Supplementary material for "Leveraging fine-mapping and non-European training data to improve trans-ethnic polygenic risk scores"

Supplementary Table Captions

Supplementary Table 1: Detailed simulation results. For each combination of generative model parameters, ancestry, and trait we report **(Polygenicity):** trait polygenicity **(European h2):** Trait heritability in Europeans **(Trans-ethnic rg):** Trans-ethnic genetic correlation. **(Method):** The method name (see below). **(#Experiments):** Number of experiments performed. **(Ancestry):** Target ancestry. **(N):** Target ancestry sample size. **(Avg. h2):** Average h2 in the target ancestry. **(s.d. h2):** s.d. h2 in the target ancestry. **(Average R2):** Average R^2**. (R2 s.e.):** The standard errors of R^2. **(R2 (normalized vs. EUR)):** R^2 (normalized against the results of the same method in Europeans). **(R2 (normalized vs. EUR) s.e.):** Standard error of R^2 (normalized against the results of the same method in Europeans). **(R2 (normalized vs. BOLT-LMM-EUR)):** R^2 (normalized vs. BOLT-LMM in non-British Europeans). **(R2 (normalized vs. BOLT-LMM-EUR) s.e.):** The standard error of R^2 (normalized vs. BOLT-LMM in non-British Europeans). **(R2 (normalized vs. BOLT-LMM-EUR) P-value):** The p-value of the statistical test testing the null hypothesis that R^2 is the same as obtained by BOLT-LMM in non-British Europeans. **(R2 (normalized vs. BOLT-LMM)):** R^2 (normalized vs. BOLT-LMM in the target ancestry). **(R2 (normalized vs. BOLT-LMM) s.e.):** The standard error of R^2 (normalized vs. BOLT-LMM in the target ancestry). **(R2 (normalized vs. BOLT-LMM) P-value):** The p-value of the statistical test testing the null hypothesis that R^2 is the same as obtained by BOLT-LMM in the target ancestry. **(R2 diff vs. BOLT-LMM-EUR):** The difference between the R^2 obtained by the current method and BOLT-LMM in non-British Europeans. **(R2 diff vs. BOLT-LMM-EUR s.e.):** The standard error of the difference between the R^2 obtained by the current method and BOLT-LMM in non-British Europeans. **(R2 diff vs. BOLT-LMM-EUR P-value):** The p-value of the statistical test testing the null hypothesis that R^2 is the same as obtained by BOLT-LMM in non-British Europeans. **(R2 diff vs. BOLT-LMM):** The difference between the R^2 obtained by the current method and BOLT-LMM in the target ancestry. **(R2 diff vs. BOLT-LMM s.e.):** The standard error of the difference between the R^2 obtained by the current method and BOLT-LMM in the target ancestry. **(R2 diff vs. BOLT-LMM P-value):** The p-value of the statistical test testing the null hypothesis that R^2 is the same as obtained by BOLT-LMM in the target ancestry. The method names are as described in the main text, with the following rules. PolyPred-S is a linear combination of PolyFun-pred and SBayesR (instead of PolyFun-pred and BOLT-LMM). A suffix "-L1" indicates that PolyFun-pred was invoked assuming only one causal SNP per locus (and did not make use of LD information). A suffix "-N" indicates that the British training sample size was reduced from N=337K to some other number. A suffix "-Nmix" indicates that the target-ancestry training sample size used for estimating mixing weights was modified from 500 to some other number. A suffix "-1000G" indicates that the LD reference panel used European data from the 1000 Genomes project. A suffix "-UK10K" indicates that the LD reference panel used UK10K.

Supplementary Table 2: Detailed simulation runtime analysis. For each combination of method and generative model parameters we report (**Method**): method name. **(#Experiments**): number of experiments performed. (**Polygenicity**): trait polygenicity. (**European h2**): h2 in Europeans. (**Average run time (sec)**): average run time in seconds. (**SE (sec)**): the standard error of the runtime in seconds. (**Average run time (hr)**): average run time in hours. (**SE (hr)**): the standard error of the runtime in hours.

Supplementary Table 3: List of 49 diseases and complex traits. For each trait, we report its UK Biobank British sample size (used for training most methods), its UK Biobank British heritability estimate and its standard error (estimated using S-LDSC with the Baseline-LF v2.2.UKB model[1]), whether the trait was one of the 7 traits included in the meta-analysis, and whether the trait exists in Biobank Japan.

Supplementary Table 4: Detailed results of analyses using UKB British training individuals applied to other UKB populations, compared vs. BOLT-LMM. For each combination of method, ancestry and trait (including meta-analyzed traits) we report (**Method**): The method name (see below). (**N**): Test set sample size; (**R2**): The squared Pearson correlation coefficient between the PRS and the trait; (**R2 s.e.**): The standard error of R^2, computed via genomic block-jackknife; (**R2 (normalized vs. BOLT-LMM-EUR)**): The R^2 value, divided by the R^2 of BOLT-LMM in non-British Europeans; (**R2 (normalized vs. BOLT-LMM-EUR) s.e.**): The standard error of the normalized R^2; (**R2 diff vs. BOLT-LMM**): The difference between R^2 and the R^2 obtained by BOLT-LMM; (**R2 diff vs. BOLT-LMM s.e.**) The standard error of the difference between R^2 and the R^2 obtained by BOLT-LMM, computed via genomic block-jackknife; (**R2 diff vs. BOLT-LMM (normalized vs. BOLT-LMM-EUR)**): The difference between normalized R^2 and the normalized R^2 obtained by BOLT-LMM; (**R2 diff vs. BOLT-LMM (normalized vs. BOLT-LMM-EUR) s.e.**): The standard error of the difference between normalized R^2 and the normalized R^2 obtained by BOLT-LMM; (**R2 vs. BOLT-LMM (normalized vs. BOLT-LMM-EUR) P-value**): The P-value of the difference between the normalized R^2 and the normalized R^2 obtained by BOLT-LMM; (**Regression slope**): Slope obtained when regressing the true phenotype on the PRS; (**Regression slope s.e.**): The standard error of the regression slope, computed via genomic block-jackknife; (**R2 ind-s.e.**): The standard error of R^2, computed via jackknife over individuals; (**Mixing weights**): The mixing weights of combined methods (blank for non-combined methods). The first value is the intercept, and the other values are PolyPred, BOLT-LMM, and BOLT-LMM-BBJ (when there are four numbers). The method names that are not explicitly defined in the main text are the following: **Methods ending with -pX** (where X is a number) are methods using a fixed mixing weight X for PolyPred; **Methods ending with with -100** use 100 individuals from the target cohort to estimate mixing weights (instead of 500 as used by most combined methods); **BOLT-LMM-727K**: BOLT-LMM using only genotyped SNPs; **LDpred-1000G-p**: LDpred using the 1000 genomes as an LD reference panel, and assuming that proportion p of causal SNPs are causal; **LDpred-1000G-cheat**: LDpred using the 1000 genomes as an LD reference panel, and using the best value of p for each trait (as determined via R^2 in the test set); **LDpred-UK10K-p**: LDpred using the UK10K cohort as an LD reference panel, and assuming that proportion p of causal SNPs are causal; **LDpred-UK10K-cheat**: LDpred using the UK10 cohort as an LD reference panel, and using the best value of p for each trait (as determined via R^2 in the test set); **PRS-CS-phi0.0001**: PRS-CS with –phi=0.0001; **PRS-CS-phi0.01**: PRS-CS with –phi=0.01; **PRS-CS-auto**: PRS-CS without specifying –phi; **PolyFun-pred-pipP**: PolyFun-pred restricted to SNPs with PIP greater than P; **PolyFun-pred-NoFun**: PolyFun-pred without using functional annotations; **P+T-pX**: P+T that uses only SNPs with BOLT-LMM P-value <X; **P+T-cheat**: P+T that uses the best value of X for each target ancestry. **PolyPred+-Ext**: PolyPred+ with mixing weights estimated in Biobank Japan; **PolyPred-pipP**: PolyPred restricted to SNPs with PIP greater than P; **PolyPred-NoFun**: PolyPred without using functional annotations; **SBayesR-2.8M**: SBayesR using 2.8M common SNPs selected by the SBayesR authors. We caution that standard errors of methods using only PIP>0.95 SNPs may not be accurate because of the small number of SNPs used. To see the numerical results of the analyses reported in the main text, please filter the **Trait** column to see only the trait 'Meta-Analysis'.

Supplementary Table 5: Detailed results of analyses using UKB British training individuals applied to other UKB populations, compared vs. PolyPred. The table is similar to Table 4, but all results are normalized and compared with respect to PolyPred instead of BOLT-LMM.

Supplementary Table 6: Ancestry-specific SNP heritability estimates in the UK Biobank, across 7 independent complex traits. For each trait (including meta-analyzed traits) we report its sample size (n), its SNP heritability estimate (h2g) and its standard error (se). All estimates were performed using GCTA[2] with HapMap 3[3] SNPs due to the relatively small sample sizes. Non-British Europeans were down-sampled to 10,000 individuals to facilitate the analysis. Meta-analyzed h2g was computed via the average h2g, and the meta-analyzed standard error was computed via the square root of the average sampling variance, divided by the square root of the number of traits.

Supplementary Table 7: Detailed results of analyses applied to Biobank Japan and to Uganda-APCDR. The table is similar to Table 4, but includes columns comparing each method to PolyPred in addition to columns comparing each method to BOLT-LMM.

Supplementary Table 8: Comparing prediction accuracy in UK Biobank Non-British Europeans and in Biobank Japan when using equal training set sample sizes. For each of 7 independent traits we report (**N**) its Biobank Japan training sample size (which was also used for the UK Biobank British training sample size in this analysis); (**h2g (UKB-EUR)**) its non-British European SNP heritability, as estimated by BOLT-REML; (**h2g (BBJ)**) its Biobank Japan SNP heritability, as estimated by BOLT-REML; (**R2-expected (UKB EUR)**) the expected R^2 in non-British Europeans as a function of training set sample size and SNP heritability, based on theory (see Supplementary Note); (**R2-expected (Biobank Japan)**) the expected R^2 in Biobank Japan as a function of training set sample size and SNP heritability, based on theory; (**R2 (UKB-EUR)**) the R^2 obtained in practice in non-British Europeans when training BOLT-LMM using a UK Biobank British training sample with the same sample size as the Biobank Japan training sample size; (**R2 (BBJ)**) the R^2 obtained in practice in 5K Biobank Japan individuals when training BOLT-LMM using a Biobank Japan training sample.

Supplementary Note

Causal vs. tagging effects

We consider a linear model $y = \sum_i x_i \beta_i + \epsilon$ where y is a trait, x_i is the number of minor alleles at SNP i, β_i is the (true) causal effect size of SNP i, and ϵ is a residual term sampled from a normal distribution. We consider a method (such as PolyFun-pred) that estimates β_i. If the generative model holds and all SNPs i are considered in the estimation procedure, then the estimated value $\hat{\beta}_i$ is a consistent estimator of β_i, and thus $\hat{\beta}_i$ represents a causal effect. In contrast, if only a subset of the SNPs, such as HapMap3 SNPs, are considered in the estimation procedure (i.e. if we incorrectly assume the generative model $y = \sum_{i \in S} x_i \beta_i + \epsilon$, where S is a subset of SNPs) then the estimated value $\hat{\beta}_i$ represents a linear combination of β_i and of the effect sizes of other SNPs.

The exact value estimated by $\hat{\beta}_i$ depends on the estimation procedure. For example, assuming an ordinary least squares estimator for simplicity, the vector $\widehat{\boldsymbol{\beta}}_S$ of estimated coefficients is a consistent estimator of $\left[I_{m-k}\ R_{S\bar{S}}^{-1} R_{S\bar{S}}\right]\boldsymbol{\beta}$, where m is the total number of SNPs, k is the number of SNPs in the set S, R_{SS} is the LD matrix of the SNPs in the set S, $R_{S\bar{S}}$ is a matrix wherein each entry i, j is the correlation between SNP i in the set S and SNP j in the set of SNPs that are not in S, and $\boldsymbol{\beta}$ is the vector of true effect sizes, assuming without loss of generality that the set S includes the first k SNPs (out of m SNPs considered). It is easy to derive this quantity by writing down the conditional expectation of $\widehat{\boldsymbol{\beta}}_S$ under an ordinary least squares estimator, given by $\mathrm{E}\left[\widehat{\boldsymbol{\beta}}_S \mid \boldsymbol{\beta}\right] = \mathrm{E}[(\boldsymbol{X}_s^T \boldsymbol{X}_s)^{-1} \boldsymbol{X}_s^T \boldsymbol{y} \mid \boldsymbol{\beta}]$, where $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ is a vector of observed phenotypes and $\boldsymbol{X}$ is the corresponding matrix of SNPs, $\boldsymbol{X}_s$ is the submatrix of $\mathbf{X}$ consisting of columns of SNPs in the set S, and we assume that ϵ is independent of $\boldsymbol{X}$.

Investigating if off-cohort loss of accuracy is driven by SNP heritability differences

We investigated if lower prediction accuracies in Biobank Japan vs. the UK Biobank can be largely explained by SNP heritability differences. We began by comparing trait heritabilities across the UK Biobank and Biobank Japan, using BOLT-REML[4] applied to UK Biobank British-ancestry individuals (average *N*=325K) and to Biobank Japan (average *N*=124K), restricting to HapMap 3 SNPs. The average heritability in the UK Biobank was 67% larger (Supplementary Table 8), indicating differences in either trait measurement, cohort ascertainment, the ability of HapMap 3 SNPs to tag East Asian causal SNPs[5], or in the true underlying heritabilities (we could not perform a similar analysis with UK Biobank East Asian individuals due to small sample sizes leading to large standard errors). We next asked if the observed differences in PRS accuracy between Biobank Japan and the UK Biobank can be explained by the 67% increased average SNP heritability in the UK Biobank. To this end, we computed the expected R^2 within each cohort as function of SNP heritability, sample size, and the effective number of independent SNPs[6,7]:

$$\mathrm{E}[R^2] = h^2 \frac{h^2}{h^2 + \frac{m}{n}}.$$

Here, h^2 is SNP heritability, n is sample size, and m is the effective number of independent SNPs (which we specified as 55,000, determined by dividing the number of HapMap 3 SNPs by their average within-HapMap 3 LD-score). We used the smaller Biobank Japan sample size in both cohorts to eliminate differences due to sample size differences (by choosing a random subset of UK Biobank British individuals as a training set). The average expected R^2 in the UK Biobank was 104% larger than in Biobank Japan (Supplementary Table 8). We then trained BOLT-LMM using subsets of the UK Biobank British sample (matching the Biobank Japan sample size for each trait) and applied the predictions to UK Biobank non-British Europeans. The average R^2 in UK Biobank non-British Europeans (when training BOLT-LMM using the reduced British training sample) was 108% larger than the average R^2 in Biobank Japan (when training BOLT-LMM using the Biobank Japan training sample) (Supplementary Table 8), strongly consistent with the 104% increase expected from theory. Finally, we determined that when training BOLT-LMM using the full UK Biobank British training set (average *N*=325K), the average R^2 in UK Biobank East Asians across the 7 independent traits is 93% larger than in Biobank Japan (Supplementary Tables 4 and 7), broadly consistent with the previous results. Assuming that the main factor differentiating the UK Biobank East Asian sample from the Biobank Japan sample is SNP heritability differences (rather than differences in MAF, LD, or causal effect sizes), these findings suggest that the main factor leading to lower prediction accuracies in Biobank Japan vs. the UK Biobank is SNP heritability differences.

To further investigate if off-cohort loss of accuracy is driven by SNP heritability differences, we compared prediction accuracies in UK Biobank East Asians and in Biobank Japan, when training BOLT-LMM using the

Biobank Japan training sample. The average relative-R^2 in UK Biobank East Asians across the 7 independent traits was 9.0% larger (Supplementary Tables 4,7), though the difference was not statistically significant (P=0.18), possibly owing to the small UK Biobank East Asian sample size.

Although these results are not conclusive, they suggest that heritability differences drive most of the differences in prediction accuracies observed between the UK Biobank and Biobank Japan. Surprisingly, these results are consistent with a model in which HapMap 3 SNPs in Biobank Japan tag approximately 50% of the causal SNPs that they tag in the UK Biobank, rather than a model in which SNP heritabilities in Biobank Japan are smaller due to smaller causal effect sizes. This is because under the second model, we would expect to see large increase in prediction accuracy in UK Biobank East Asians vs. Biobank Japan when training BOLT-LMM using Biobank Japan (compared with only a 9.0% increase observed in practice). A partial explanation is that the HapMap 3 SNP set consists of a combination of two genotyping chips, one of which is explicitly designed to optimize tagging in Europeans[8].

References

1. Weissbrod, O. *et al.* Functionally informed fine-mapping and polygenic localization of complex trait heritability. *Nat. Genet.* 1–9 (2020).
2. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
3. HapMap3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52 (2010).
4. Loh, P.-R. *et al.* Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat. Genet.* **47**, 1385–1392 (2015).
5. Bhangale, T. R., Rieder, M. J. & Nickerson, D. A. Estimating coverage and power for genetic association studies using near-complete variation data. *Nat. Genet.* **40**, 841–843 (2008).
6. Daetwyler, H. D., Villanueva, B. & Woolliams, J. A. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PloS One* **3**, e3395 (2008).
7. Visscher, P. M. & Hill, W. G. The limits of individual identification from sample allele frequencies: theory and statistical analysis. *PLoS Genet* **5**, e1000628 (2009).
8. Duan, S., Zhang, W., Cox, N. J. & Dolan, M. E. FstSNP-HapMap3: a database of SNPs with high population differentiation for HapMap3. *Bioinformation* **3**, 139 (2008).