep learning annotations have not previously been applied to improving PRS model performance. IMPACT explained greater phenotypic variance on average (mean R^2= 4.2%, sem = 1.0%) than the top deep learning annotations (3.2%, sem = 0.8%, one-tailed paired wilcoxon *P* = 0.03) and was a significant improvement for four of five traits (four one-tailed difference of means *P* < 0.006), while only trending higher for asthma (*P* = 0.13). IMPACT also explained greater phenotypic variance on average than SEG (0.9%, sem = 0.2%, one-tailed paired wilcoxon *P* = 0.03) and this difference was individually detected for each of five traits (all one-tailed difference of means *P* < 3.4e-6). This trend was not as strong when comparing IMPACT to CTS (R^2= 2.6%, sem = 0.5%, one-tailed paired wilcoxon *P* = 0.06), although this difference was individually detected for three of five traits (three one-tailed difference of means *P* < 1.1e-4). We performed a similar bootstrap analysis as above, yielding

similar results; for only RA and asthma did IMPACT-PRS not produce consistently greater R^2 estimates than CTS-PRS (**ST20**).

Functionally-informed PRS might to some extent compensate for population-specific LD differences between populations. Hence, we hypothesized that IMPACT-informed PRS would improve standard PRS moreso in the trans-ethnic prediction framework, in which EUR PRS models predict EAS phenotypes, than in a within-population framework, in which EAS PRS models predict EAS phenotypes. Here, we define within-population PRS as PRS_{EAS} and trans-ethnic PRS as PRS_{EUR} to avoid confusion. In order to directly compare PRS model improvements between PRS_{EAS} and PRS_{EUR}, we evaluated prediction accuracy on the same individuals. Briefly, we partitioned the BBJ cohort to reserve 5,000 individuals for PRS testing, derived GWAS summary statistics from the remaining individuals, and performed P+T PRS modeling and prediction as done above (**Figure 5D, SF21-23, ST21-22, Online Methods**). For functionally-informed PRS_{EAS}, we selected lead IMPACT annotations from S-LDSC results using GWAS summary statistics, as done above, on the partition of the BBJ cohort excluding the 5,000 PRS test individuals. We defined improvement as the percent increase in R^2 from standard to functionally-informed PRS; therefore, differences in PRS performance due to intrinsic factors, such as GWAS power or genotyping platform, cancel out. In both scenarios, we observed substantial positive improvements: averaged across the 21 traits in the trans-ethnic setting (mean percent increase in R^2= 47.3%, sem = 8.1%, one-tailed z test P < 2.7e-9) and in the within-population setting (mean percent increase in R^2= 20.9%, sem = 6.6%, one-tailed z test P < 7.5e-4). Indeed, this revealed a significantly greater improvement in the trans-ethnic application than in the within-population application across the 21 traits (one-tailed paired wilcoxon P < 0.012, **Figure 5E**). To ensure that the disease predictive power of our PRS models

was not driven by a few loci of large effect, we performed a block jackknife over the genome to establish confidence intervals around the R^2 estimates as well as the relative improvement of IMPACT PRS over standard P+T PRS R^2 estimates (**Online Methods, SF24**). We observed narrow intervals around the estimates; for functionally-informed PRS$_{EUR}$ and functionally-informed PRS$_{EAS}$, we observed the average 95% confidence interval around R^2 estimates to be 0.001 and around the relative R^2 improvement to be 0.11 in PRS$_{EUR}$ and 0.07 in PRS$_{EAS}$. These results suggest that the disease predictive power of IMPACT-informed P+T models are not driven by a few loci of large effect. Moreover, our results for case/control diseases are not affected by estimating marginal effect sizes on the logistic scale, rather than the liability scale[64] (**Online Methods, SF25, SF26, Supplement**).

Overall, our results reveal that functional prioritization of SNPs using IMPACT improves both trans-ethnic and within-population PRS models, but is especially advantageous for the trans-ethnic application of PRS. We believe there are at least three important mechanisms at play leading to this improvement. First, restricting P+T PRS to variants that are more likely to be functional increases the likelihood of selecting a causal variant with disease predictive power in the target population. Previous studies support that the identification of causal variants can improve PRS accuracy[3,5,65]. Second, as shown in **Figure 3B**, the per-SNP heritability captured by IMPACT annotations tends to be similar in EUR and EAS populations, thereby ensuring that IMPACT-informed SNP prioritization schemes using EUR data are still effective in EAS. Third, as shown in **Figure 4C**, SNPs prioritized by IMPACT have more consistent multi-ethnic marginal effect sizes, which means that these SNPs are less likely to be solely associated by linkage and therefore might improve performance. In conclusion, our results nominate the prioritization of SNPs according to functional annotations, especially using IMPACT, as a potential tentative

533 solution for the lack of trans-ethnic portability of PRS models. While individuals of European
534 ancestry dominate current genetic studies, population-nonspecific cell-type-specific IMPACT
535 annotations can help transfer highly powered EUR genetic data to study still underserved
536 populations.
537
538 **Discussion**
539
540 In this study, we created a compendium of 707 cell-type-specific regulatory annotations
541 (**Web Resources**) capturing disproportionately large amounts of polygenic heritability in 95
542 complex traits and diseases in EUR and EAS populations. We then proposed a three-step
543 framework to assess how well prioritization of regulatory variants with functional data can
544 improve multi-ethnic genetic comparisons. First, we showed that heritability-enriched
545 regulatory elements between EUR and EAS populations capture indistinguishable proportions
546 of heritability across 29 complex traits. Second, we showed that functional prioritization of
547 variants selects those with more highly correlated marginal effect sizes between populations,
548 while negligibly affecting the distribution of F_{st}; this might explain the improvement driven by
549 functional prioritization in P+T PRS models which use marginal effect sizes. Third, we showed
550 that variant prioritization with IMPACT annotations results in consistently improved PRS
551 prediction accuracy, especially for the trans-ethnic application; potentially due to overcoming
552 large population-specific influences such as LD which is an important challenge of multi-
553 population models.
554 Designing genetic models for each complex trait or disease that capture risk for the full
555 diversity of the human population will be challenging. This necessitates approaches that
556 effectively transfer predictive genetic information from well studied populations to less well
557 studied populations. Without such approaches, the potential clinical benefits of PRS risk to

preferentially benefit populations with larger training GWAS datasets, e.g. European populations. As it will ultimately be useful to develop PRS scores that can be applied widely to many populations and admixed individuals[66,67], IMPACT may have the potential to be a tool that can prioritize key variants for this purpose. We argue for the use of biologically diverse IMPACT annotations to capture relevant genetic signal and compensate, to some extent, for differences in LD across populations. To begin to address this, we investigated PRS using EUR summary statistics and genotyping data from five populations (AFR, AMR, EAS, EUR, and SAS) in 1000 Genomes and found that IMPACT-informed PRS moderately reduces the inter-population variation of PRS values compared to standard P+T (one-tailed paired wilcoxon $P = 0.003$, 52.0% reduction in mean F-statistic for EUR PRS (**SF27**) and one-tailed paired wilcoxon $P = 0.002$, 64.6% reduction in mean F-statistic for EAS PRS (**SF28**)), suggesting functional prioritization can stabilize PRS values (**Online Methods**). However, other challenges such as differences in allele frequencies will need to be addressed in future studies.

Our work and that of others advocate for larger genetic studies in understudied populations[3] and the use of orthogonal LD-independent functional data to improve the disease predictive power of genetic models in such populations, as even increasing GWAS power cannot mitigate the bias introduced by LD. Our study should not in any way be interpreted as a justification for reducing the emphasis on the need for diversity in human genetic studies. A future which offers high powered GWAS in understudied populations will transform the study of trans-ethnic portability from an issue of EUR-biased health disparities to a question of population-specific genetic and environmental effects.

Our work provides insight into the potential clinical implementation of PRS and broader genetic applications that aim to integrate multi-ethnic data. This study suggests that functional

data may be leveraged to improve portability of genetic models; however, the issue of portability need not be restricted to two different continental populations as shown in this study, but rather will be relevant to any PRS model in which the target individual is not perfectly matched to the ancestry of the training population. While we did not assess a PRS model using meta-analyzed summary statistics from two or more populations in this study, we believe that this approach could be effective in identifying shared regulatory variants, especially for populations with limited GWAS sample size.

We believe that IMPACT may prioritize phenotype-driving regulatory variation. We have shown IMPACT to be more effective at capturing genetic variation of complex traits than commonly used functional annotations such as experimentally-derived cell-type-specific histone marks, gene sets, and deep learning regulatory annotations. We hypothesize the utility of IMPACT comes from 1) cell-type-specificity of TF binding models which locate key classes of regulatory elements and 2) the integration of thousands of experimentally-derived annotations, which presumably removes noise and enriches for biological signal present in each individual annotation. Here, we did not demonstrate the potential utility of IMPACT to perform functional fine-mapping to reduce credible sets beyond our previous work[30], due to lack of sufficient gold standards with causal experimental validation and the limitation to genome-wide significant variants. The specific application of IMPACT in multi-ethnic fine-mapping needs to be further investigated.

We must consider several important limitations of our work. First, our functional insights are limited by biases in publicly available TF ChIP-seq data, as IMPACT cannot evaluate TF-cell type pairs for which training data does not exist. These biases include preference toward workhorse cell lines over primary cells or cell types that are rarer or more difficult to assay.

Furthermore, these biases include preference toward TFs with evidence of cell type expression and regulation, specific antibodies, and known sequence motifs for compatibility with IMPACT. These biases directly affect our ability to capture trait-relevant biology, leading to systematically better heritability enrichment for autoimmune diseases and hematological traits for which the relevant cell type is easier to assay, e.g. blood, and worse enrichment for brain-related traits for which the relevant tissue is difficult to assay. Future work may be needed to adapt the IMPACT framework to model the epigenetic signatures of functional marks beyond TF binding to capture a broader array of trait-relevant biological processes. In the future, the cell-type-specific functional training data for IMPACT may be replaced by newer experimental strategies to map enhancers. For example, high-throughput CRISPR screens paired with assays for open chromatin could be used to precisely redefine regulatory landscapes. Second, we used multi-ethnic data to argue for the utility of our approach. However, the robustness of multi-ethnic comparisons for a given phenotype rely on properties surrounding the recruitment of individuals or the exact genotyping platform used in various biobanks, which may result in cohort-bias that inflates within-population PRS prediction accuracy. For example, BBJ is a disease ascertainment cohort, in which each individual has any one of 47 common diseases[68,69]; therefore, BBJ control samples are not comparable to healthy controls of UKBB. Other biases may arise from clinical differences in phenotyping. Also, we only considered a single non-EUR population in this study, although the disparity in trans-ethnic portability, and hence resulting benefit from functional annotations, may be greater in other non-EUR populations. Therefore, the results presented here may only be used to interpret the improved portability of genetic data between EUR and EAS populations. Further work is required to assess potential improvements in portability between EUR and other populations.

In conclusion, we demonstrated that IMPACT annotations improve the comparison of genetic data between populations and trans-ethnic portability of PRS models using ancestrally unmatched data. While a long-term goal of the field must be to diversify GWAS and other genetic studies in non-European populations, it is imperative that genetic models be developed that work in multiple populations. Such initiatives will necessitate the use of population-independent functional annotations, such as IMPACT, in order to capture shared biological mechanisms regulated by complex genetic variation.

Supplemental Data

See *Amariuta_Ishigaki_Supplement_Revised2NG.pdf* and *Amariuta_Ishigaki_Supplementary_Tables_Revised2NG.xlsx*

Online Methods

See *Amariuta_Ishigaki_Online_Methods_Revised2NG.pdf*

Acknowledgements

This work is supported in part by funding from the National Institutes of Health (NHGRI T32 HG002295, UH2AR067677, 1U01HG009088, U01 HG009379, and 1R01AR063759) and the Doris Duke Charitable Foundation grant #2013097.

Declaration of Interests

The authors declare no competing financial interests.

Data Availability

We provide IMPACT-707 annotations at https://github.com/immunogenomics/IMPACT/tree/master/IMPACT707

Code Availability

We provide code to recreate our analyses at https://github.com/immunogenomics/IMPACT/tree/master/IMPACT707/AnalysisCode

Web Resources

1. IMPACT Github repository: https://github.com/immunogenomics/IMPACT
2. IMPACT 707 annotations: https://github.com/immunogenomics/IMPACT/tree/master/IMPACT707

3. Analysis code: https://github.com/immunogenomics/IMPACT/tree/master/IMPACT707/AnalysisCode
4. HOMER: http://homer.ucsd.edu/homer/motif/
5. S-LDSC: https://github.com/bulik/ldsc
6. 1000 Genomes: http://www.internationalgenome.org/
7. Cell-type-specifically expressed gene set annotations and LD scores: https://data.broadinstitute.org/alkesgroup/LDSCORE/LDSC_SEG_ldscores/
8. Cell-type-specific histone modification ChIP-seq datasets: https://data.broadinstitute.org/alkesgroup/LDSCORE/
9. Plink: https://www.cog-genomics.org/plink2
10. Riken website: http://jenger.riken.jp/en/
11. Price Lab: https://data.broadinstitute.org/alkesgroup/sumstats_formatted/
12. Neale Lab: http://www.nealelab.is/uk-biobank
13. GWAS Catalog: https://www.ebi.ac.uk/gwas/
14. Deep Learning: https://data.broadinstitute.org/alkesgroup/LDSCORE/DeepLearning/

References

1. Kichaev, G. & Pasaniuc, B. Leveraging Functional-Annotation Data in Trans-ethnic Fine-Mapping Studies. *Am. J. Hum. Genet.* **97**, 260–271 (2015).
2. Lam, M. *et al.* Comparative genetic architectures of schizophrenia in East Asian and European populations. doi:10.1101/445874
3. Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).
4. Sirugo, G., Williams, S. M. & Tishkoff, S. A. The Missing Diversity in Human Genetic Studies. *Cell* **177**, 26–31 (2019).
5. Vilhjálmsson, B. J. *et al.* Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
6. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).

7. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
8. International Schizophrenia Consortium *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
9. Chatterjee, N., Shi, J. & García-Closas, M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat. Rev. Genet.* **17**, 392–406 (2016).
10. Stahl, E. A. *et al.* Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat. Genet.* **44**, 483–489 (2012).
11. Chatterjee, N. *et al.* Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. *Nat. Genet.* **45**, 400–5, 405e1–3 (2013).
12. Khera, A. V. *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219–1224 (2018).
13. Schumacher, F. R. *et al.* Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat. Genet.* **50**, 928–936 (2018).
14. Sharp, S. A. *et al.* Development and Standardization of an Improved Type 1 Diabetes Genetic Risk Score for Use in Newborn Screening and Incident Diagnosis. *Diabetes Care* **42**, 200–207 (2019).
15. Kullo, I. J. *et al.* Incorporating a Genetic Risk Score Into Coronary Heart Disease Risk Estimates: Effect on Low-Density Lipoprotein Cholesterol Levels (the MI-GENES Clinical Trial). *Circulation* **133**, 1181–1188 (2016).
16. Natarajan, P. *et al.* Polygenic Risk Score Identifies Subgroup With Higher Burden of Atherosclerosis and Greater Relative Benefit From Statin Therapy in the Primary Prevention Setting. *Circulation* **135**, 2091–2101 (2017).

17. Márquez-Luna, C., Loh, P.-R., South Asian Type 2 Diabetes (SAT2D) Consortium, SIGMA Type 2 Diabetes Consortium & Price, A. L. Multiethnic polygenic risk scores improve risk prediction in diverse populations. *Genet. Epidemiol.* **41**, 811–823 (2017).
18. Duncan, L. *et al.* Analysis of polygenic risk score usage and performance in diverse human populations. *Nat. Commun.* **10**, 3328 (2019).
19. Curtis, D. Polygenic risk score for schizophrenia is more strongly associated with ancestry than with schizophrenia. *Psychiatr. Genet.* **28**, 85–89 (2018).
20. Martin, A. R. *et al.* Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *Am. J. Hum. Genet.* **100**, 635–649 (2017).
21. Hu, Y. *et al.* Leveraging functional annotations in genetic risk prediction for human complex diseases. *PLoS Comput. Biol.* **13**, e1005589 (2017).
22. Márquez-Luna, C. *et al.* Modeling functional enrichment improves polygenic prediction accuracy in UK Biobank and 23andMe data sets. *bioRxiv* 375337 (2018). doi:10.1101/375337
23. Okada, Y. *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376–381 (2014).
24. Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nature Genetics* **50**, 390–400 (2018).
25. Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018).
26. Schaub, M. A., Boyle, A. P., Kundaje, A., Batzoglou, S. & Snyder, M. Linking disease associations with regulatory information in the human genome. *Genome Res.* **22**, 1748–

1759 (2012).

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

28. Reshef, Y. A. *et al.* Detecting genome-wide directional effects of transcription factor binding on polygenic disease risk. *Nat. Genet.* **50**, 1483–1493 (2018).

29. Liu, X., Li, Y. I. & Pritchard, J. K. Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. *Cell* **177**, 1022–1034.e6 (2019).

30. Amariuta, T. *et al.* IMPACT: Genomic Annotation of Cell-State-Specific Regulatory Elements Inferred from the Epigenome of Bound Transcription Factors. *Am. J. Hum. Genet.* **104**, 879–895 (2019).

31. Kawakami, E., Nakaoka, S., Ohta, T. & Kitano, H. Weighted enrichment method for prediction of transcription regulators from transcriptome and global chromatin immunoprecipitation data. *Nucleic Acids Res.* **44**, 5010–5021 (2016).

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

33. Roadmap Epigenomics, C. *et al.* Heravi-428 Moussavi A, Kheradpour P, Zhang Z, Wang J, et al. Integrative analysis of 111 reference human 429 epigenomes. *Nature* **518**, 317–330 (2015).

34. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).

35. Gazal, S. *et al.* Linkage disequilibrium–dependent architecture of human complex traits shows action of negative selection. *Nat. Genet.* **49**, 1421–1427 (2017).

36. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association

761 studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012
762 (2019).
763 37. Akiyama, M. *et al.* Characterizing rare and low-frequency height-associated variants in the
764 Japanese population. *Nat. Commun.* **10**, 4393 (2019).
765 38. Ishigaki, K., Akiyama, M., Kanai, M. & Takahashi, A. Large scale genome-wide association
766 study in a Japanese population identified 45 novel susceptibility loci for 22 diseases.
767 *bioRxiv* (2019).
768 39. Akiyama, M. *et al.* Genome-wide association study identifies 112 new loci for body mass
769 index in the Japanese population. *Nat. Genet.* **49**, 1458–1467 (2017).
770 40. Peterson, R. E. *et al.* Genome-wide Association Studies in Ancestrally Diverse Populations:
771 Opportunities, Methods, Pitfalls, and Recommendations. *Cell* **179**, 589–603 (2019).
772 41. Gurdasani, D., Barroso, I., Zeggini, E. & Sandhu, M. S. Genomics of disease risk in globally
773 diverse populations. *Nat. Rev. Genet.* **20**, 520–535 (2019).
774 42. Brown, B. C., Ye, C. J., Price, A. L. & Zaitlen, N. Transethnic Genetic-Correlation Estimates
775 from Summary Statistics. *Am. J. Hum. Genet.* **99**, 76–88 (2016).
776 43. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human
777 height. *Nat. Genet.* **42**, 565–569 (2010).
778 44. Yang, J., Zeng, J., Goddard, M. E., Wray, N. R. & Visscher, P. M. Concepts, estimation and
779 interpretation of SNP-based heritability. *Nat. Genet.* **49**, 1304–1310 (2017).
780 45. Drake, L. Y. *et al.* B cells play key roles in th2-type airway immune responses in mice
781 exposed to natural airborne allergens. *PLoS One* **10**, e0121660 (2015).
782 46. Amariuta, T., Luo, Y., Knevel, R., Okada, Y. & Raychaudhuri, S. Advances in genetics toward
783 identifying pathogenic cell states of rheumatoid arthritis. *Immunol. Rev.* (2019).

784 doi:10.1111/imr.12827
785 47. Buttari, B., Profumo, E. & Riganò, R. Crosstalk between red blood cells and the immune
786 system and its impact on atherosclerosis. *Biomed Res. Int.* **2015**, 616834 (2015).
787 48. Anderson, H. L., Brodsky, I. E. & Mangalmurti, N. S. The Evolving Erythrocyte: Red Blood
788 Cells as Modulators of Innate Immunity. *J. Immunol.* **201**, 1343–1351 (2018).
789 49. Lui, J. C. & Baron, J. Mechanisms limiting body growth in mammals. *Endocr. Rev.* **32**, 422–
790 440 (2011).
791 50. Maier, A. B., van Heemst, D. & Westendorp, R. G. J. Relation between body height and
792 replicative capacity of human fibroblasts in nonagenarians. *J. Gerontol. A Biol. Sci. Med.*
793 *Sci.* **63**, 43–45 (2008).
794 51. Murphy, R. A. *et al.* Adipose tissue, muscle, and function: potential mediators of
795 associations between body weight and mortality in older adults with type 2 diabetes.
796 *Diabetes Care* **37**, 3213–3219 (2014).
797 52. Heymsfield, S. B., Gallagher, D., Mayer, L., Beetsch, J. & Pietrobelli, A. Scaling of human
798 body composition to stature: new insights into body mass index. *Am. J. Clin. Nutr.* **86**, 82–
799 91 (2007).
800 53. Kichaev, G. *et al.* Leveraging Polygenic Functional Enrichment to Improve GWAS Power.
801 *Am. J. Hum. Genet.* **104**, 65–75 (2019).
802 54. Kelley, D. R. *et al.* Sequential regulatory activity prediction across chromosomes with
803 convolutional neural networks. *Genome Res.* **28**, 739–750 (2018).
804 55. Zhou, J. & Troyanskaya, O. G. Predicting effects of noncoding variants with deep learning-
805 based sequence model. *Nat. Methods* **12**, 931–934 (2015).
806 56. Dey, K. K., Van de Geijn, B., Kim, S. S. & Hormozdiari, F. Evaluating the informativeness of

807 deep learning annotations for human complex diseases. *bioRxiv* (2019).
808 57. Galinsky, K. J. *et al.* Estimating cross-population genetic correlations of causal effect sizes.
809 *Genet. Epidemiol.* **43**, 180–188 (2019).
810 58. Shi, H. *et al.* Population-specific causal disease effect sizes in functionally important
811 regions impacted by selection. *bioRxiv* 803452 (2019). doi:10.1101/803452
812 59. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits.
813 *Nat. Genet.* **47**, 1236–1241 (2015).
814 60. Lu, Q. *et al.* A Powerful Approach to Estimating Annotation-Stratified Genetic Covariance
815 via GWAS Summary Statistics. *Am. J. Hum. Genet.* **101**, 939–964 (2017).
816 61. Gusev, A. *et al.* Atlas of prostate cancer heritability in European and African-American men
817 pinpoints tissue-specific regulation. *Nat. Commun.* **7**, 10979 (2016).
818 62. Gibbs, R. A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74
819 (2015).
820 63. Chen, K. M., Cofer, E. M., Zhou, J. & Troyanskaya, O. G. Selene: a PyTorch-based deep
821 learning library for sequence data. *Nature Methods* **16**, 315–318 (2019).
822 64. Gillett, A. C., Vassos, E. & Lewis, C. M. Transforming Summary Statistics from Logistic
823 Regression to the Liability Scale: Application to Genetic and Environmental Risk Scores.
824 *Hum. Hered.* **83**, 210–224 (2018).
825 65. Lloyd-Jones, L. R. *et al.* Improved polygenic prediction by Bayesian multiple regression on
826 summary statistics. *Nat. Commun.* **10**, 5086 (2019).
827 66. Bitarello, B. D. & Mathieson, I. Polygenic scores for height in admixed populations. *bioRxiv*
828 (2020).
829 67. Marnetto, D. *et al.* Ancestry deconvolution and partial polygenic score can improve

susceptibility predictions in recently admixed individuals. *Nat. Commun.* **11**, 1628 (2020).

68. Nagai, A. *et al.* Overview of the BioBank Japan Project: Study design and profile. *J. Epidemiol.* **27**, S2–S8 (2017).

69. Hirata, M. *et al.* Cross-sectional analysis of BioBank Japan clinical data: A large cohort of 200,000 patients with 47 common diseases. *J. Epidemiol.* **27**, S9–S21 (2017).