

 Open access • Posted Content • DOI:10.1101/375337

## **Modeling functional enrichment improves polygenic prediction accuracy in UK Biobank and 23andMe data sets — [Source link](#)**

Carla Marquez-Luna, Steven Gazal, Po-Ru Loh, Nicholas A. Furlotte ...+2 more authors

**Institutions:** Harvard University, Brigham and Women's Hospital

**Published on:** 24 Jul 2018 - bioRxiv (Cold Spring Harbor Laboratory)

Related papers:

- [LDpred-funct: incorporating functional priors improves polygenic prediction accuracy in UK Biobank and 23andMe data sets](#)
- [Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores](#)
- [Common polygenic variation contributes to risk of schizophrenia and bipolar disorder](#)
- [The UK Biobank resource with deep phenotyping and genomic data](#)
- [Second-generation PLINK: rising to the challenge of larger and richer datasets](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/modeling-functional-enrichment-improves-polygenic-prediction-4wwkogkcv6>

# 1 Modeling functional enrichment improves polygenic prediction 2 accuracy in UK Biobank and 23andMe data sets

3 Carla Márquez-Luna<sup>1,3,4,\*</sup>, Steven Gazal<sup>2,3</sup>, Po-Ru Loh<sup>2,3,5</sup>, Samuel S. Kim<sup>2,6</sup>, Nicholas Furlotte<sup>7</sup>,  
Adam Auton<sup>7</sup>, 23andMe Research Team<sup>7</sup>, Alkes L. Price<sup>1,2,3,\*</sup>

4 <sup>1</sup>Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA.

5 <sup>2</sup>Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA.

6 <sup>3</sup>Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA.

7 <sup>4</sup>Charles R. Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New  
8 York, NY, USA.

9 <sup>5</sup>Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School,  
10 Boston, MA, USA

11 <sup>6</sup>Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cam-  
12 bridge, MA, USA.

13 <sup>7</sup>23andMe Inc., Mountain View, CA, USA.

14 \* corresponding authors

## 15 **Abstract**

16 Genetic variants in functional regions of the genome are enriched for complex trait heritabil-  
17 ity. Here, we introduce a new method for polygenic prediction, LDpred-funct, that leverages  
18 trait-specific functional enrichments to increase prediction accuracy. We fit priors using the  
19 recently developed baseline-LD model, which includes coding, conserved, regulatory and LD-  
20 related annotations. We analytically estimate posterior mean causal effect sizes and then use  
21 cross-validation to regularize these estimates, improving prediction accuracy for sparse architec-  
22 tures. LDpred-funct attained higher prediction accuracy than other polygenic prediction methods  
23 in simulations using real genotypes. We applied LDpred-funct to predict 21 highly heritable traits  
24 in the UK Biobank. We used association statistics from British-ancestry samples as training data  
25 (avg  $N=365K$ ) and samples of other European ancestries as validation data (avg  $N=22K$ ), to  
26 minimize confounding. LDpred-funct attained a +9% relative improvement in average predic-  
27 tion accuracy (avg prediction  $R^2=0.145$ ; highest  $R^2=0.413$  for height) compared to LDpred (the  
28 best method that does not incorporate functional information), consistent with simulations. For  
29 height, meta-analyzing training data from UK Biobank and 23andMe cohorts (total  $N=1107K$ ;

30 higher heritability in UK Biobank cohort) increased prediction  $R^2$  to 0.429. Our results show  
31 that modeling functional enrichment improves polygenic prediction accuracy, consistent with the  
32 functional architecture of complex traits.

## 33 Introduction

34 Genetic variants in functional regions of the genome are enriched for complex trait heritability<sup>1-6</sup>.  
35 In this study, we aim to leverage functional enrichment to improve polygenic prediction<sup>7,8</sup>. Sev-  
36 eral studies have shown that incorporating prior distributions on causal effect sizes can improve  
37 prediction accuracy<sup>9-12</sup>, compared to standard Best Linear Unbiased Prediction (BLUP) or Prun-  
38 ing+Thresholding methods<sup>13-15</sup>. Recent efforts to incorporate functional information have produced  
39 promising results<sup>16,17</sup>, but may be limited by dichotomizing between functional and non-functional  
40 variants<sup>16</sup> or restricting their analyses to genotyped variants<sup>17</sup>.

41 Here, we introduce a new method, LDpred-funct, for leveraging trait-specific functional enrich-  
42 ments to increase polygenic prediction accuracy. We fit functional priors using our recently devel-  
43 oped baseline-LD model<sup>18</sup>, which includes coding, conserved, regulatory and LD-related annotations.  
44 LDpred-funct first analytically estimates posterior mean causal effect sizes, accounting for functional  
45 priors and LD between variants. LDpred-funct then uses cross-validation within validation samples  
46 to regularize causal effect size estimates in bins of different magnitude, improving prediction accuracy  
47 for sparse architectures. We show that LDpred-funct attains higher polygenic prediction accuracy  
48 than other methods in simulations with real genotypes, analyses of 21 highly heritable UK Biobank  
49 traits, and meta-analyses of height using training data from UK Biobank and 23andMe cohorts.

## 50 Methods

### 51 Polygenic prediction methods

52 We compared 5 main prediction methods: Pruning+Thresholding<sup>14,15</sup> (P+T), LDpred<sup>12</sup>, P+T with  
53 functionally informed LASSO shrinkage<sup>16</sup> (P+T-funct-LASSO), our new LDpred-funct-inf method,  
54 and our new LDpred-funct method; we also included LDpred-inf<sup>12</sup>, which is known to attain lower  
55 prediction accuracy than LDpred<sup>12</sup>, in some of our secondary analyses. P+T, LDpred-inf and LD-  
56 pred are polygenic prediction methods that do not use functional annotations. P+T-funct-LASSO  
57 is a modification of P+T that corrects marginal effect sizes for winner's curse, accounting for func-  
58 tional annotations. LDpred-funct-inf is an improvement of LDpred-inf that incorporates functionally  
59 informed priors on causal effect sizes. LDpred-funct is an improvement of LDpred-funct-inf that uses  
60 cross-validation to regularize posterior mean causal effect size estimates, improving prediction accu-  
61 racy for sparse architectures. Each method is described in greater detail below. In both simulations  
62 and analyses of real traits, we used squared correlation ( $R^2$ ) between predicted phenotype and true  
63 phenotype in a held-out set of samples as our primary measure of prediction accuracy.

64

65 **P+T.** The P+T method builds a polygenic risk score (PRS) using a subset of independent SNPs  
66 obtained via informed LD-pruning<sup>15</sup> (also known as LD-clumping) followed by P-value thresholding<sup>14</sup>.  
67 Specifically, the method has two parameters,  $R_{LD}^2$  and  $P_T$ , and proceeds as follows. First, the method  
68 prunes SNPs based on a pairwise threshold  $R_{LD}^2$ , removing the less significant SNP in each pair.  
69 Second, the method restricts to SNPs with an association P-value below the significance threshold  
70  $P_T$ . Letting  $M$  be the number of SNPs remaining after LD-clumping, polygenic risk scores (PRS)  
71 are computed as

$$PRS(P_T) = \sum_{i=1}^M \mathbf{1}_{\{P_i < P_T\}} \tilde{\beta}_i g_i, \quad (1)$$

72 where  $\tilde{\beta}_i$  are normalized marginal effect size estimates and  $g_i$  is a vector of normalized genotypes for  
73 SNP  $i$ . The parameters  $R_{LD}^2$  and  $P_T$  are commonly tuned using validation data to optimize predic-  
74 tion accuracy<sup>14,15</sup>. While in theory this procedure is susceptible to overfitting, in practice, validation  
75 sample sizes are typically large, and  $R_{LD}^2$  and  $P_T$  are selected from a small discrete set of parameter  
76 choices, so that overfitting is considered to have a negligible effect<sup>7,14,15,19</sup>. Accordingly, in this work,  
77 we consider  $R_{LD}^2 \in \{0.1, 0.2, 0.5, 0.8\}$  and  $P_T \in \{1, 0.3, 0.1, 0.03, 0.01, 0.003, 0.001, 3 * 10^{-4}, 10^{-4}, 3 * 10^{-5}, 10^{-5}, 10^{-6}, 10^{-7}, 10^{-8}\}$ , and we always report results corresponding to the best choices of these  
78 parameters. The P+T method is implemented in the PLINK software (see Web Resources).  
79

80  
81 **LDpred-inf.** The LDpred-inf method estimates posterior mean causal effect sizes under an  
82 infinitesimal model, accounting for LD<sup>12</sup>. The infinitesimal model assumes that normalized causal  
83 effect sizes have prior distribution  $\beta_i \sim N(0, \sigma^2)$ , where  $\sigma^2 = h_g^2/M$ ,  $h_g^2$  is the SNP-heritability, and  
84  $M$  is the number of SNPs. The posterior mean causal effect sizes are

$$E(\beta|\tilde{\beta}, \mathbf{D}) = \left( \frac{N}{1 - h_l^2} * \mathbf{D} + \frac{1}{\sigma^2} \mathbf{I} \right)^{-1} N * \tilde{\beta}, \quad (2)$$

85 where  $\mathbf{D}$  is the LD matrix between markers,  $\mathbf{I}$  is the identity matrix,  $N$  is the training sample size,  
86  $\tilde{\beta}$  is the vector of marginal association statistics, and  $h_l^2 \approx kh^2/M$  is the heritability of the  $k$  SNPs  
87 in the region of LD; following ref. 12 we use the approximation  $1 - h_l^2 \approx 1$ , which is appropriate  
88 when  $M \gg k$ .  $\mathbf{D}$  is typically estimated using validation data, restricting to non-overlapping LD  
89 windows. We used the default LD window size, which is  $M/3000$ .  $h_g^2$  can be estimated from raw  
90 genotype/phenotype data<sup>20,21</sup> (the approach that we use here; see below), or can be estimated from  
91 summary statistics using the aggregate estimator as described in ref. 12. To approximate the nor-  
92 malized marginal effect size ref. 12 uses the p-values to obtain absolute Z scores and then multiplies  
93 absolute Z scores by the sign of the estimated effect size. When sample sizes are very large, p-  
94 values may be rounded to zero, in which case we approximate normalized marginal effect sizes  $\tilde{\beta}_i$  by

95  $\widehat{b}_i \frac{\sqrt{2 * p_i * (1 - p_i)}}{\sqrt{\sigma_Y^2}}$ , where  $\widehat{b}_i$  is the per-allele marginal effect size estimate,  $p_i$  is the minor allele frequency  
 96 of SNP  $i$ , and  $\sigma_Y^2$  is the phenotypic variance in the training data. This applies to all the methods  
 97 that use normalized effect sizes. Although the published version of LDpred requires a matrix inver-  
 98 sion (Equation 2), we have implemented a computational speedup that computes the posterior mean  
 99 causal effect sizes by efficiently solving<sup>22</sup> the system of linear equations  $(\frac{1}{\sigma^2} \mathbf{I} + \mathbf{N} * \mathbf{D})E(\boldsymbol{\beta} | \tilde{\boldsymbol{\beta}}, \mathbf{D}) = \mathbf{N} \tilde{\boldsymbol{\beta}}$ .

100  
 101 **LDpred.** The LDpred method is an extension of LDpred-inf that uses a point-normal prior to es-  
 102 timate posterior mean effect sizes via Markov Chain Monte Carlo (MCMC)<sup>12</sup>. It assumes a Gaussian  
 103 mixture prior:  $\beta_i \sim N(0, h_g^2 / M * p)$  with probability  $p$ , and  $\beta_i \sim 0$  with probability  $1 - p$ , where  $p$  is the  
 104 proportion of causal SNPs. The method is optimized by considering different values of  $p$  (1E-4, 3E-4,  
 105 1E-3, 3E-3, 0.01, 0.03, 0.1, 0.3, 1). We excluded SNPs from long-range LD regions (reported in ref. 23),  
 106 as our secondary analyses showed that including these regions was suboptimal, consistent with ref. 24.

107  
 108 **P+T-funct-LASSO.** Ref. 16 proposed an extension of P+T that corrects the marginal effect  
 109 sizes of SNPs for winner’s curse and incorporates external functional annotation data (P+T-funct-  
 110 LASSO). The winner’s curse correction is performed by applying a LASSO shrinkage to the marginal  
 111 association statistics of the PRS:

$$PRS_{LASSO}(P_T) = \sum_{i=1}^M \text{sign}(\tilde{\beta}_i) (|\tilde{\beta}_i| - \lambda(P_T)) \mathbf{1}_{\{P_i < P_T\}} g_i, \quad (3)$$

112 where  $\lambda(P_T) = \Phi^{-1}(1 - \frac{P_T}{2}) \text{sd}(\tilde{\beta}_i)$ , where  $\Phi^{-1}$  is the inverse standard normal CDF. Functional anno-  
 113 tations are incorporated via two disjoint SNPs sets, representing ”high-prior” SNPs (HP) and ”low-  
 114 prior” SNPs (LP), respectively. We define the HP SNP set for P+T-funct-LASSO as the set of SNPs  
 115 in the top 10% of expected per-SNP heritability under the baseline-LD model<sup>18</sup>, which includes cod-  
 116 ing, conserved, regulatory and LD-related annotations, whose enrichments are jointly estimated using  
 117 stratified LD score regression<sup>5,18</sup> (see Baseline-LD model annotations section). We also performed  
 118 secondary analyses using the top 5% (P+T-funct-LASSO-top5%). We define  $PRS_{LASSO,HP}(P_{HP})$   
 119 to be the PRS restricted to the HP SNP set, and  $PRS_{LASSO,LP}(P_{LP})$  to be the PRS restricted to  
 120 the LP SNP set, where  $P_{HP}$  and  $P_{LP}$  are the optimal significance thresholds for the HP and LP SNP  
 121 sets, respectively. We define  $PRS_{LASSO}(P_{HP}, P_{LP}) = PRS_{LASSO,HP}(P_{HP}) + PRS_{LASSO,LP}(P_{LP})$ .  
 122 We also performed secondary analyses where we allow an additional regularization to the two PRS:  
 123  $PRS_{LASSO}(P_{HP}, P_{LP}) = \alpha_1 PRS_{LASSO,HP}(P_{HP}) + \alpha_2 PRS_{LASSO,LP}(P_{LP})$ ; we refer to this method  
 124 as P+T-funct-LASSO-weighted.

125  
 126 **LDpred-funct-inf.** We modify LDpred-inf to incorporate functionally informed priors on causal

127 effect sizes using the baseline-LD model<sup>18</sup>, which includes coding, conserved, regulatory and LD-  
 128 related annotations, whose enrichments are jointly estimated using stratified LD score regression<sup>5,18</sup>.  
 129 Specifically, we assume that normalized causal effect sizes have prior distribution  $\beta_i \sim N(0, c * \sigma_i^2)$ ,  
 130 where  $\sigma_i^2$  is the expected per-SNP heritability under the baseline-LD model (fit using training data  
 131 only) and  $c$  is a normalizing constant such that  $\sum_{i=1}^M \mathbb{1}_{\{\sigma_i^2 > 0\}} c \sigma_i^2 = h_g^2$ ; SNPs with  $\sigma_i^2 \leq 0$  are  
 132 removed, which is equivalent to setting  $\sigma_i^2 = 0$ . The posterior mean causal effect sizes are

$$E[\beta|\tilde{\beta}, \mathbf{D}, \sigma_1^2, \dots, \sigma_{M_+}^2] = \mathbf{W}^{-1} N * \tilde{\beta} = \left[ N * \mathbf{D} + \frac{1}{c} \begin{pmatrix} \frac{1}{\sigma_1^2} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\sigma_{M_+}^2} \end{pmatrix} \right]^{-1} N * \tilde{\beta}, \quad (4)$$

133 where  $M_+$  is the number of SNPs with  $\sigma_i^2 > 0$ . The posterior mean causal effect sizes are computed by  
 134 solving the system of linear equations  $\mathbf{W}E[\beta|\tilde{\beta}, \mathbf{D}, \sigma_1^2, \dots, \sigma_{M_+}^2] = N * \tilde{\beta}$ .  $h_g^2$  is estimated as described  
 135 above (see LDpred-inf).  $\mathbf{D}$  is estimated using validation data, restricting to windows of size  $0.15\%M_+$ .  
 136

137 **LDpred-funct.** We modify LDpred-funct-inf to regularize posterior mean causal effect sizes using  
 138 cross-validation. We rank the SNPs by their (absolute) posterior mean causal effect sizes, partition  
 139 the SNPs into  $K$  bins (analogous to ref. 25) where each bin has roughly the same sum of squared  
 140 posterior mean effect sizes, and determine the relative weights of each bin based on predictive value  
 141 in the validation data. Intuitively if a bin is dominated by non-causal SNPs, the inferred relative  
 142 weight will be lower than for a bin with a high proportion of causal SNPs. This non-parametric  
 143 shrinkage approach can optimize prediction accuracy regardless of the genetic architecture. In detail,  
 144 let  $S = \sum_i E[\beta_i|\tilde{\beta}_i]^2$ . To define each bin, we first rank the posterior mean effect sizes based on their  
 145 squared values  $E[\beta_i|\tilde{\beta}_i]^2$ . We define bin  $b_1$  as the smallest set of top SNPs with  $\sum_{i \in b_1} E[\beta_i|\tilde{\beta}_i]^2 \geq \frac{S}{K}$ ,  
 146 and iteratively define bin  $b_k$  as the smallest set of additional top SNPs with  $\sum_{i \in b_1, \dots, b_k} E[\beta_i|\tilde{\beta}_i]^2 \geq \frac{kS}{K}$ .  
 147 Let  $PRS(k) = \sum_{i \in b_k} E[\beta_i|\tilde{\beta}_i]g_i$ . We define

$$PRS_{LDpred-funct} = \sum_{k=1}^K \alpha_k PRS(k), \quad (5)$$

148 where the bin-specific weights  $\alpha_k$  are optimized using validation data via 10-fold cross-validation. For  
 149 each held-out fold in turn, we split the data so we estimate the weights  $\alpha_k$  using the samples from the  
 150 other nine folds (90% of the validation) and compute PRS on the held-out fold using these weights  
 151 (10% of the validation). We then compute the average prediction  $R^2$  across the 10 held-out folds. To  
 152 avoid overfitting when  $K$  is very close to  $N$ , we set the number of bins ( $K$ ) to be between 1 and 100,  
 153 such that it is proportional to  $h_g^2$  and the number of samples used to estimate the  $K$  weights in each

154 fold is at least 100 times larger than  $K$ :

$$K = \min(100, \lceil \frac{0.9N * h_g^2}{100} \rceil), \quad (6)$$

155 where  $N$  is the number of validation samples. For highly heritable traits ( $h_g^2 \sim 0.5$ ), LDpred-  
156 funct reduces to the LDpred-funct-inf method if there are  $\sim 200$  validation samples or fewer; for less  
157 heritable traits ( $h_g^2 \sim 0.1$ ), LDpred-funct reduces to the LDpred-funct-inf method if there are  $\sim 1,000$   
158 validation samples or fewer. In simulations, we set  $K$  to 40 (based on 7,585 validation samples; see  
159 below), approximately concordant with Equation 6. The value of 100 in the denominator of Equation  
160 6 was coarsely optimized in simulations, but was not optimized using real trait data.

161 **Standard errors.** Standard errors for the prediction  $R^2$  of each method and the difference in  
162 prediction  $R^2$  between two methods were computed via block-jackknife using 200 genomic jackknife  
163 blocks<sup>5</sup>; this is more conservative than computing standard errors based on the number of validation  
164 samples, which does not account for variation across a finite number of SNPs. For each method,  
165 we first optimized any relevant tuning parameters using the entire genome and then analyzed each  
166 jackknife block using those tuning parameters.

## 167 Simulations

168 We simulated quantitative phenotypes using real genotypes from the UK Biobank interim release  
169 (see below). We used up to 50,000 unrelated British-ancestry samples as training samples, and 7,585  
170 samples of other European ancestries as validation samples (see below). We made these choices to  
171 minimize confounding due to shared population stratification or cryptic relatedness between training  
172 and validation samples (which, if present, could overstate the prediction accuracy that could be ob-  
173 tained in independent samples<sup>26</sup>), while preserving a large number of training samples. We restricted  
174 our simulations to 459,284 imputed SNPs on chromosome 1 (see below), fixed the number of causal  
175 SNPs at 2,000 or 5,000 (we also performed secondary simulations with 1,000 or 10,000 causal vari-  
176 ants), and fixed the SNP-heritability  $h_g^2$  at 0.5. We sampled normalized causal effect sizes  $\beta_i$  for causal  
177 SNPs from a normal distribution with variance equal to  $\frac{\sigma_i^2}{p}$ , where  $p$  is the proportion of causal SNPs  
178 and  $\sigma_i^2$  is the expected causal per-SNP heritability under the baseline-LD model<sup>18</sup>, fit using strati-  
179 fied LD score regression (S-LDSC)<sup>5,18</sup> applied to height summary statistics computed from unrelated  
180 British-ancestry samples from the UK Biobank interim release ( $N=113,660$ ). We computed per-allele  
181 effect sizes  $b_i$  as  $b_i = \frac{\beta_i}{\sqrt{2p_i(1-p_i)}}$ , where  $p_i$  is the minor allele frequency for SNP  $i$  estimated using the  
182 validation genotypes. We simulated phenotypes as  $Y_j = \sum_i^M b_i g_{ij} + \epsilon_j$ , where  $\epsilon_j \sim N(0, 1 - h_g^2)$ . We  
183 set the training sample size to either 10,000, 20,000 or 50,000. The motivation to perform simulations  
184 using one chromosome is to be able to extrapolate performance at larger sample sizes<sup>12</sup> according to



185 the ratio  $N/M$ , where  $N$  is the training sample size. We compared each of the five methods described  
186 above. For LDpred-funct-inf and LDpred-funct, for each simulated trait we used S-LDSC (applied to  
187 training data only) to estimate baseline-LD model parameters. For LDpred-funct, we report  $R^2$  as  
188 the average prediction  $R^2$  across the 10 held-out folds.

## 189 Full UK Biobank data set

190 The full UK Biobank data set includes 459,327 European-ancestry samples and  $\sim 20$  million imputed  
191 SNPs<sup>23</sup> (after filtering as in ref. 20, excluding indels and structural variants). We selected 21 UK  
192 Biobank traits (14 quantitative traits and 7 binary traits) with phenotyping rate  $> 80\%$  ( $> 80\%$  of  
193 females for age at menarche,  $> 80\%$  of males for balding), SNP-heritability  $h_g^2 > 0.2$  for quantitative  
194 traits, observed-scale SNP-heritability  $h_g^2 > 0.1$  for binary traits, and low correlation between traits  
195 (as described in ref. 20). We restricted training samples to 409,728 British-ancestry samples<sup>23</sup>,  
196 including related individuals (avg  $N=365$ K phenotyped training samples; see Table S1 for quantitative  
197 traits and Table S2 for binary traits). We computed association statistics from training samples using  
198 BOLT-LMM v2.3<sup>20</sup>. We have made these association statistics publicly available (see Web Resources).  
199 We restricted validation samples to 25,112 samples of non-British European ancestry, after removing  
200 validation samples that were related ( $> 0.05$ ) to training samples and/or other validation samples (avg  
201  $N=22$ K phenotyped validation samples; see Table S1 and S2). As in our simulations, we made these  
202 choices to minimize confounding due to shared population stratification or cryptic relatedness between  
203 training and validation samples (which, if present, could overstate the prediction accuracy that could  
204 be obtained in independent samples<sup>26</sup>), while preserving a large number of training samples. We  
205 analyzed 6,334,603 genome-wide imputed SNPs, after removing SNPs with minor allele frequency  
206  $< 1\%$ , removing SNPs with imputation accuracy  $< 0.9$ , and removing A/T and C/G SNPs to  
207 eliminate potential strand ambiguity. We used  $h_g^2$  estimates from BOLT-LMM v2.3<sup>20</sup> as input to  
208 LDpred, LDpred-funct-inf and LDpred-funct.

## 209 UK Biobank interim release

210 The UK Biobank interim release includes 145,416 European-ancestry samples<sup>27</sup>. We used the UK  
211 Biobank interim release both in simulations using real genotypes, and in a subset of analyses of height  
212 phenotypes (to investigate how prediction accuracy varies with training sample size).

213 In our analyses of height phenotypes, we restricted training samples to 113,660 unrelated ( $\leq 0.05$ )  
214 British-ancestry samples for which height phenotypes were available. We computed association statis-  
215 tics by adjusting for 10 PCs<sup>28</sup>, estimated using FastPCA<sup>29</sup> (see Web Resources). For consistency,  
216 we used the same set of 25,030 validation samples of non-British European ancestry with height  
217 phenotypes as defined above. We analyzed 5,957,957 genome-wide SNPs, after removing SNPs with

218 minor allele frequency  $< 1\%$ , removing SNPs with imputation accuracy  $< 0.9$ , removing SNPs that  
219 were not present in the 23andMe height data set (see below), and removing A/T and C/G SNPs to  
220 eliminate potential strand ambiguity. .

221 In our simulations, we restricted training samples to up to 50,000 of the 113,660 unrelated British-  
222 ancestry samples, and restricted validation samples to 8,441 samples of non-British European ancestry,  
223 after removing validation samples that were related ( $> 0.05$ ) to training samples and/or other valida-  
224 tion samples. We restricted the 5,957,957 genome-wide SNPs (see above) to chromosome 1, yielding  
225 459,284 SNPs after QC.

## 226 **23andMe height summary statistics**

227 The 23andMe data set consists of summary statistics computed from 698,430 European-ancestry  
228 samples (23andMe customers who consented to participate in research) at 9,898,287 imputed SNPs,  
229 after removing SNPs with minor allele frequency  $< 1\%$  and that passed QC filters (which include  
230 filters on imputation quality,  $\text{avg.rsq} < 0.5$  or  $\text{min.rsq} < 0.3$  in any imputation batch, and imputation  
231 batch effects). Analyses were restricted to the set of individuals with  $> 97\%$  European ancestry,  
232 as determined via an analysis of local ancestry<sup>30</sup>. Summary association statistics were computed  
233 using linear regression adjusting for age, gender, genotyping platform, and the top five principal  
234 components to account for residual population structure. The summary association statistics will be  
235 made available to qualified researchers (see Web Resources).

236 We analyzed 5,957,935 genome-wide SNPs, after removing SNPs with minor allele frequency  $< 1\%$ ,  
237 removing SNPs with imputation accuracy  $< 0.9$ , removing SNPs that were not present in the full  
238 UK Biobank data set (see above), and removing A/T and C/G SNPs to eliminate potential strand  
239 ambiguity.

## 240 **Meta-analysis of full UK Biobank and 23andMe height data sets**

241 We meta-analyzed height summary statistics from the full UK Biobank and 23andMe data sets. We  
242 define

$$PRS_{meta} = \gamma_1 PRS_1 + \gamma_2 PRS_2, \quad (7)$$

243 where  $PRS_i$  is the PRS obtained using training data from cohort  $i$ . The PRS can be obtained using  
244 P+T, P+T-funct-LASSO, LDpred-inf or LDpred-funct. The meta-analysis weights  $\gamma_i$  can either be  
245 specified via fixed-effect meta-analysis (e.g.  $\gamma_i = \frac{N_i}{\sum N_i}$ ) or optimized using validation data<sup>19</sup>. We  
246 use the latter approach, which can improve prediction accuracy (e.g. if the cohorts differ in their  
247 heritability as well as their sample size). In our primary analyses, we fit the weights  $\gamma_i$  in-sample  
248 and report prediction accuracy using adjusted  $R^2$  to account for in-sample fitting<sup>19</sup>. We also report

249 results using 10-fold cross-validation: for each held-out fold in turn, we estimate the weights  $\gamma_i$  using  
250 the other nine folds and compute PRS on the held-out fold using these weights. We then compute  
251 the average prediction  $R^2$  across the 10 held-out folds.

252 When using LDpred-funct as the prediction method, we perform the meta-analysis as follows.  
253 First, we use LDpred-funct-inf to fit meta-analysis weights  $\gamma_i$ . Then, we use  $\gamma_i$  to compute (meta-  
254 analysis) weighted posterior mean causal effect sizes (PMCES) via  $PMCES = \gamma_1 PMCES_1 +$   
255  $\gamma_2 PMCES_2$ , which are binned into  $k$  bins. Then, we estimate bin-specific weights  $\alpha_k$  (used to com-  
256 pute (meta-analysis + bin-specific) weighted posterior mean causal effect sizes  $\sum_{k=1}^K \alpha_k PMCES(k)$ )  
257 using validation data via 10-fold cross validation.

## 258 Baseline-LD model annotations

259 The baseline-LD model (v1.1) contains a broad set of 75 functional annotations (including coding,  
260 conserved, regulatory and LD-related annotations), whose enrichments are jointly estimated using  
261 stratified LD score regression<sup>5,18</sup>. For each trait, we used the  $\tau_c$  values estimated for that trait to  
262 compute  $\sigma_i^2$ , the expected per-SNP heritability of SNP  $i$  under the baseline-LD model, as

$$\sigma_i^2 = \sum_c a_c(i) \tau_c, \quad (8)$$

263 where  $a_c(i)$  is the value of annotation  $c$  at SNP  $i$ .

264 Joint effect sizes  $\tau_c$  for each annotation  $c$  are estimated via

$$E[\chi_i^2] = N \sum_c \tau_c l(i, c) + 1, \quad (9)$$

265 where  $l(i, c)$  is the LD score of SNP  $i$  with respect to annotation  $a_c$  and  $\chi_i^2$  is the chi-square statistic  
266 for SNP  $i$ . We note that  $\tau_c$  quantifies effects that are unique to annotation  $c$ . In all analyses of real  
267 phenotypes,  $\tau_c$  and  $\sigma_i^2$  were estimated using training samples only.

268 In our primary analyses, we used 489 unrelated European samples from phase 3 of the 1000  
269 Genomes Project<sup>31</sup> as the reference data set to compute LD scores, as in ref. 18.

270 To verify that our 1000 Genomes reference data set produces reliable LD estimates, we repeated  
271 our LDpred-funct analyses using S-LDSC with 3,567 unrelated individuals from UK10K<sup>32</sup> as the  
272 reference data set (as in ref. 33), ensuring a closer ancestry match with British-ancestry UK Biobank  
273 samples. We also repeated our LDpred-funct analyses using S-LDSC with the baseline-LD+LDAK  
274 model (instead of the baseline-LD model), with UK10K as the reference data set. The baseline-  
275 LD+LDAK model (introduced in ref. 33) consists of the baseline-LD model plus one additional  
276 continuous annotation constructed using LDAK weights<sup>34</sup>, which has values  $(p_j(1 - p_j))^{1+\alpha} w_j$ ,

277 where  $\alpha = -0.25$ ,  $p_j$  is the allele frequency of SNP  $j$ , and  $w_j$  is the LDAK weight of SNP  $j$  computed  
278 using UK10K data.

## 279 Results

### 280 Simulations

281 We performed simulations using real genotypes from the UK Biobank interim release and simulated  
282 phenotypes (see Methods). We simulated quantitative phenotypes with SNP-heritability  $h_g^2 = 0.5$ ,  
283 using 476,613 imputed SNPs from chromosome 1. We selected either 2,000 or 5,000 variants to  
284 be causal; we refer to these as "sparse" and "polygenic" architectures, respectively. We sampled  
285 normalized causal effect sizes from normal distributions with variances based on expected causal  
286 per-SNP heritabilities under the baseline-LD model<sup>18</sup>, fit using stratified LD score regression (S-  
287 LDSC)<sup>5,18</sup> applied to height summary statistics from British-ancestry samples from the UK Biobank  
288 interim release. We randomly selected 10,000, 20,000 or 50,000 unrelated British-ancestry samples as  
289 training samples, and we used 7,585 unrelated samples of non-British European ancestry as validation  
290 samples. By restricting simulations to chromosome 1 ( $\approx 1/10$  of SNPs), we can extrapolate results  
291 to larger sample sizes ( $\approx 10\times$  larger; see Application to 21 UK Biobank traits), analogous to previous  
292 work<sup>12</sup>.

293 We compared prediction accuracies ( $R^2$ ) for five main methods: P+T<sup>14,15</sup>, LDpred<sup>12</sup>, P+T-  
294 funct-LASSO<sup>16</sup>, LDpred-funct-inf and LDpred-funct (see Methods). Results are reported in Figure 1  
295 (main simulations) and Figure S1 (additional values of number of causal variants); numerical results  
296 are reported in Table S3 and Table S4. Among methods that do not use functional information, the  
297 prediction accuracy of LDpred was higher than P+T (particularly for the polygenic architecture),  
298 consistent with previous work<sup>8,12</sup> (see Table S5 and Table S6 for optimal tuning parameters).

299  
300 Incorporating functional information via LDpred-funct-inf (a method that does not model spar-  
301 sity) produced improvements that varied with sample size (+4.7% relative improvement for sparse  
302 architecture and +4.8% relative improvement for polygenic architecture at  $N=50K$  training samples,  
303 compared to LDpred; smaller improvements at smaller sample sizes). These results are consistent  
304 with the fact that LDpred is known to be sensitive to model assumptions at large sample sizes<sup>12</sup>.  
305 Accounting for sparsity using LDpred-funct further improved prediction accuracy, particularly for  
306 the sparse architecture (+7.3% relative improvement for sparse architecture and +5.4% relative im-  
307 provement for polygenic architecture at  $N=50K$  training samples, compared to LDpred; smaller  
308 improvements at smaller sample sizes). LDpred-funct attained substantially higher prediction accu-  
309 racy than P+T-funct-LASSO in most settings (+11% relative improvement for sparse architecture

310 and +18% relative improvement for polygenic architecture at  $N=50K$  training samples; smaller im-  
311 provements at smaller sample sizes). The difference in prediction accuracy between LDpred and each  
312 other method, as well as the difference in prediction accuracy between LDpred-funct and each other  
313 method, was statistically significant in most cases (see Table S4). Simulations with 1,000 or 10,000  
314 causal variants generally recapitulated these findings, although P+T-funct-LASSO performed better  
315 than LDpred-funct for the extremely sparse architecture (Table S3).

316 We performed three secondary analyses. First, we assessed the calibration of each method by  
317 checking whether a regression of true vs. predicted phenotype yielded a slope of 1. We determined  
318 that LDpred-funct was well-calibrated (regression slope 0.98-0.99), LDpred was fairly well-calibrated  
319 (regression slope 0.85-1.00), and other methods were not well-calibrated (Table S7). Second, we  
320 assessed the sensitivity of LDpred-funct to the choice of  $K=40$  posterior mean causal effect size bins  
321 to regularize effect sizes in our main simulations. We determined that results were not sensitive to  
322 this parameter (Table S8); slightly higher values of  $K$  performed slightly better, but we did not finely  
323 optimize this parameter. Third, we evaluated a "cheating" version of LDpred-funct that utilized the  
324 true baseline-LD model parameters used to simulate the data, instead of estimating these parameters  
325 from the data (LDpred-funct-cheat). LDpred-funct-cheat performed only slightly better than LDpred-  
326 funct, indicating that LDpred-funct is not sensitive to imperfect estimation of functional enrichment  
327 parameters (see Table S9).

## 328 Application to 21 UK Biobank traits

329 We applied P+T, LDpred, P+T-funct-LASSO, LDpred-funct-inf and LDpred-funct to 21 UK Biobank  
330 traits (14 quantitative traits and 7 binary traits; Table S1 and Table S2). We analyzed training  
331 samples of British ancestry (avg  $N=365K$ ) and validation samples of non-British European ancestry  
332 (avg  $N=22K$ ). We included 6,334,603 imputed SNPs in our analyses (see Methods). We computed  
333 summary statistics and  $h_g^2$  estimates from training samples using BOLT-LMM v2.3<sup>20</sup> (see Table S10).  
334 We estimated trait-specific functional enrichment parameters for the baseline-LD model<sup>18</sup> by running  
335 S-LDSC<sup>5,18</sup> on these summary statistics. Results for quantitative traits are reported in Figure 2 and  
336 Table S11, and results for binary traits are reported in Figure 3 and Table S12. Differences between  
337 each main prediction method and LDpred (and block-jackknife standard errors on these differences)  
338 are reported in Table S13, and averages across all 21 traits for main and secondary prediction methods  
339 are reported in Table S14.

340 Among methods that do not use functional information, LDpred outperformed P+T (+18% rel-  
341 ative improvement in avg prediction  $R^2$ ), consistent with simulations under a polygenic architecture  
342 (see Table S15 and Table S16 for optimal tuning parameters) and with previous work<sup>8,12</sup>. LDpred also  
343 outperformed LDpred-inf, a method that does not model sparsity (see Table S14). The exclusion of

344 long-range LD regions (see Methods) was critical to LDpred performance, as running LDpred without  
345 excluding long-range LD regions (as implemented in a previous version of this paper<sup>35</sup>) performed  
346 much worse (see Table S14).

347 Incorporating functional information via LDpred-funct-inf (a method that does not model spar-  
348 sity) performed only slightly better than LDpred (+0.9% relative improvement in avg prediction  $R^2$ ).  
349 Accounting for sparsity using LDpred-funct substantially improved prediction accuracy (+8.7% rel-  
350 ative improvement in avg prediction  $R^2$  vs. LDpred,  $P = 0.006$  for difference using one-sided z-test  
351 based on block-jackknife standard error in Table S13; avg prediction  $R^2=0.145$ ; highest  $R^2=0.413$  for  
352 height), consistent with simulations. The relative improvement in avg prediction  $R^2$  for LDpred-funct  
353 vs. LDpred was larger for quantitative traits (+9.2%; higher prediction  $R^2$  for 14/14 traits) than for  
354 binary traits (+6.6%; higher prediction  $R^2$  for 2/7 traits), consistent with the higher average  $h_g^2$  for  
355 quantitative traits (0.33) than for binary traits (0.19; observed scale), which corresponds to higher  
356 effective sample size (see simulation results in Figure 1) and higher absolute prediction  $R^2$  (Figure  
357 2 vs. Figure 3). Accordingly, the improvement of LDpred-funct vs. LDpred across all 21 traits was  
358 smaller when averaging relative improvements in prediction  $R^2$  for each trait individually (+6.3%), a  
359 computation that more heavily weights traits with low prediction  $R^2$ . LDpred-funct also performed  
360 substantially better than P+T-funct-LASSO (+19% relative improvement in avg prediction  $R^2$ ),  
361 consistent with simulations under a polygenic architecture.

362 We performed several secondary analyses. First, we assessed the calibration of each method  
363 by checking whether a regression of true vs. predicted phenotype yielded a slope of 1. As in our  
364 simulations, we determined that LDpred-funct was well-calibrated (average regression slope: 0.98),  
365 LDpred was fairly well-calibrated (average regression slope: 0.89), and other methods were not well-  
366 calibrated (Table S17). Second, we assessed the sensitivity of LDpred-funct to the average value of  
367  $K = 58$  posterior mean causal effect size bins to regularize effect sizes in these analyses (see Equation  
368 6 and Table S10). We determined that results were not sensitive to the number of bins (Table S18).  
369 Third, we assessed the sensitivity of LDpred-funct to validation sample size; we note that our main  
370 analyses involved very large validation sample sizes (up to 25,032; Table S1 and Table S2), which  
371 aids the regularization step of LDpred-funct. We determined that results were little changed when  
372 restricting to smaller validation sample sizes (as low as 1,000; see Table S19). Fourth, we determined  
373 that functional enrichment information is far less useful when restricting to genotyped variants (e.g.  
374 -6.9% relative change in avg prediction  $R^2$  for LDpred-funct vs. LDpred when both methods are  
375 restricted to typed variants; Table S14), likely because tagging variants may not belong to enriched  
376 functional annotations. Fifth, we evaluated a modification of P+T-funct-LASSO in which different  
377 weights were allowed for the two predictors (P+T-funct-LASSO-weighted; see Methods), but results  
378 were little changed (+1.1% relative improvement in avg prediction  $R^2$  vs. P+T-funct-LASSO; Table

379 S14). Sixth, we obtained similar results for P+T-funct-LASSO when defining the "high-prior" (HP)  
380 SNP set using the top 5% of SNPs with the highest per-SNP heritability, instead of the top 10% (see  
381 Table S14). Seventh, we determined that incorporating baseline-LD model functional enrichments  
382 that were meta-analyzed across traits (31 traits from ref. 18), instead of the trait-specific functional  
383 enrichments used in our primary analyses, slightly reduced the prediction accuracy of LDpred-funct-  
384 inf (Table S14). Eighth, we determined that using our previous baseline model<sup>5</sup>, instead of the  
385 baseline-LD model<sup>18</sup>, slightly reduced the prediction accuracy of LDpred-funct-inf and LDpred-funct  
386 (Table S14). Ninth, we determined that inferring functional enrichments using only the SNPs that  
387 passed QC filters and were used for prediction had no impact on the prediction accuracy of LDpred-  
388 funct-inf (Table S14). Tenth, we determined that using UK10K (instead of 1000 Genomes) as the LD  
389 reference panel had virtually no impact on prediction accuracy (Table S14).

### 390 **Application to height in meta-analysis of UK Biobank and 23andMe cohorts**

391 We applied P+T, LDpred-inf, P+T-funct-LASSO, LDpred-funct-inf and LDpred-funct to predict  
392 height in a meta-analysis of UK Biobank and 23andMe cohorts (see Methods). Training sample sizes  
393 were equal to 408,092 for UK Biobank and 698,430 for 23andMe, for a total of 1,106,522 training  
394 samples. For comparison purposes, we also computed predictions using the UK Biobank and 23andMe  
395 training data sets individually, as well as a training data set consisting of 113,660 British-ancestry  
396 samples from the UK Biobank interim release. (The analysis using the 408,092 UK Biobank training  
397 samples was nearly identical to the analysis of Figure 2, except that we used a different set of 5,957,935  
398 SNPs, for consistency throughout this set of comparisons; see Methods.) We used 25,030 UK Biobank  
399 samples of non-British European ancestry as validation samples in all analyses.

400 Results are reported in Figure 4 and Table S20. The relative improvements attained by LDpred-  
401 funct-inf and LDpred-funct were broadly similar across all four training data sets (also see Figure  
402 2), implying that these improvements are not specific to the UK Biobank data set. Interestingly,  
403 compared to the full UK Biobank training data set ( $R^2=0.413$  for LDpred-funct), prediction accuracies  
404 were only slightly higher for the meta-analysis training data set ( $R^2=0.429$  for LDpred-funct), and  
405 were lower for the 23andMe training data set ( $R^2=0.328$  for LDpred-funct), consistent with the  $\approx 30\%$   
406 higher heritability in UK Biobank as compared to 23andMe and other large cohorts<sup>18,20,21</sup>; the higher  
407 heritability in UK Biobank could potentially be explained by lower environmental heterogeneity. We  
408 note that in the meta-analysis, we optimized the meta-analysis weights using validation data (similar  
409 to ref. 19), instead of performing a fixed-effect meta-analysis. This approach accounts for differences  
410 in heritability as well as sample size, and attained a +5.9% relative improvement in prediction  $R^2$   
411 compared to fixed-effects meta-analysis (see Table S20).

## 412 Discussion

413 We have shown that leveraging trait-specific functional enrichments inferred by S-LDSC with the  
414 baseline-LD model<sup>18</sup> substantially improves polygenic prediction accuracy. Across 21 UK Biobank  
415 traits, we attained a +9% relative improvement in average prediction  $R^2$  using a method that leverages  
416 functional enrichment and performs an additional regularization step to account for sparsity (LDpred-  
417 funct), compared to the most accurate method that does not model functional enrichment (LDpred).  
418 We note that our main analyses used baseline-LD model v1.1, but using the updated baseline-LD  
419 model v2.1 yields slightly higher prediction  $R^2$  for LDpred-funct-inf and LDpred-funct (Table S14).

420 Previous work has highlighted the potential advantages of leveraging functional enrichment to  
421 improve prediction accuracy<sup>16,17</sup>. We included one such method<sup>16</sup> (which we call P+T-funct-LASSO)  
422 in our analyses, determining that LDpred-funct attains a +19% average relative improvement vs.  
423 P+T-funct-LASSO across 21 UK Biobank traits. More recently, ref. 17 introduced AnnoPred, which  
424 uses a Bayesian framework to incorporate functional annotations. However, ref. 17 considered only  
425 genotyped variants and binary annotations. As noted above, functional enrichment information is  
426 far less useful when restricting to genotyped variants (Table S14), likely because tagging variants  
427 may not belong to enriched functional annotations; thus, the utility of AnnoPred in more general  
428 settings is currently unknown. To assess this, we applied AnnoPred to the 21 UK Biobank traits (see  
429 Table S14 and Table S21. We determined that AnnoPred performed slightly but non-significantly  
430 worse than LDpred-funct (-2.3% relative change in avg prediction  $R^2$  for AnnoPred vs. LDpred-  
431 funct,  $P = 0.17$  for difference using one-sided z-test based on block-jackknife standard error in  
432 Table S21). We emphasize that our study is, to our knowledge, the first study that combines binary  
433 and continuous-valued functional annotations to improve polygenic risk prediction using imputed  
434 variants.

435 Our work has several limitations. First, LDpred-funct analyzes summary statistic training data  
436 (which are publicly available for a broad set of diseases and traits<sup>36</sup>), but methods that use raw  
437 genotypes/phenotypes as training data have the potential to attain higher accuracy<sup>20</sup>; incorporating  
438 functional enrichment information into prediction methods that use raw genotypes/phenotypes as  
439 training data remains a direction for future research. Second, the regularization step employed by  
440 LDpred-funct to account for sparsity relies on heuristic cross-validation instead of inferring posterior  
441 mean causal effect sizes under a prior sparse functional model; we made this choice because the  
442 appropriate choice of sparse functional model is unclear, and because inference of posterior means via  
443 MCMC may be subject to convergence issues. As a consequence, the improvement of LDpred-funct  
444 over LDpred-funct-inf may be contingent on the number of validation samples available for cross-  
445 validation; in particular, for very small validation samples, the number of cross-validation bins is  
446 equal to 1 (Equation 6) and LDpred-funct is identical to LDpred-funct-inf. However, we determined



447 that results of LDpred-funct were little changed when restricting to smaller validation sample sizes  
448 (as low as 1,000; see Table S19). Third, we have considered only single-trait analyses, but leveraging  
449 genetic correlations among traits has considerable potential to improve prediction accuracy<sup>37,38</sup>.  
450 Fourth, we have not considered how to leverage functional enrichment for polygenic prediction in  
451 related individuals<sup>39</sup>. Fifth, we have not investigated the application of our methods to polygenic  
452 prediction in diverse populations<sup>19,40,41</sup>, for which very similar functional enrichments have been  
453 reported<sup>42,43</sup>. Finally, the improvements in prediction accuracy that we reported are a function of the  
454 baseline-LD model<sup>18</sup>, but there are many possible ways to improve this model, e.g. by incorporating  
455 tissue-specific enrichments<sup>1-6,44-47</sup>, modeling MAF-dependent architectures<sup>48-50</sup>, and/or employing  
456 alternative approaches to modeling LD-dependent effects<sup>34</sup>; we anticipate that future improvements  
457 to the baseline-LD model will yield even larger improvements in prediction accuracy. As an initial  
458 step to explore alternative approaches to modeling LD-dependent effects, we repeated our analyses  
459 using the baseline-LD+LDAK model (introduced in ref. 33), which consists of the baseline-LD model  
460 plus one additional continuous annotation constructed using LDAK weights<sup>34</sup>. (Recent work has  
461 shown that incorporating LDAK weights increases polygenic prediction accuracy in analyses that  
462 do not include the baseline-LD model<sup>51</sup>.) We determined that results were virtually unchanged (avg  
463 prediction  $R^2=0.1350$  for baseline-LD+LDAK vs.  $0.1354$  for baseline-LD using LDpred-funct-inf with  
464 UK10K SNPs; see Table S14 and Table S22). Despite these limitations and open directions for future  
465 research, our work demonstrates that leveraging functional enrichment using the baseline-LD model  
466 substantially improves polygenic prediction accuracy.

## 467 **Acknowledgements**

468 We thank the research participants and employees of 23andMe for making this work possible. We are  
469 grateful to S. Sunyaev, S. Chun, L. O'Connor, O. Weissbrod and H. Finucane for helpful discussions.  
470 This research was conducted using the UK Biobank Resource under Application #16549 and was  
471 funded by NIH grants R01 GM105857, R01 MH101244 and U01 HG009379.

472 Collaborators for the 23andMe research team are: Michelle Agee, Babak Alipanahi, Robert K.  
473 Bell, Katarzyna Bryc, Sarah L. Elson, Pierre Fontanillas, David A. Hinds, Jennifer C. McCreight,  
474 Karen E. Huber, Aaron Kleinman, Nadia K. Litterman, Matthew H. McIntyre, Joanna L. Mountain,  
475 Elizabeth S. Noblin, Carrie A.M. Northover, Steven J. Pitts, J. Fah Sathirapongsasuti, Olga V.  
476 Sazonova, Janie F. Shelton, Suyash Shringarpure, Chao Tian, Joyce Y. Tung, Vladimir Vacic, and  
477 Catherine H. Wilson.

## 478 Author contributions

479 C.M.L. and A.L.P. designed experiments. C.M.L. performed experiments. C.M.L., S.G., P.R.L.,  
480 S.S.K., N.F. and A.A. analyzed data. C.M.L. and A.L.P. wrote the manuscript with assistance from  
481 S.G., P.R.L. S.S.K., N.F. and A.L.P.

## 482 Web Resources

483 Software implementing the LDpred-funct-inf and LDpred-funct: [https://www.hsph.harvard.edu/](https://www.hsph.harvard.edu/alkes-price/software)  
484 [alkes-price/software](https://www.hsph.harvard.edu/alkes-price/software)  
485 LDscore regression software: <https://github.com/bulik/ldsc>  
486 UK Biobank Resource: <http://www.ukbiobank.ac.uk/>  
487 BOLT-LMM v2.3 software <http://data.broadinstitute.org/alkesgroup/BOLT-LMM/>  
488 BOLT-LMM v2.3 association statistics: [https://data.broadinstitute.org/alkesgroup/UKBB/](https://data.broadinstitute.org/alkesgroup/UKBB/UKBB_409K/)  
489 [UKBB\\_409K/](https://data.broadinstitute.org/alkesgroup/UKBB/UKBB_409K/)  
490 23andMe height association statistics: The full summary statistics for the 23andMe height GWAS  
491 will be made available through 23andMe to qualified researchers under an agreement with 23andMe  
492 that protects the privacy of the 23andMe participants. Please visit [https://research.23andme.](https://research.23andme.com/collaborate/#publication)  
493 [com/collaborate/#publication](https://research.23andme.com/collaborate/#publication) for more information and to apply to access the data.

## 494 References

- 495 [1] Matthew T Maurano, Richard Humbert, Eric Rynes, Robert E Thurman, Eric Haugen, Hao  
496 Wang, Alex P Reynolds, Richard Sandstrom, Hongzhu Qu, Jennifer Brody, et al. Systematic  
497 localization of common disease-associated variation in regulatory dna. *Science*, page 1222794,  
498 2012.
- 499 [2] Gosia Trynka, Cynthia Sandor, Buhm Han, Han Xu, Barbara E Stranger, X Shirley Liu, and  
500 Soumya Raychaudhuri. Chromatin marks identify critical cell types for fine mapping complex  
501 trait variants. *Nature genetics*, 45(2):124, 2013.
- 502 [3] Joseph K Pickrell. Joint analysis of functional genomic data and genome-wide association studies  
503 of 18 human traits. *American Journal of Human Genetics*, 94(4):559–573, 04 2014.
- 504 [4] Roadmap Epigenomics Consortium, Anshul Kundaje, Wouter Meuleman, Jason Ernst, Misha  
505 Bilenky, Angela Yen, Alireza Heravi-Moussavi, Pouya Kheradpour, Zhizhuo Zhang, Jianrong  
506 Wang, Michael J. Ziller, Viren Amin, John W. Whitaker, Matthew D. Schultz, Lucas D. Ward,

507 Abhishek Sarkar, Gerald Quon, Richard S. Sandstrom, Matthew L. Eaton, Yi-Chieh Wu, An-  
508 dreas R. Pfenning, Xinchun Wang, Melina Claussnitzer, Yaping Liu, Cristian Coarfa, R. Alan  
509 Harris, Noam Shores, Charles B. Epstein, Elizabeta Gjoneska, Danny Leung, Wei Xie, R. David  
510 Hawkins, Ryan Lister, Chibo Hong, Philippe Gascard, Andrew J. Mungall, Richard Moore, Eric  
511 Chuah, Angela Tam, Theresa K. Canfield, R. Scott Hansen, Rajinder Kaul, Peter J. Sabo,  
512 Mukul S. Bansal, Annaick Carles, Jesse R. Dixon, Kai-How Farh, Soheil Feizi, Rosa Karlic,  
513 Ah-Ram Kim, Ashwinikumar Kulkarni, Daofeng Li, Rebecca Lowdon, GiNell Elliott, Tim R.  
514 Mercer, Shane J. Neph, Vitor Onuchic, Paz Polak, Nisha Rajagopal, Pradipta Ray, Richard C.  
515 Sallari, Kyle T. Siebenthal, Nicholas A. Sinnott-Armstrong, Michael Stevens, Robert E. Thur-  
516 man, Jie Wu, Bo Zhang, Xin Zhou, Arthur E. Beaudet, Laurie A. Boyer, Philip L. De Jager,  
517 Peggy J. Farnham, Susan J. Fisher, David Haussler, Steven J. M. Jones, Wei Li, Marco A.  
518 Marra, Michael T. McManus, Shamil Sunyaev, James A. Thomson, Thea D. Tlsty, Li-Huei Tsai,  
519 Wei Wang, Robert A. Waterland, Michael Q. Zhang, Lisa H. Chadwick, Bradley E. Bernstein,  
520 Joseph F. Costello, Joseph R. Ecker, Martin Hirst, Alexander Meissner, Aleksandar Milosavl-  
521 jevic, Bing Ren, John A. Stamatoyannopoulos, Ting Wang, and Manolis Kellis. Integrative  
522 analysis of 111 reference human epigenomes. *Nature*, 518:317 EP –, 02 2015.

523 [5] Hilary K Finucane, Brendan Bulik-Sullivan, Alexander Gusev, Gosia Trynka, Yakir Reshef, Po-  
524 Ru Loh, Verneri Anttila, Han Xu, Chongzhi Zang, Kyle Farh, Stephan Ripke, Felix R Day,  
525 ReproGen Consortium, Schizophrenia Working Group of the Psychiatric Genomics Consortium,  
526 The RACI Consortium, Shaun Purcell, Eli Stahl, Sara Lindstrom, John R B Perry, Yukinori  
527 Okada, Soumya Raychaudhuri, Mark J Daly, Nick Patterson, Benjamin M Neale, and Alkes L  
528 Price. Partitioning heritability by functional annotation using genome-wide association summary  
529 statistics. *Nature Genetics*, 47:1228 EP –, 09 2015.

530 [6] Kyle Kai-How Farh, Alexander Marson, Jiang Zhu, Markus Kleinewietfeld, William J Hous-  
531 ley, Samantha Beik, Noam Shores, Holly Whitton, Russell JH Ryan, Alexander A Shishkin,  
532 et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*,  
533 518(7539):337, 2015.

534 [7] Nilanjan Chatterjee, Jianxin Shi, and Montserrat García-Closas. Developing and evaluating  
535 polygenic risk prediction models for stratified disease prevention. *Nat Rev Genet*, 17(7):392–406,  
536 July 2016.

537 [8] Amit V. Khera, Mark Chaffin, Krishna G. Aragam, Mary E. Haas, Carolina Roselli, Seung Hoan  
538 Choi, Pradeep Natarajan, Eric S. Lander, Steven A. Lubitz, Patrick T. Ellinor, and Sekar  
539 Kathiresan. Genome-wide polygenic scores for common diseases identify individuals with risk  
540 equivalent to monogenic mutations. *Nature Genetics*, 50(9):1219–1224, 2018.

- 541 [9] Xiang Zhou, Peter Carbonetto, and Matthew Stephens. Polygenic modeling with bayesian sparse  
542 linear mixed models. *PLOS Genetics*, 9(2):1–14, 02 2013.
- 543 [10] Gerhard Moser, Sang Hong Lee, Ben J. Hayes, Michael E. Goddard, Naomi R. Wray, and  
544 Peter M. Visscher. Simultaneous discovery, estimation and prediction analysis of complex traits  
545 using a bayesian mixture model. *PLOS Genetics*, 11(4):1–22, 04 2015.
- 546 [11] Doug Speed and David J Balding. Multiblup: improved snp-based prediction for complex traits.  
547 *Genome Research*, 24(9):1550–1557, 09 2014.
- 548 [12] Bjarni J Vilhjálmsson, Jian Yang, Hilary K Finucane, Alexander Gusev, Sara Lindström, Stephan  
549 Ripke, Giulio Genovese, Po-Ru Loh, Gaurav Bhatia, Ron Do, et al. Modeling linkage disequi-  
550 librium increases accuracy of polygenic risk scores. *The American Journal of Human Genetics*,  
551 97(4):576–592, 2015.
- 552 [13] C. R. Henderson. Best linear unbiased estimation and prediction under a selection model. *Bio-*  
553 *metrics*, 31(2):423–447, 1975.
- 554 [14] International Schizophrenia Consortium, Shaun M. Purcell, Naomi R. Wray, Jennifer L. Stone,  
555 Peter M. Visscher, Michael C. O’Donovan, Patrick F. Sullivan, and Pamela Sklar. Common poly-  
556 genic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, 460(7256):748–  
557 752, August 2009.
- 558 [15] Eli A Stahl, Daniel Wegmann, Gosia Trynka, Javier Gutierrez-Achury, Ron Do, Benjamin F  
559 Voight, Peter Kraft, Robert Chen, Henrik J Kallberg, Fina AS Kurreeman, et al. Bayesian infer-  
560 ence analyses of the polygenic architecture of rheumatoid arthritis. *Nature genetics*, 44(5):483–  
561 489, 2012.
- 562 [16] Jianxin Shi, Ju-Hyun Park, Jubao Duan, Berndt, et al. Winner’s Curse Correction and Vari-  
563 able Thresholding Improve Performance of Polygenic Risk Modeling Based on Genome-Wide  
564 Association Study Summary-Level Data. *PLOS Genetics*, 12(12):e1006493, December 2016.
- 565 [17] Yiming Hu, Qiongshi Lu, Ryan Powles, Xinwei Yao, Can Yang, Fang Fang, Xinran Xu, and  
566 Hongyu Zhao. Leveraging functional annotations in genetic risk prediction for human complex  
567 diseases. *PLOS Computational Biology*, 13(6):1–16, 06 2017.
- 568 [18] Steven Gazal, Hilary K Finucane, Nicholas A Furlotte, Po-Ru Loh, Pier Francesco Palamara,  
569 Xuanyao Liu, Armin Schoech, Brendan Bulik-Sullivan, Benjamin M Neale, Alexander Gusev, and  
570 Alkes L Price. Linkage disequilibrium–dependent architecture of human complex traits shows  
571 action of negative selection. *Nature Genetics*, 49:1421 EP –, 09 2017.

- 572 [19] Carla Márquez-Luna, Po-Ru Loh, South Asian Type 2 Diabetes (SAT2D) Consortium, The  
573 SIGMA Type 2 Diabetes Consortium, and Alkes L. Price. Multiethnic polygenic risk scores  
574 improve risk prediction in diverse populations. *Genetic Epidemiology*, 41(8):811–823, 2017.
- 575 [20] Po-Ru Loh, Gleb Kichaev, Steven Gazal, Armin P. Schoech, and Alkes L. Price. Mixed-model  
576 association for biobank-scale datasets. *Nature Genetics*, 50(7):906–908, 2018.
- 577 [21] Tian Ge, Chia-Yen Chen, Benjamin M. Neale, Mert R. Sabuncu, and Jordan W. Smoller.  
578 Phenome-wide heritability analysis of the UK Biobank. *PLoS Genetics*, 13(4):e1006711, April  
579 2017.
- 580 [22] Gilbert Strang. *Linear Algebra and Its Applications*. Academic Press, Inc., 2nd edition, 1980.
- 581 [23] Clare Bycroft, Colin Freeman, Desislava Petkova, Gavin Band, Lloyd T. Elliott, Kevin Sharp,  
582 Allan Motyer, Damjan Vukcevic, Olivier Delaneau, Jared O’Connell, Adrian Cortes, Samantha  
583 Welsh, Alan Young, Mark Effingham, Gil McVean, Stephen Leslie, Naomi Allen, Peter Donnelly,  
584 and Jonathan Marchini. The uk biobank resource with deep phenotyping and genomic data.  
585 *Nature*, 562(7726):203–209, 2018.
- 586 [24] Luke R. Lloyd-Jones, Jian Zeng, Julia Sidorenko, Loic Yengo, Gerhard Moser, Kathryn E. Kem-  
587 per, Huanwei Wang, Zhili Zheng, Reedik Magi, Tonu Esko, Andres Metspalu, Naomi R. Wray,  
588 Michael E. Goddard, Jian Yang, and Peter M. Visscher. Improved polygenic prediction by  
589 bayesian multiple regression on summary statistics. *bioRxiv*, 2019.
- 590 [25] Sung Chun, Maxim Imakaev, Daniel Hui, Nikolaos A Patsopoulos, Benjamin M Neale, Sekar  
591 Kathiresan, Nathan O Stitzel, and Shamil R Sunyaev. Non-parametric polygenic risk prediction  
592 using partitioned gwas summary statistics. *bioRxiv*, 2019.
- 593 [26] Naomi R. Wray, Jian Yang, Ben J. Hayes, Alkes L. Price, Michael E. Goddard, and Peter M.  
594 Visscher. Pitfalls of predicting complex traits from snps. *Nature Reviews Genetics*, 14:507 EP  
595 –, 06 2013.
- 596 [27] Cathie Sudlow, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul  
597 Downey, Paul Elliott, Jane Green, Martin Landray, et al. Uk biobank: an open access resource  
598 for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS*  
599 *medicine*, 12(3):e1001779, 2015.
- 600 [28] Kevin J. Galinsky, Po-Ru Loh, Swapan Mallick, Nick J. Patterson, and Alkes L. Price. Population  
601 structure of uk biobank and ancient eurasians reveals adaptation at genes influencing blood  
602 pressure. *The American Journal of Human Genetics*, 99(5):1130–1139, 11 2016.

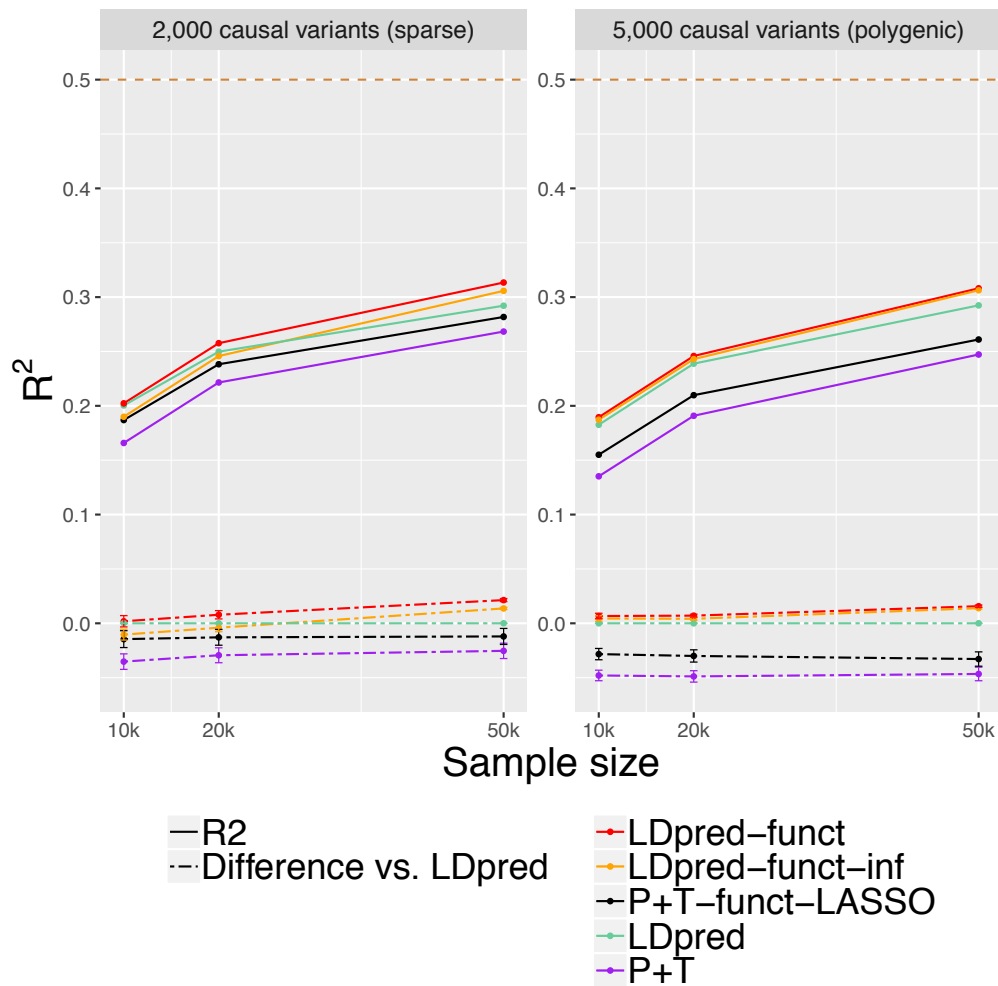
- 603 [29] Kevin J. Galinsky, Gaurav Bhatia, Po-Ru Loh, Stoyan Georgiev, Sayan Mukherjee, Nick J.  
604 Patterson, and Alkes L. Price. Fast Principal-Component Analysis Reveals Convergent Evolution  
605 of ADH1b in Europe and East Asia. *The American Journal of Human Genetics*, 98(3):456–472,  
606 March 2016.
- 607 [30] Eric Y Durand, Chuong B Do, Joanna L Mountain, and J. Michael Macpherson. Ancestry  
608 composition: A novel, efficient pipeline for ancestry deconvolution. *bioRxiv*, 2014.
- 609 [31] 1000 Genomes Project Consortium et al. A global reference for human genetic variation. *Nature*,  
610 526(7571):68, 2015.
- 611 [32] UK10K Consortium et al. The uk10k project identifies rare variants in health and disease.  
612 *Nature*, 526(7571):82, 2015.
- 613 [33] Steven Gazal, Carla Marquez-Luna, Hilary K. Finucane, and Alkes L. Price. Reconciling s-ldsc  
614 and ldak functional enrichment estimates. *Nature Genetics*, 2019.
- 615 [34] Doug Speed, Na Cai, Michael R Johnson, Sergey Nejentsev, David J Balding, UCLEB Consor-  
616 tium, et al. Reevaluation of snp heritability in complex human traits. *Nature genetics*, 49(7):986,  
617 2017.
- 618 [35] Carla Márquez-Luna, Steven Gazal, Po-Ru Loh, Nicholas Furlotte, Adam Auton, 23andMe Re-  
619 search Team, and Alkes L Price. Modeling functional enrichment improves polygenic prediction  
620 accuracy in uk biobank and 23andme data sets. *bioRxiv*, 2018.
- 621 [36] Bogdan Pasaniuc and Alkes L Price. Dissecting the genetics of complex traits using summary  
622 association statistics. *Nature Reviews Genetics*, 18(2):117, 2017.
- 623 [37] Robert Maier, Gerhard Moser, Guo-Bo Chen, Stephan Ripke, Cross-Disorder Working Group of  
624 the Psychiatric Genomics Consortium, William Coryell, James B. Potash, William A. Scheftner,  
625 Jianxin Shi, Myrna M. Weissman, Christina M. Hultman, Mikael LandÄ©n, Douglas F. Levin-  
626 son, Kenneth S. Kendler, Jordan W. Smoller, Naomi R. Wray, and S. Hong Lee. Joint analysis  
627 of psychiatric disorders increases accuracy of risk prediction for schizophrenia, bipolar disorder,  
628 and major depressive disorder. *Am. J. Hum. Genet.*, 96(2):283–294, February 2015.
- 629 [38] Robert M. Maier, Zhihong Zhu, Sang Hong Lee, Maciej Trzaskowski, Douglas M. Ruderfer,  
630 Eli A. Stahl, Stephan Ripke, Naomi R. Wray, Jian Yang, Peter M. Visscher, and Matthew R.  
631 Robinson. Improving genetic prediction by leveraging genetic correlations among human diseases  
632 and traits. *Nature Communications*, 9(1):989, 2018.

- 633 [39] George Tucker, Po-Ru Loh, Iona M. MacLeod, Ben J. Hayes, Michael E. Goddard, Bonnie  
634 Berger, and Alkes L. Price. Two-Variance-Component Model Improves Genetic Prediction in  
635 Family Datasets. *Am. J. Hum. Genet.*, 97(5):677–690, November 2015.
- 636 [40] Alicia R. Martin, Masahiro Kanai, Yoichiro Kamatani, Yukinori Okada, Benjamin M. Neale, and  
637 Mark J. Daly. Clinical use of current polygenic risk scores may exacerbate health disparities.  
638 *Nature Genetics*, 51(4):584–591, 2019.
- 639 [41] Deepti Gurdasani, Inês Barroso, Eleftheria Zeggini, and Manjinder S. Sandhu. Genomics of  
640 disease risk in globally diverse populations. *Nature Reviews Genetics*, 2019.
- 641 [42] Gleb Kichaev and Bogdan Pasaniuc. Leveraging Functional-Annotation Data in Trans-ethnic  
642 Fine-Mapping Studies. *The American Journal of Human Genetics*, 97(2):260–271, August 2015.
- 643 [43] Masahiro Kanai, Masato Akiyama, Atsushi Takahashi, Nana Matoba, Yukihide Momozawa,  
644 Masashi Ikeda, Nakao Iwata, Shiro Ikegawa, Makoto Hirata, Koichi Matsuda, Michiaki Kubo,  
645 Yukinori Okada, and Yoichiro Kamatani. Genetic analysis of quantitative traits in the Japanese  
646 population links cell types to complex human diseases. *Nature Genetics*, 50(3):390–400, March  
647 2018.
- 648 [44] Diego Calderon, Anand Bhaskar, David A. Knowles, David Golan, Towfique Raj, Audrey Q.  
649 Fu, and Jonathan K. Pritchard. Inferring Relevant Cell Types for Complex Traits by Using  
650 Single-Cell Gene Expression. *Am. J. Hum. Genet.*, 101(5):686–699, November 2017.
- 651 [45] Halit Ongen, Andrew A. Brown, Olivier Delaneau, Nikolaos I. Panousis, Alexandra C. Nica,  
652 GTEEx Consortium, and Emmanouil T. Dermitzakis. Estimating the causal tissues for complex  
653 traits and diseases. *Nat. Genet.*, 49(12):1676–1683, December 2017.
- 654 [46] Hilary K. Finucane, Yakir A. Reshef, Verner Anttila, Kamil Slowikowski, Alexander Gusev,  
655 Andrea Byrnes, Steven Gazal, Po-Ru Loh, Caleb Lareau, Noam Shores, Giulio Genovese, Arpiar  
656 Saunders, Evan Macosko, Samuela Pollack, Brainstorm Consortium, John R. B. Perry, Jason D.  
657 Buenrostro, Bradley E. Bernstein, Soumya Raychaudhuri, Steven McCarroll, Benjamin M. Neale,  
658 and Alkes L. Price. Heritability enrichment of specifically expressed genes identifies disease-  
659 relevant tissues and cell types. *Nat. Genet.*, 50(4):621–629, April 2018.
- 660 [47] Daniel Backenroth, Zihuai He, Krzysztof Kiryluk, Valentina Boeva, Lynn Pethukova, Ekta Khu-  
661 rana, Angela Christiano, Joseph D. Buxbaum, and Iuliana Ionita-Laza. FUN-LDA: A Latent  
662 Dirichlet Allocation Model for Predicting Tissue-Specific Functional Effects of Noncoding Vari-  
663 ation: Methods and Applications. *Am. J. Hum. Genet.*, 102(5):920–942, May 2018.

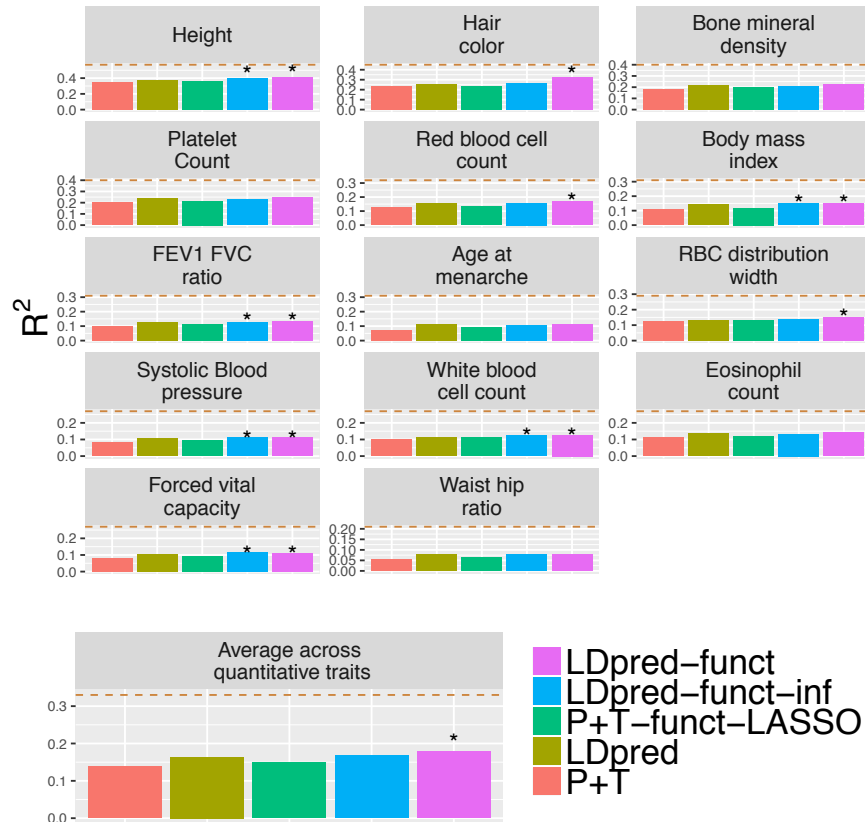
- 664 [48] Jian Zeng, Ronald de Vlaming, Yang Wu, Matthew R. Robinson, Luke R. Lloyd-Jones, Loic  
665 Yengo, Chloe X. Yap, Angli Xue, Julia Sidorenko, Allan F. McRae, Joseph E. Powell, Grant W.  
666 Montgomery, Andres Metspalu, Tonu Esko, Greg Gibson, Naomi R. Wray, Peter M. Visscher,  
667 and Jian Yang. Signatures of negative selection in the genetic architecture of human complex  
668 traits. *Nature Genetics*, 50(5):746–753, 2018.
- 669 [49] Steven Gazal, Po-Ru Loh, Hilary K. Finucane, Andrea Ganna, Armin Schoech, Shamil Sunyaev,  
670 and Alkes L. Price. Functional architecture of low-frequency variants highlights strength of  
671 negative selection across coding and non-coding annotations. *Nature Genetics*, 50(11):1600–  
672 1607, 2018.
- 673 [50] Armin P. Schoech, Daniel M. Jordan, Po-Ru Loh, Steven Gazal, Luke J. O’Connor, Daniel J.  
674 Balick, Pier F. Palamara, Hilary K. Finucane, Shamil R. Sunyaev, and Alkes L. Price. Quan-  
675 tification of frequency-dependent genetic architectures in 25 uk biobank traits reveals action of  
676 negative selection. *Nature Communications*, 10(1):790, 2019.
- 677 [51] Doug Speed and David J. Balding. Sumher better estimates the snp heritability of complex traits  
678 from summary statistics. *Nature Genetics*, 51(2):277–284, 2019.



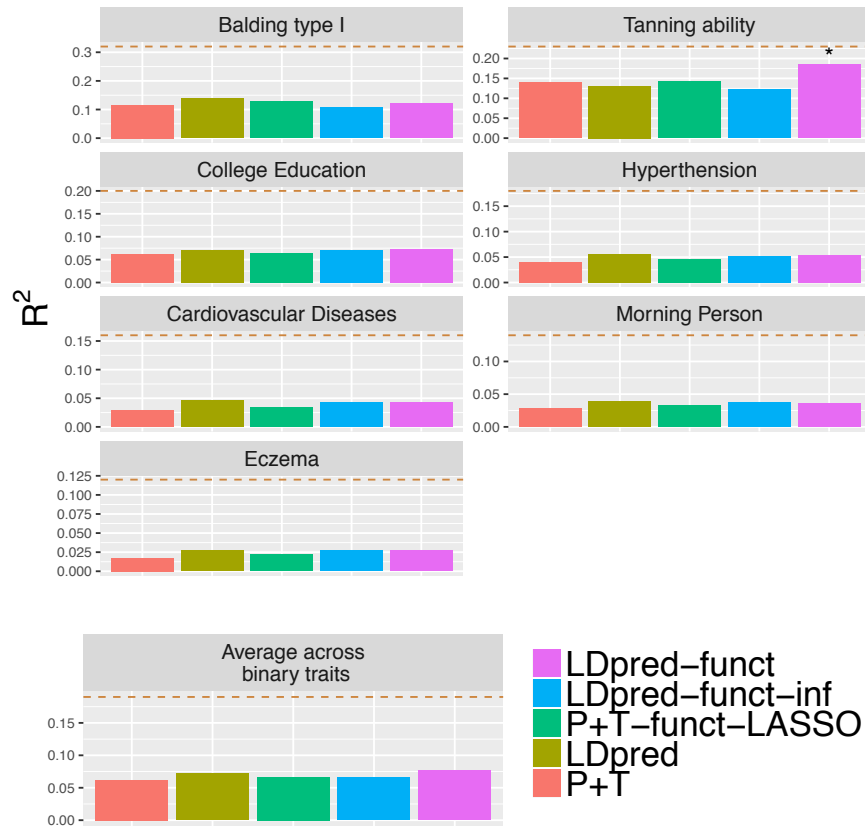
679 **Figures**



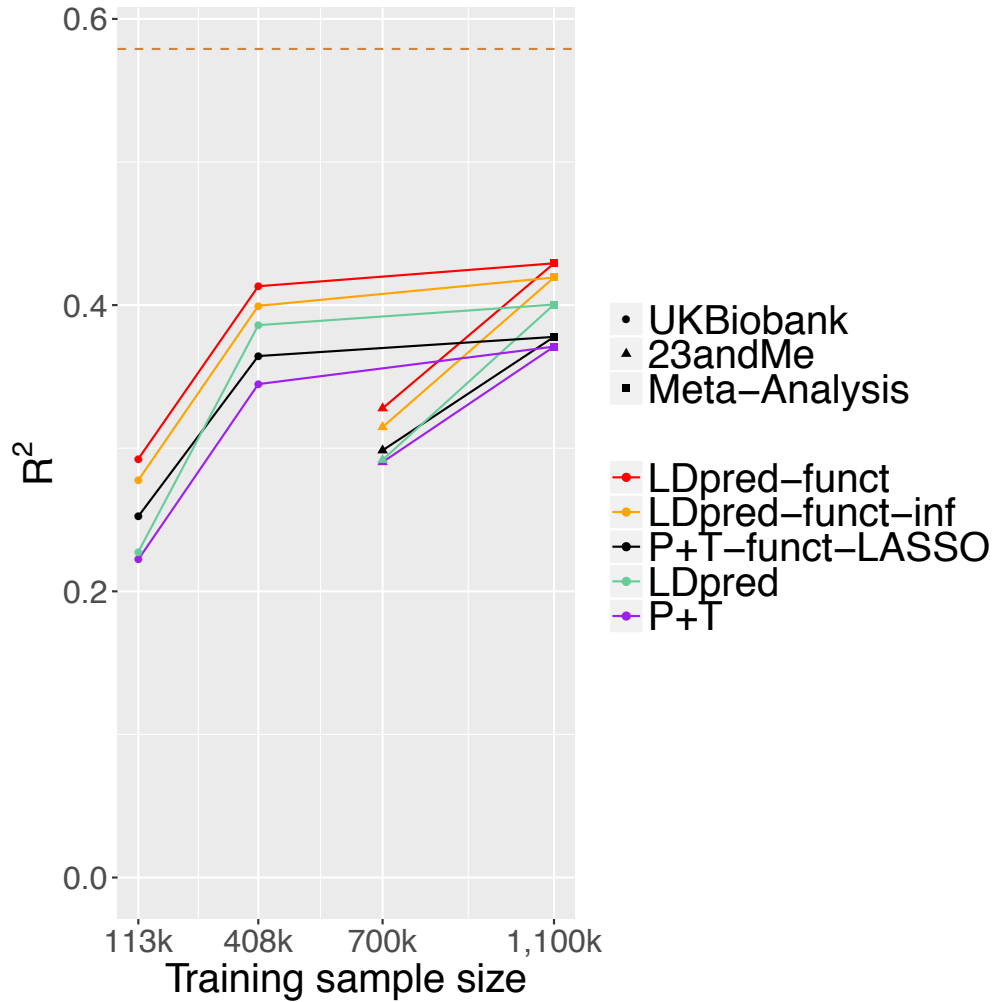
**Figure 1: Accuracy of 5 polygenic prediction methods in simulations using UK Biobank genotypes.** We report results for P+T, LDpred, P+T-funct-LASSO, LDpred-funct-inf and LDpred-funct in chromosome 1 simulations with 2,000 causal variants (sparse architecture) and 5,000 causal variants (polygenic architecture). Results are averaged across 100 simulations. Top dashed line denotes simulated SNP-heritability of 0.5. Bottom dashed lines denote differences vs. LDpred; error bars represent 95% confidence intervals. Results for other values of the number of causal variants are reported in Figure S1, and numerical results are reported in Table S3 and Table S4.



**Figure 2: Accuracy of 5 polygenic prediction methods across 14 UK Biobank quantitative traits.** We report results for P+T, LDpred, P+T-funct-LASSO, LDpred-funct-inf and LDpred-funct. Dashed lines denote estimates of SNP-heritability. Numerical results are reported in Table S11. \* denotes methods that significantly outperform LDpred ( $P < 0.05$  for difference using one-sided z-test based on block-jackknife standard error in Table S13).

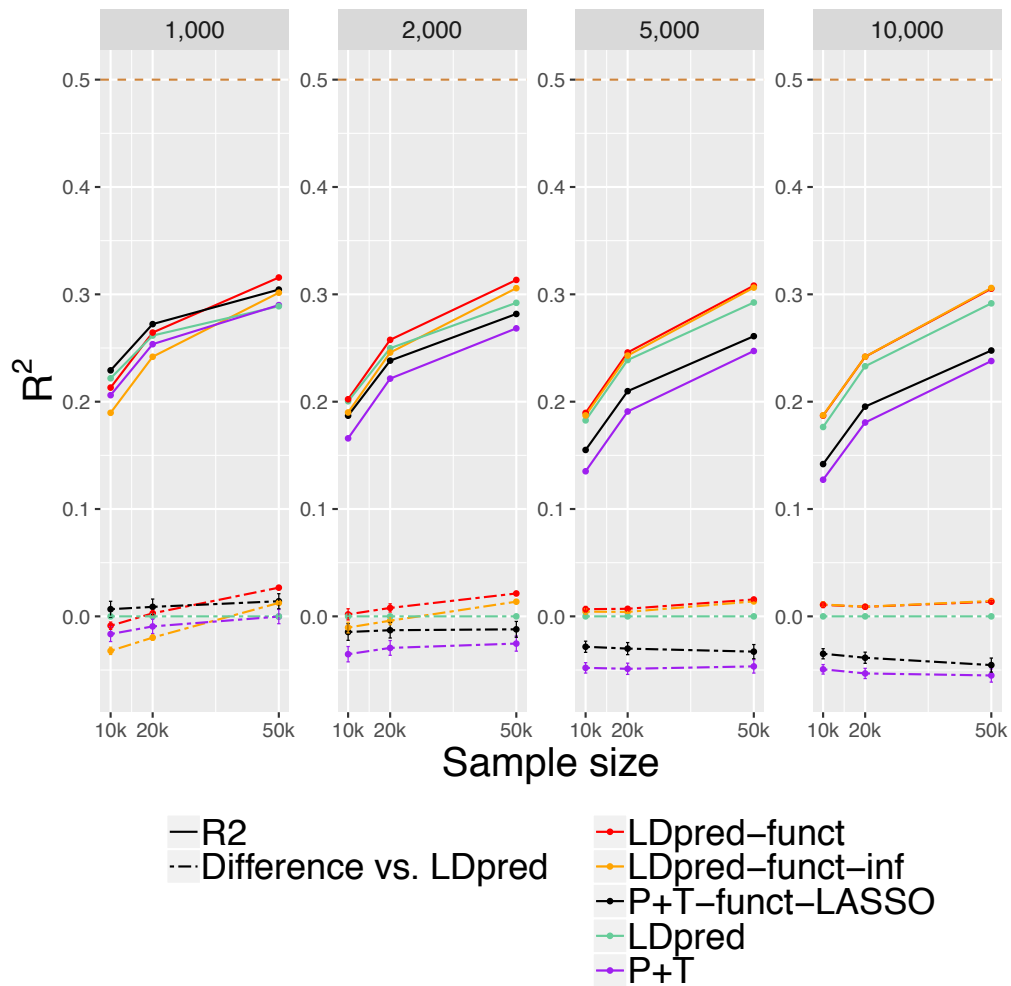


**Figure 3: Accuracy of 5 polygenic prediction methods across 7 UK Biobank binary traits.** We report results for P+T, LDpred, P+T-funct-LASSO, LDpred-funct-inf and LDpred-funct. Dashed lines denote estimates of SNP-heritability. Numerical results are reported in Table S12. \* denotes methods that significantly outperform LDpred ( $P < 0.05$  for difference using one-sided z-test based on block-jackknife standard error in Table S13).



**Figure 4: Accuracy of 5 prediction methods in height meta-analysis of UK Biobank and 23andMe cohorts.** We report results for P+T, LDpred, P+T-funct-LASSO, LDpred-funct-inf and LDpred-funct, for each of 4 training data sets: UK Biobank interim release (113,660 training samples), UK Biobank (408,092 training samples), 23andMe (698,430 training samples) and meta-analysis of UK Biobank and 23andMe (1,107,430 training samples). Nested training data sets are connected by solid lines (e.g. UK Biobank (408k) and 23andMe are both connected to Meta-Analysis, but not to each other). Dashed line denotes estimate of SNP-heritability in UK Biobank. Numerical results are reported in Table S20.

680 **Supplementary Figures**



**Figure S1: Accuracy of 5 polygenic prediction methods in simulations using UK Biobank genotypes, for 4 values of the number of causal variants.** We report results for P+T, LDpred, P+T-funct-LASSO, LDpred-funct-inf and LDpred-funct in chromosome 1 simulations with 1,000 causal variants (extremely sparse architecture), 2,000 causal variants (sparse architecture), 5,000 causal variants (polygenic architecture) and 10,000 causal variants (extremely polygenic architecture). Results are averaged across 100 simulations. Top dashed line denotes simulated SNP-heritability of 0.5. Bottom dashed lines denote differences vs. LDpred-inf; error bars represent 95% confidence intervals. Numerical results are reported in Table S3 and Table S4.

681 **Supplementary Tables**

Trait	$h_g^2$	Training N	Validation N (ancestry distribution)
1 Height	0.57	408092	25030 (43.5% Irish, 56.5% Other)
2 Hair color	0.45	403024	24773 (43.5% Irish, 56.5% Other)
3 Platelet count	0.40	395747	24277 (43.5% Irish, 56.5% Other)
4 Bone mineral density	0.40	397274	24167 (43.6% Irish, 56.4% Other)
5 Red blood cell count	0.32	396464	24305 (43.5% Irish, 56.5% Other)
6 Age at menarche	0.31	214860	13999 (39.7% Irish, 60.3% Other)
7 FEV1 FVC ratio	0.31	331786	19929 (42.5% Irish, 57.5% Other)
8 Body mass index	0.31	407667	25000 (43.5% Irish, 56.5% Other)
9 RBC distribution width	0.29	394258	24175 (43.5% Irish, 56.5% Other)
10 Forced vital capacity	0.27	331786	19929 (42.5% Irish, 57.5% Other)
11 Eosinophil count	0.27	391787	24030 (43.4% Irish, 56.6% Other)
12 White blood cell count	0.27	395835	24293 (43.5% Irish, 56.5% Other)
13 Systolic Blood pressure	0.27	376437	23127 (43.2% Irish, 56.8% Other)
14 Waist hip ratio	0.21	408196	25032 (43.5% Irish, 56.5% Other)

**Table S1: List of 14 UK Biobank quantitative traits.** We list the training sample size and validation sample size for each trait.  $h_g^2$  estimates are obtained using BOLT-LMM v2.3 using the training data set.

	Trait	$h_g^2$	Training		Validation			Prevalence
			N	Prevalence	N (ancestry distribution)			
1	Balding Type I	0.32	186506	0.32	10578 (48.9% Irish, 51.1% Other)			0.34
2	Tanning	0.23	400721	0.61	24608 (43.5% Irish, 56.5% Other)			0.60
3	College Education	0.20	405140	0.31	24749 (43.5% Irish, 56.5% Other)			0.49
4	Hypertension	0.18	408323	0.27	25041 (43.5% Irish, 56.5% Other)			0.25
5	Cardiovascular Diseases	0.16	408963	0.32	25111 (43.5% Irish, 56.5% Other)			0.29
6	Morning Person	0.14	365245	0.63	22768 (43.4% Irish, 56.6% Other)			0.58
7	Eczema	0.12	408454	0.23	25052 (43.5% Irish, 56.5% Other)			0.23

**Table S2: List of 7 UK Biobank binary traits.** We list the training sample size, validation sample size and prevalence for each trait.  $h_g^2$  estimates are obtained using BOLT-LMM v2.3 using the training data set.

# Causal variants	Model	Training sample size		
		10,000	20,000	50,000
		Average $R^2$ (s.e.)	Average $R^2$ (s.e.)	Average $R^2$ (s.e.)
1,000	P+T	0.2061 ( 0.0022)	0.2536 ( 0.0021)	0.2900 ( 0.0019)
	LDpred	0.2218 ( 0.0024)	0.2616 ( 0.0021)	0.2889 ( 0.0018)
	P+T-funct-LASSO	0.2292 ( 0.0024)	0.2723 ( 0.0024)	0.3044 ( 0.002)
	LDpred-funct-inf	0.1896 ( 0.0018)	0.2419 ( 0.0019)	0.3015 ( 0.0019)
	LDpred-funct	0.2131 ( 0.002)	0.2644 ( 0.0021)	0.3157 ( 0.002)
2,000	P+T	0.1658 ( 0.0022)	0.2215 ( 0.0026)	0.2683 ( 0.0029)
	LDpred	0.2004 ( 0.0028)	0.2498 ( 0.0023)	0.2921 ( 0.0015)
	P+T-funct-LASSO	0.1869 ( 0.0026)	0.2383 ( 0.0028)	0.2817 ( 0.0031)
	LDpred-funct-inf	0.1900 ( 0.0015)	0.2458 ( 0.0015)	0.3057 ( 0.0016)
	LDpred-funct	0.2023 ( 0.0016)	0.2576 ( 0.0016)	0.3134 ( 0.0017)
5,000	P+T	0.1352 ( 0.0016)	0.1909 ( 0.002)	0.2472 ( 0.0024)
	LDpred	0.1826 ( 0.0017)	0.2388 ( 0.0013)	0.2924 ( 0.0013)
	P+T-funct-LASSO	0.1550 ( 0.0018)	0.2098 ( 0.0021)	0.261 ( 0.0026)
	LDpred-funct-inf	0.1872 ( 0.0012)	0.243 ( 0.0013)	0.3063 ( 0.0014)
	LDpred-funct	0.1895 ( 0.0012)	0.2458 ( 0.0013)	0.3081 ( 0.0014)
10,000	P+T	0.1273 ( 0.0015)	0.1806 ( 0.002)	0.2379 ( 0.0024)
	LDpred	0.1764 ( 0.0016)	0.233 ( 0.0012)	0.2916 ( 0.0012)
	P+T-funct-LASSO	0.1419 ( 0.0017)	0.1954 ( 0.0022)	0.2477 ( 0.0026)
	LDpred-funct-inf	0.1873 ( 0.0012)	0.2419 ( 0.0012)	0.3059 ( 0.0013)
	LDpred-funct	0.1870 ( 0.0013)	0.2418 ( 0.0012)	0.3053 ( 0.0012)

**Table S3: Accuracy of 5 polygenic prediction methods in simulations using UK Biobank genotypes, for 4 values of the number of causal variants.** We report results for P+T, LDpred, P+T-funct-LASSO, LDpred-funct-inf and LDpred-funct in chromosome 1 simulations with 1,000 causal variants (extremely sparse architecture), 2,000 causal variants (sparse architecture), 5,000 causal variants (polygenic architecture) and 10,000 causal variants (extremely polygenic architecture). Results are averaged across 100 simulations. We report standard errors in parentheses.



(a)		Training sample size		
# Causal variants	Model	10,000	20,000	50,000
		Diff. $R^2$ ( <i>s.e.</i> )	Diff. $R^2$ ( <i>s.e.</i> )	Diff. $R^2$ ( <i>s.e.</i> )
1,000	P+T	0.0069 (0.0018)	0.0106 (0.0016)	0.0254 (0.0015)
	P+T-funct-LASSO	-0.0162 (0.002)	-0.0081 (0.0018)	0.011 (0.0016)
	LDpred	-0.0087 (0.0017)	0.0028 (0.0013)	0.0267 (8e-04)
	LDpred-funct-inf	0.0235 (8e-04)	0.0225 (6e-04)	0.0142 (6e-04)
	LDpred-funct	0	0	0
2,000	P+T	0.0365 (0.0019)	0.0361 (0.0022)	0.0451 (0.0026)
	P+T-funct-LASSO	0.0153 (0.0023)	0.0194 (0.0024)	0.0317 (0.0027)
	LDpred	0.0019 (0.0026)	0.0078 (0.0019)	0.0213 (7e-04)
	LDpred-funct-inf	0.0123 (5e-04)	0.0118 (5e-04)	0.0077 (4e-04)
	LDpred-funct	0	0	0
5,000	P+T	0.0544 (0.0016)	0.055 (0.0018)	0.0609 (0.0021)
	P+T-funct-LASSO	0.0345 (0.0017)	0.036 (0.0019)	0.0471 (0.0023)
	LDpred	0.0067 (0.0013)	0.007 (7e-04)	0.0157 (5e-04)
	LDpred-funct-inf	0.0023 (3e-04)	0.0029 (3e-04)	0.0018 (2e-04)
	LDpred-funct	0	0	0
10,000	P+T	0.0597 (0.0016)	0.0612 (0.002)	0.0674 (0.0024)
	P+T-funct-LASSO	0.0451 (0.0017)	0.0464 (0.0022)	0.0576 (0.0026)
	LDpred	0.0107 (0.0013)	0.0089 (5e-04)	0.0136 (5e-04)
	LDpred-funct-inf	-4e-04 (2e-04)	-1e-04 (2e-04)	-7e-04 (2e-04)
	LDpred-funct	0	0	0

(b)		Training sample size		
# Causal variants	Model	10,000	20,000	50,000
		Diff. $R^2$ ( <i>s.e.</i> )	Diff. $R^2$ ( <i>s.e.</i> )	Diff. $R^2$ ( <i>s.e.</i> )
1,000	P+T	-0.0165 (0.0035)	-0.0094 (0.0034)	-2e-04 (0.0033)
	LDpred	0	0	0
	P+T-funct-LASSO	0.0067 (0.0037)	0.0088 (0.0037)	0.0141 (0.0035)
	LDpred-funct-inf	-0.0321 (0.0017)	-0.0198 (0.0012)	0.0125 (6e-04)
	LDpred-funct	-0.0087 (0.0017)	0.0028 (0.0013)	0.0267 (8e-04)
2,000	P+T	-0.0352 (0.0036)	-0.0294 (0.0035)	-0.0254 (0.0036)
	LDpred	0	0	0
	P+T-funct-LASSO	-0.0146 (0.0039)	-0.0129 (0.0036)	-0.0121 (0.0037)
	LDpred-funct-inf	-0.0104 (0.0025)	-0.004 (0.0019)	0.0137 (5e-04)
	LDpred-funct	0.0019 (0.0026)	0.0078 (0.0019)	0.0213 (7e-04)
5,000	P+T	-0.048 (0.0024)	-0.0488 (0.0026)	-0.0466 (0.0031)
	LDpred	0	0	0
	P+T-funct-LASSO	-0.0283 (0.0026)	-0.03 (0.0028)	-0.0329 (0.0033)
	LDpred-funct-inf	0.0044 (0.0013)	0.0041 (7e-04)	0.0139 (4e-04)
	LDpred-funct	0.0067 (0.0013)	0.007 (7e-04)	0.0157 (5e-04)
10,000	P+T	-0.0493 (0.0022)	-0.0532 (0.0024)	-0.0551 (0.0031)
	LDpred	0	0	0
	P+T-funct-LASSO	-0.0348 (0.0024)	-0.0386 (0.0026)	-0.0454 (0.0033)
	LDpred-funct-inf	0.0111 (0.0012)	0.009 (4e-04)	0.0143 (5e-04)
	LDpred-funct	0.0107 (0.0013)	0.0089 (5e-04)	0.0136 (5e-04)

**Table S4: Differences between polygenic prediction methods in simulations using UK Biobank genotypes, for 4 values of the number of causal variants.** We report results for P+T, LDpred, P+T-funct-LASSO, LDpred-funct-inf and LDpred-funct in chromosome 1 simulations with 1,000 causal variants (extremely sparse architecture), 2,000 causal variants (sparse architecture), 5,000 causal variants (polygenic architecture) and 10,000 causal variants (extremely polygenic architecture). Results are averaged across 100 simulations. We report standard errors in parentheses. (a) Difference between  $R^2$  for LDpred-funct vs.  $R^2$  for each method. (b) Difference between  $R^2$  for each method vs.  $R^2$  for LDpred.

# Causal	Training sample size		
	10,000	20,000	50,000
1,000	0.03	0.1	1
2,000	0.03	0.1	1
5,000	0.03	0.1	1
10,000	0.1	0.3	1

**Table S5: Model parameter values for LDpred in simulations.** We report the optimal value of  $p$  which is the fraction of non-zero effects in the prior, and LD-radius assumed was 2000 SNPs. The analyses from LDpred exclude long-range LD regions reported in ref. 23.

# Causal		Training sample size		
		10,000	20,000	50,000
1,000	P+T	0.0001	0.0001	0.0001
	P+T-funct-LASSO HP SNP Set	0.1000	0.1000	0.3000
	P+T-funct-LASSO LP SNP Set	0.0100	0.0100	0.0100
2,000	P+T	0.0010	0.0010	0.0010
	P+T-funct-LASSO HP SNP Set	0.1000	0.1000	0.3000
	P+T-funct-LASSO LP SNP Set	0.0100	0.0100	0.0100
5,000	P+T	0.0100	0.0100	0.0100
	P+T-funct-LASSO HP SNP Set	0.3000	0.3000	0.3000
	P+T-funct-LASSO LP SNP Set	0.1000	0.1000	0.1000
10,000	P+T	0.1000	0.1000	0.0100
	P+T-funct-LASSO HP SNP Set	0.3000	0.3000	1.0000
	P+T-funct-LASSO LP SNP Set	0.1000	0.1000	0.1000

**Table S6: Model parameter values for P+T and P+T-funct-LASSO in simulated traits.** We report the optimal p-value threshold for Pruning + Thresholding (P+T), optimal p-value threshold for P+T-funct-LASSO high prior SNP (HP) set and optimal p-value threshold for P+T-funct-LASSO low prior SNP (LP) set. Optimal  $R_{LD}^2$  values was 0.1.

# Causal variants	Model	Training sample size		
		10,000	20,000	50,000
		Average $R^2$ (s.e.)	Average $R^2$ (s.e.)	Average $R^2$ (s.e.)
1,000	P+T	0.9371 ( 0.0282)	0.9806 ( 0.0294)	0.8189 ( 0.0486)
	LDpred	0.992 ( 0.0146)	0.947 ( 0.0083)	0.8521 ( 0.004)
	P+T-funct-LASSO	1.5051 ( 0.0589)	1.3703 ( 0.0386)	1.0151 ( 0.075)
	LDpred-funct-inf	0.4708 ( 0.0025)	0.454 ( 0.002)	0.4345 ( 0.0024)
	LDpred-funct	0.9803 ( 6e-04)	0.9847 ( 4e-04)	0.9877 ( 4e-04)
2,000	P+T	0.7644 ( 0.0309)	0.791 ( 0.0257)	0.7976 ( 0.0209)
	LDpred	0.9688 ( 0.037)	0.9346 ( 0.0257)	0.8483 ( 0.0044)
	P+T-funct-LASSO	1.3572 ( 0.0382)	1.2138 ( 0.0544)	1.0448 ( 0.0284)
	LDpred-funct-inf	0.4656 ( 0.004)	0.457 ( 0.0028)	0.4396 ( 0.0021)
	LDpred-funct	0.9787 ( 0.001)	0.9837 ( 7e-04)	0.9882 ( 4e-04)
5,000	P+T	0.4546 ( 0.0207)	0.5954 ( 0.0172)	0.6728 ( 0.0158)
	LDpred	0.9984 ( 0.0067)	0.9671 ( 0.0071)	0.8538 ( 0.0044)
	P+T-funct-LASSO	0.8085 ( 0.0267)	0.8994 ( 0.012)	0.909 ( 0.0213)
	LDpred-funct-inf	0.47 ( 0.0035)	0.4584 ( 0.0023)	0.4424 ( 0.0015)
	LDpred-funct	0.9776 ( 9e-04)	0.9839 ( 5e-04)	0.9881 ( 4e-04)
10,000	P+T	0.3196 ( 0.0136)	0.4655 ( 0.016)	0.586 ( 0.0116)
	LDpred	0.9903 ( 0.0156)	0.9449 ( 0.0059)	0.847 ( 0.0041)
	P+T-funct-LASSO	0.6824 ( 0.0182)	0.8142 ( 0.0182)	0.8178 ( 0.017)
	LDpred-funct-inf	0.4654 ( 0.0028)	0.4528 ( 0.0025)	0.4365 ( 0.0024)
	LDpred-funct	0.9761 ( 7e-04)	0.9824 ( 6e-04)	0.9874 ( 4e-04)

**Table S7: Calibration of 5 polygenic prediction methods in simulations using UK Biobank genotypes, for 4 values of the number of causal variants.** We report calibration slopes for P+T, LDpred, P+T-funct-LASSO, LDpred-funct-inf and LDpred-funct in chromosome 1 simulations with 1,000 causal variants (extremely sparse architecture), 2,000 causal variants (sparse architecture), 5,000 causal variants (polygenic architecture) and 10,000 causal variants (extremely polygenic architecture). Results are averaged across 100 simulations.

# Causal variants	Model	Training sample size		
		10,000	20,000	50,000
		Average $R^2$ ( <i>s.e.</i> )	Average $R^2$ ( <i>s.e.</i> )	Average $R^2$ ( <i>s.e.</i> )
1,000	LDpred-funct-inf	0.1896 ( 0.0018)	0.2419 ( 0.0019)	0.3015 ( 0.0019)
	LDpred-funct-inf-5	0.208 ( 0.002)	0.2585 ( 0.002)	0.3104 ( 0.0019)
	LDpred-funct-inf-10	0.2101 ( 0.002)	0.261 ( 0.002)	0.3124 ( 0.002)
	LDpred-funct-inf-20	0.2116 ( 0.002)	0.263 ( 0.002)	0.314 ( 0.002)
	LDpred-funct-inf-30	0.2126 ( 0.002)	0.2638 ( 0.002)	0.315 ( 0.002)
	LDpred-funct-inf-40	0.2131 ( 0.002)	0.2644 ( 0.0021)	0.3157 ( 0.002)
	LDpred-funct-inf-50	0.2141 ( 0.002)	0.2652 ( 0.0021)	0.3161 ( 0.002)
	LDpred-funct-inf-60	0.2145 ( 0.0021)	0.2655 ( 0.0021)	0.3172 ( 0.002)
	LDpred-funct-inf-70	0.2157 ( 0.0021)	0.266 ( 0.0021)	0.317 ( 0.0021)
	LDpred-funct-inf-80	0.216 ( 0.002)	0.2665 ( 0.0021)	0.3173 ( 0.0021)
LDpred-funct-inf-90	0.2164 ( 0.0021)	0.2667 ( 0.0021)	0.3176 ( 0.0021)	
LDpred-funct-inf-100	0.2165 ( 0.0021)	0.267 ( 0.0021)	0.3174 ( 0.0021)	
2,000	LDpred-funct-inf	0.1900 ( 0.0015)	0.2458 ( 0.0015)	0.3057 ( 0.0016)
	LDpred-funct-inf-5	0.1994 ( 0.0016)	0.254 ( 0.0016)	0.3101 ( 0.0016)
	LDpred-funct-inf-10	0.2005 ( 0.0016)	0.2554 ( 0.0016)	0.3113 ( 0.0017)
	LDpred-funct-inf-20	0.2016 ( 0.0016)	0.2566 ( 0.0016)	0.3124 ( 0.0017)
	LDpred-funct-inf-30	0.2018 ( 0.0016)	0.2572 ( 0.0016)	0.3129 ( 0.0017)
	LDpred-funct-inf-40	0.2023 ( 0.0016)	0.2576 ( 0.0016)	0.3134 ( 0.0017)
	LDpred-funct-inf-50	0.2023 ( 0.0016)	0.2575 ( 0.0016)	0.3136 ( 0.0017)
	LDpred-funct-inf-60	0.2025 ( 0.0016)	0.258 ( 0.0017)	0.3137 ( 0.0017)
	LDpred-funct-inf-70	0.2027 ( 0.0016)	0.2579 ( 0.0017)	0.3135 ( 0.0017)
	LDpred-funct-inf-80	0.2031 ( 0.0016)	0.2583 ( 0.0017)	0.3133 ( 0.0017)
LDpred-funct-inf-90	0.2028 ( 0.0016)	0.2579 ( 0.0017)	0.3134 ( 0.0018)	
LDpred-funct-inf-100	0.2031 ( 0.0016)	0.2582 ( 0.0017)	0.313 ( 0.0018)	
5,000	LDpred-funct-inf	0.1872 ( 0.0012)	0.243 ( 0.0013)	0.3063 ( 0.0014)
	LDpred-funct-inf-5	0.1895 ( 0.0012)	0.2451 ( 0.0013)	0.3075 ( 0.0014)
	LDpred-funct-inf-10	0.1898 ( 0.0012)	0.2456 ( 0.0013)	0.3079 ( 0.0014)
	LDpred-funct-inf-20	0.1897 ( 0.0012)	0.2461 ( 0.0013)	0.3083 ( 0.0014)
	LDpred-funct-inf-30	0.1898 ( 0.0012)	0.2461 ( 0.0013)	0.3084 ( 0.0014)
	LDpred-funct-inf-40	0.1895 ( 0.0012)	0.2458 ( 0.0013)	0.3081 ( 0.0014)
	LDpred-funct-inf-50	0.1894 ( 0.0012)	0.2457 ( 0.0013)	0.3081 ( 0.0014)
	LDpred-funct-inf-60	0.1893 ( 0.0012)	0.2454 ( 0.0013)	0.3077 ( 0.0014)
	LDpred-funct-inf-70	0.1891 ( 0.0012)	0.245 ( 0.0013)	0.3073 ( 0.0014)
	LDpred-funct-inf-80	0.1888 ( 0.0012)	0.2447 ( 0.0013)	0.3071 ( 0.0014)
LDpred-funct-inf-90	0.1885 ( 0.0012)	0.2444 ( 0.0013)	0.3066 ( 0.0014)	
LDpred-funct-inf-100	0.188 ( 0.0012)	0.244 ( 0.0013)	0.3062 ( 0.0014)	
10,000	LDpred-funct-inf	0.1873 ( 0.0012)	0.2419 ( 0.0012)	0.3059 ( 0.0013)
	LDpred-funct-inf-5	0.1883 ( 0.0012)	0.2428 ( 0.0012)	0.3064 ( 0.0013)
	LDpred-funct-inf-10	0.1882 ( 0.0012)	0.2428 ( 0.0012)	0.3064 ( 0.0012)
	LDpred-funct-inf-20	0.1878 ( 0.0012)	0.2427 ( 0.0012)	0.3061 ( 0.0012)
	LDpred-funct-inf-30	0.1873 ( 0.0013)	0.2422 ( 0.0012)	0.3056 ( 0.0013)
	LDpred-funct-inf-40	0.187 ( 0.0013)	0.2418 ( 0.0012)	0.3053 ( 0.0012)
	LDpred-funct-inf-50	0.1865 ( 0.0012)	0.2414 ( 0.0012)	0.3049 ( 0.0013)
	LDpred-funct-inf-60	0.186 ( 0.0013)	0.2409 ( 0.0012)	0.3043 ( 0.0013)
	LDpred-funct-inf-70	0.1855 ( 0.0013)	0.2406 ( 0.0012)	0.3039 ( 0.0013)
	LDpred-funct-inf-80	0.1851 ( 0.0012)	0.2399 ( 0.0012)	0.3036 ( 0.0012)
LDpred-funct-inf-90	0.1846 ( 0.0013)	0.2393 ( 0.0012)	0.3027 ( 0.0013)	
LDpred-funct-inf-100	0.1841 ( 0.0013)	0.2387 ( 0.0012)	0.3027 ( 0.0013)	

**Table S8: Sensitivity of LDpred-funct results to number of bins used for regularization in simulations using UK Biobank genotypes.** We report results with the number of posterior mean causal effect size bins used for regularization ( $K$ ) set to 10, 20, 50 or 100. LDpred-funct- $K$  denotes each respective value of  $K$ . We also report results for LDpred-funct-inf, which is identical to LDpred-funct with  $K$  set to 1. Results are averaged across 100 simulations. We report standard errors in parentheses.

# Causal variants	Model	Training sample size		
		10,000	20,000	50,000
		Average $R^2$ (s.e.)	Average $R^2$ (s.e.)	Average $R^2$ (s.e.)
1,000	LDpred-funct-inf	0.1896 ( 0.0018)	0.2419 ( 0.0019)	0.3015 ( 0.0019)
	LDpred-funct	0.2131 ( 0.002)	0.2644 ( 0.0021)	0.3157 ( 0.002)
	LDpred-funct-inf-cheat	0.1926 ( 0.0018)	0.2456 ( 0.0019)	0.3074 ( 0.002)
	LDpred-funct-cheat	0.2221 ( 0.0021)	0.2714 ( 0.0022)	0.3228 ( 0.0021)
2,000	LDpred-funct-inf	0.1900 ( 0.0015)	0.2458 ( 0.0015)	0.3057 ( 0.0016)
	LDpred-funct	0.2023 ( 0.0016)	0.2576 ( 0.0016)	0.3134 ( 0.0017)
	LDpred-funct-inf-cheat	0.1943 ( 0.0015)	0.2498 ( 0.0016)	0.3108 ( 0.0016)
	LDpred-funct-cheat	0.2109 ( 0.0016)	0.2646 ( 0.0017)	0.3193 ( 0.0017)
5,000	LDpred-funct-inf	0.1872 ( 0.0012)	0.243 ( 0.0013)	0.3063 ( 0.0014)
	LDpred-funct	0.1895 ( 0.0012)	0.2458 ( 0.0013)	0.3081 ( 0.0014)
	LDpred-funct-inf-cheat	0.1928 ( 0.0013)	0.2479 ( 0.0013)	0.3102 ( 0.0014)
	LDpred-funct-cheat	0.1972 ( 0.0014)	0.252 ( 0.0013)	0.3121 ( 0.0014)
10,000	LDpred-funct-inf	0.1873 ( 0.0012)	0.2419 ( 0.0012)	0.3059 ( 0.0013)
	LDpred-funct	0.1870 ( 0.0013)	0.2418 ( 0.0012)	0.3053 ( 0.0012)
	LDpred-funct-inf-cheat	0.1937 ( 0.0012)	0.2474 ( 0.0012)	0.3097 ( 0.0012)
	LDpred-funct-cheat	0.194 ( 0.0013)	0.2482 ( 0.0013)	0.3096 ( 0.0013)

**Table S9: Accuracy of LDpred-funct method in simulations using UK Biobank genotypes under different BaselineLD estimates, for 4 values of the number of causal variants.** LDpred-funct-cheat refers to a "cheating" version of LDpred-funct that utilized the true baseline-LD model parameters used to simulate the data. Results are averaged across 100 simulations.

	Trait	Training $N$	$h_g^2$	$c$	bins
1	Height	408092	0.57	0.45	100
2	Hair color	403024	0.45	0.22	100
3	Platelet count	395747	0.40	0.29	88
4	Bone mineral density	397274	0.40	0.26	87
5	Red blood cell count	396464	0.32	0.21	70
6	Age at menarche	214860	0.31	0.20	40
7	FEV1 FVC ratio	331786	0.31	0.24	56
8	Body mass index	407667	0.31	0.27	70
9	RBC distribution width	394258	0.29	0.20	63
10	Eosinophil count	391787	0.27	0.18	60
11	Forced vital capacity	331786	0.27	0.22	50
12	White blood cell count	395835	0.27	0.21	60
13	Systolic Blood pressure	376437	0.27	0.21	56
14	Waist hip ratio	408196	0.21	0.15	48
1	Balding type I	186506	0.32	0.11	31
2	Tanning ability	400721	0.23	0.09	53
3	College Education	405140	0.20	0.15	45
4	Hypertension	408323	0.18	0.14	41
5	Cardiovascular Diseases	408963	0.16	0.12	37
6	Morning Person	365245	0.14	0.11	29
7	Eczema	408454	0.12	0.09	27

**Table S10: Parameter values for 21 UK Biobank traits.** The 14 quantitative traits are listed first, followed by the 7 binary traits. For each trait, we list the training sample size,  $h_g^2$  estimate (from BOLT-LMM v2.3; used by LDpred, LDpred-funct-inf and LDpred-funct), the  $c$  parameter (used by LDpred-funct-inf and LDpred-funct) and number of bins for LDpred-funct.

Trait	h2g	P+T	LDpred	P+T-funct-LASSO	LDpred-funct-inf	LDpred-funct
1 Height	0.57	0.3462 (0.0164)	0.3763 (0.0193)	0.3667 (0.0167)	0.4003 (0.0194)	0.4128 (0.0261)
2 Hair color	0.45	0.2339 (0.086)	0.2519 (0.1072)	0.2389 (0.0844)	0.2624 (0.1096)	0.329 (0.1358)
3 Platelet count	0.40	0.1994 (0.0192)	0.2392 (0.024)	0.215 (0.0203)	0.2315 (0.0201)	0.246 (0.0269)
4 Bone mineral density	0.40	0.1871 (0.0177)	0.2188 (0.0219)	0.1993 (0.0178)	0.2137 (0.0188)	0.2256 (0.025)
5 Red blood cell count	0.32	0.1247 (0.0117)	0.1526 (0.0159)	0.1326 (0.0123)	0.1571 (0.0139)	0.1659 (0.0202)
6 Age at menarche	0.31	0.0747 (0.0076)	0.1108 (0.0098)	0.0899 (0.0087)	0.1079 (0.0089)	0.1122 (0.0183)
7 FEV1 FVC ratio	0.31	0.1029 (0.0083)	0.125 (0.0099)	0.1142 (0.0089)	0.1311 (0.0091)	0.133 (0.017)
8 Body mass index	0.31	0.1087 (0.0057)	0.1446 (0.0074)	0.1189 (0.0064)	0.1508 (0.0071)	0.1499 (0.0151)
9 RBC distribution width	0.29	0.1237 (0.0123)	0.1324 (0.0151)	0.1346 (0.013)	0.1421 (0.0147)	0.1533 (0.0202)
10 Forced vital capacity	0.27	0.0817 (0.0059)	0.1072 (0.0071)	0.0935 (0.0062)	0.1145 (0.0067)	0.1134 (0.0148)
11 Eosinophil count	0.27	0.1131 (0.0097)	0.1359 (0.0239)	0.1189 (0.0103)	0.1335 (0.0126)	0.1409 (0.0191)
12 White blood cell count	0.27	0.0994 (0.0078)	0.1143 (0.0095)	0.1109 (0.0085)	0.1239 (0.0093)	0.127 (0.0161)
13 Systolic Blood pressure	0.27	0.0802 (0.0061)	0.1049 (0.0067)	0.0919 (0.0066)	0.1114 (0.0064)	0.1112 (0.0133)
14 Waist hip ratio	0.21	0.0567 (0.0045)	0.0762 (0.007)	0.0645 (0.0049)	0.0793 (0.005)	0.0806 (0.0116)
15 Average across quantitative traits	0.33	0.1380 (0.0107)	0.1636 (0.0105)	0.1493 (0.0082)	0.1685 (0.0101)	0.1786 (0.0117)

**Table S11: Accuracy of 5 polygenic prediction methods across 14 UK Biobank quantitative traits.** We report results for P+T, LDpred, P+T-funct-LASSO, LDpred-funct-inf and LDpred-funct. Optimal parameters for each method are reported in Table S16, Table S15 and Table S10. We report block jackknife standard error over 200 equally sized blocks of adjacent SNPs.



Trait	h2g	P+T	LDpred	P+T-funct-LASSO	LDpred-funct-inf	LDpred-funct
1 Balding type I	0.32	0.1158 (0.015)	0.138 (0.0653)	0.1269 (0.0157)	0.1075 (0.0132)	0.1221 (0.0235)
2 Tanning ability	0.23	0.1405 (0.0516)	0.1308 (0.0678)	0.143 (0.0446)	0.1229 (0.0631)	0.1842 (0.0784)
3 College Education	0.20	0.0612 (0.0057)	0.0699 (0.0086)	0.0637 (0.006)	0.0716 (0.0059)	0.0728 (0.0109)
4 Hyperthension	0.18	0.0403 (0.0038)	0.0551 (0.0054)	0.0458 (0.0044)	0.0523 (0.0043)	0.0534 (0.0094)
5 Cardiovascular Diseases	0.16	0.0282 (0.0028)	0.0457 (0.0078)	0.0333 (0.0034)	0.0423 (0.0037)	0.0427 (0.0084)
6 Morning Person	0.14	0.0289 (0.0027)	0.0385 (0.0046)	0.0333 (0.0029)	0.0372 (0.0032)	0.0365 (0.008)
7 Eczema	0.12	0.0172 (0.0023)	0.0273 (0.0121)	0.0222 (0.0029)	0.0274 (0.0026)	0.0272 (0.0064)
8 Average across binary traits	0.19	0.0617 (0.0104)	0.0722 (0.0139)	0.0669 (0.0075)	0.0659 (0.0096)	0.0770 (0.0119)

**Table S12: Accuracy of 5 polygenic prediction methods across 7 UK Biobank binary traits.** We report results for P+T, LDpred, P+T-funct-LASSO, LDpred-funct-inf and LDpred-funct. Optimal parameters for each method are reported in Table S16, Table S15 and Table S10. We report block jackknife standard error over 200 equally sized blocks of adjacent SNPs.

Trait	$h_g^2$	P+T	P+T-funct-LASSO	LDpred	LDpred-funct-inf	LDpred-funct
1 Height	0.575	-0.029 (0.0075)	-0.0077 (0.0073)	0	0.0238 (0.0066)	0.036 (0.0069)
2 Hair color	0.446	-0.0107 (0.0252)	-0.0099 (0.0298)	0	0.0103 (0.0127)	0.0756 (0.0378)
3 Platelet count	0.401	-0.0381 (0.007)	-0.0204 (0.0071)	0	-0.0077 (0.0062)	0.0058 (0.0059)
4 Bone mineral density	0.398	-0.0294 (0.0106)	-0.0163 (0.0085)	0	-0.005 (0.0077)	0.0066 (0.007)
5 Balding type I	0.323	-0.0245 (0.0561)	-0.0101 (0.0575)	0	-0.0296 (0.0574)	-0.0151 (0.0568)
6 Red blood cell count	0.319	-0.0294 (0.0105)	-0.0216 (0.0066)	0	0.0045 (0.0047)	0.0135 (0.0043)
7 Age at menarche	0.313	-0.0396 (0.0047)	-0.0233 (0.0038)	0	-0.0025 (0.0035)	7e-04 (0.0035)
8 FEV1 FVC ratio	0.309	-0.0254 (0.0089)	-0.0125 (0.0037)	0	0.0061 (0.0036)	0.0082 (0.004)
9 Body mass index	0.307	-0.0375 (0.0034)	-0.0259 (0.003)	0	0.0062 (0.0023)	0.005 (0.0025)
10 RBC distribution width	0.286	-0.0144 (0.007)	-0.0024 (0.005)	0	0.0098 (0.0076)	0.0209 (0.0075)
11 Forced vital capacity	0.274	-0.0308 (0.0044)	-0.0178 (0.0033)	0	0.0073 (0.0027)	0.0061 (0.0027)
12 Eosinophil count	0.274	-0.026 (0.0174)	-0.0188 (0.0179)	0	-0.0023 (0.0184)	0.0053 (0.0181)
13 White blood cell count	0.273	-0.0184 (0.0044)	-0.0058 (0.0032)	0	0.0095 (0.0034)	0.0126 (0.0043)
14 Systolic Blood pressure	0.267	-0.023 (0.0038)	-0.0082 (0.0027)	0	0.0064 (0.0021)	0.0057 (0.0021)
15 Tanning ability	0.235	-0.0026 (0.0194)	-9e-04 (0.0283)	0	-0.0078 (0.014)	0.0534 (0.0297)
16 Waist hip ratio	0.21	-0.0197 (0.0049)	-0.0115 (0.0045)	0	0.0032 (0.0042)	0.0044 (0.0043)
17 College Education	0.198	-0.0086 (0.0063)	-0.006 (0.0062)	0	0.0019 (0.0059)	0.0031 (0.0061)
18 Hypertension	0.179	-0.0148 (0.0026)	-0.0082 (0.0021)	0	-0.0027 (0.0021)	-0.0016 (0.0022)
19 Cardiovascular Diseases	0.16	-0.0181 (0.0066)	-0.0121 (0.0061)	0	-0.0034 (0.0063)	-0.0029 (0.0062)
20 Morning Person	0.137	-0.0123 (0.0032)	-0.0077 (0.0028)	0	-0.0013 (0.0027)	-0.002 (0.0028)
21 Eczema	0.118	-0.0124 (0.0112)	-0.0061 (0.0109)	0	-0.0015 (0.0111)	-0.001 (0.011)
22 Average across traits	0.286	-0.0221 (0.0042)	-0.0121 (0.0045)	0	0.0012 (0.0037)	0.0114 (0.0045)

**Table S13: Absolute differences between polygenic prediction methods across 21 UK Biobank traits.** We report results for P+T, LDpred, P+T-funct-LASSO, LDpred-funct-inf and LDpred-funct. We report the difference between prediction  $R^2$  for each method vs. prediction  $R^2$  for LDpred. Block-jackknife standard errors are reported in parentheses.

	Method	Average $R^2$
1	P+T	0.1126
2	LDpred	0.1331
3	P+T-funct-LASSO	0.1218
4	LDpred-funct-inf	0.1343
5	LDpred-funct	0.1447
6	LDpred-inf	0.1133
7	LDpred (without excluding long-range LD regions)	0.0839
8	LDpred (typed SNPs only)	0.1299
9	LDpred-funct-inf (typed SNPs only)	0.1135
10	LDpred-funct (typed SNPs only)	0.1209
11	P+T-funct-LASSO-weighted	0.1231
12	P+T-funct-LASSO (5%)	0.1219
13	LDpred-funct-inf (meta31)	0.1303
14	LDpred-funct-inf (baseline)	0.1313
15	LDpred-funct (baseline)	0.1411
16	LDpred-funct-inf(QCfilters)	0.1339
17	LDpred-funct-inf(UK10K)	0.1354
18	LDpred-funct-inf(UK10K, baseline-LD+LDAK)	0.1350
19	AnnoPred	0.1413
20	LDpred-funct-inf (Baseline-LD v2.1)	0.1360
21	LDpred-funct (Baseline-LD v2.1)	0.1469

**Table S14: Accuracy of secondary polygenic prediction methods across 21 UK Biobank traits.**

For each method, we report the average prediction  $R^2$  across 21 UK Biobank traits. Rows 1-5 correspond to the "Average across traits" panel of Figure 2. Row 6 correspond to the average prediction  $R^2$  from LDpred-inf. Row 7 correspond to the average prediction  $R^2$  from LDpred that includes SNPs from long-range LD regions. Rows 8-10 are methods that analyze only genotyped SNPs (601,728 genotyped SNPs after QC). Rows 11-12 are slightly modified versions of P+T-funct-LASSO. Row 13 uses baseline-LD model functional enrichments that were meta-analyzed across 31 traits. Row 14-15 uses the baseline model, instead of the baseline-LD model. Row 16 restricts the baseline-LD model to the 6,334,603 SNPs that passed QC filters and were used for prediction. Row 17 infers baseline-LD model parameters using UK10K SNPs, instead of 1000 Genomes SNPs. Row 18 uses UK10K SNPs and uses the baseline-LD+LDAK model, instead of the baseline-LD model. Row 19 corresponds to the average prediction  $R^2$  from AnnoPred. Row 20 corresponds to the average prediction  $R^2$  for LDpred-funct-inf using baseline-LD model v2.1 (instead of baseline-LD model v1.1, which is used in our main analyses). Row 21 corresponds to the average prediction  $R^2$  for LDpred-funct using baseline-LD model v2.1 (instead of baseline-LD model v1.1, which is used in our main analyses).

	Trait	$h_g^2$	$p$
1	Height	0.57	0.3000
2	Hair color	0.45	0.3000
3	Platelet count	0.40	0.1000
4	Bone mineral density	0.40	0.1000
5	Balding type I	0.32	0.0100
6	Red blood cell count	0.32	0.1000
7	Age at menarche	0.31	0.0300
8	FEV1 FVC ratio	0.31	0.1000
9	Body mass index	0.31	0.1000
10	RBC distribution width	0.29	0.1000
11	Forced vital capacity	0.27	0.0300
12	Eosinophil count	0.27	0.0300
13	White blood cell count	0.27	0.1000
14	Systolic Blood pressure	0.27	0.1000
15	Tanning ability	0.23	0.1000
16	Waist hip ratio	0.21	0.0300
17	College Education	0.20	0.0300
18	Hypertension	0.18	0.0300
19	Cardiovascular Diseases	0.16	0.0100
20	Morning Person	0.14	0.0100
21	Eczema	0.12	0.0030

**Table S15: Model parameter values for LDpred applied to 21 UK Biobank traits.**  $h_g^2$  estimate (from BOLT-LMM v2.3),  $p$  is the fraction of non-zero effects in the prior, and LD-radius assumed was 2000 SNPs. The main analyses from LDpred exclude long-range LD regions reported in ref. 23, given that including these regions proved to be sub-optimal (see Table S14).

	Phenotype	$h_g^2$	P+T	P-values threshold for	
				P+T-funct-LASSO HP SNP set	P+T-funct-LASSO LP SNP set
1	Height	0.57	0.0100	0.30	0.10
2	Hair color	0.45	0.0010	0.10	0.01
3	Platelet count	0.40	0.0100	0.10	0.10
4	Bone mineral density	0.40	0.0010	0.10	0.10
5	Balding type I	0.32	0.0001	0.10	0.01
6	Red blood cell count	0.32	0.0010	0.10	0.10
7	Age at menarche	0.31	0.0100	0.10	0.10
8	FEV1 FVC ratio	0.31	0.0010	0.10	0.10
9	Body mass index	0.31	0.1000	0.30	0.10
10	RBC distribution width	0.29	0.0010	0.10	0.01
11	Forced vital capacity	0.27	0.0100	0.10	0.10
12	Eosinophil count	0.27	0.0010	0.10	0.10
13	White blood cell count	0.27	0.0100	0.10	0.10
14	Systolic Blood pressure	0.27	0.0100	0.10	0.10
15	Tanning ability	0.23	0.0010	0.10	0.01
16	Waist hip ratio	0.21	0.0100	0.10	0.10
17	College Education	0.20	1.0000	0.30	0.30
18	Hypertension	0.18	0.0100	0.10	0.10
19	Cardiovascular Diseases	0.16	0.1000	0.10	0.10
20	Morning Person	0.14	0.0100	0.10	0.10
21	Eczema	0.12	0.0100	0.10	0.01

**Table S16: Model parameter values for P+T and P+T-funct-LASSO in 21 UK Biobank traits.**

We report the optimal p-value threshold for Pruning + Thresholding (P+T), optimal p-value threshold for P+T-funct-LASSO high prior SNP (HP) set and optimal p-value threshold for P+T-funct-LASSO low prior SNP (LP) set. Optimal  $R_{LD}^2$  values was 0.1.

	Phenotype	$h_g^2$	P+T	P+T-funct-LASSO	LDpred	LDpred-funct-inf	LDpred-funct
1	Height	0.575	0.2228	0.3034	0.7595	0.7367	0.9938
2	Hair color	0.446	0.2505	0.3058	0.7254	0.7182	0.9920
3	Platelet count	0.401	0.2429	0.3423	0.8451	0.8115	0.9895
4	Bone mineral density	0.398	0.2871	0.3477	0.8192	0.8246	0.9865
5	Balding type I	0.323	0.3693	0.5050	0.8994	0.8781	0.9776
6	Red blood cell count	0.319	0.2898	0.3458	0.8583	0.8202	0.9822
7	Age at menarche	0.313	0.1990	0.3430	1.0227	0.8706	0.9782
8	FEV1 FVC ratio	0.309	0.3021	0.3593	0.8843	0.8527	0.9740
9	Body mass index	0.307	0.1687	0.3541	0.9138	0.8599	0.9813
10	RBC distribution width	0.286	0.2839	0.4189	0.8399	0.8123	0.9833
11	Forced vital capacity	0.274	0.2237	0.3783	0.9085	0.8665	0.9770
12	Eosinophil count	0.274	0.2781	0.3298	0.9082	0.8518	0.9830
13	White blood cell count	0.273	0.2352	0.3707	0.9033	0.8538	0.9793
14	Systolic Blood pressure	0.267	0.2200	0.3637	0.9050	0.8453	0.9808
15	Tanning ability	0.235	0.2437	0.2873	0.8312	0.8292	0.9905
16	Waist hip ratio	0.210	0.2057	0.3344	0.8453	0.8500	0.9758
17	College Education	0.198	0.1345	0.2610	1.0159	0.8520	0.9728
18	Hypertension	0.179	0.2140	0.3557	0.9817	0.8077	0.9710
19	Cardiovascular Diseases	0.160	0.1213	0.3296	0.9376	0.7953	0.9643
20	Morning Person	0.137	0.2158	0.3720	1.0803	0.8751	0.9651
21	Eczema	0.118	0.1752	0.4971	0.7496	0.7611	0.9634
22	Average across traits	0.286	0.2325	0.3574	0.8873	0.8273	0.9791

**Table S17: Calibration comparison for the 5 methods applied to 21 UK Biobank traits.** We report calibration slopes for each method, where a value close to 1 represents a well calibrated prediction.

Trait	LDpred-funct-inf	LDpred-funct-10	LDpred-funct-20	LDpred-funct-50	LDpred-funct-75	LDpred-funct-100
1 Height	0.4003	0.4113	0.4116	0.4126	0.4127	0.4128
2 Hair color	0.2624	0.2998	0.3059	0.3174	0.3199	0.3290
3 Platelet count	0.2315	0.2445	0.2453	0.2445	0.2446	0.2448
4 Bone mineral density	0.2137	0.2266	0.2266	0.2271	0.2265	0.2256
5 Balding type I	0.1075	0.1217	0.1235	0.1220	0.1198	0.1185
6 Red blood cell count	0.1571	0.1651	0.1655	0.1660	0.1660	0.1649
7 Age at menarche	0.1082	0.1118	0.1116	0.1122	0.1112	0.1070
8 FEV1 FVC ratio	0.1311	0.1353	0.1348	0.1343	0.1336	0.1315
9 Body mass index	0.1508	0.1501	0.1504	0.1494	0.1481	0.1473
10 RBC distribution width	0.1421	0.1527	0.1535	0.1535	0.1530	0.1517
11 Forced vital capacity	0.1145	0.1160	0.1155	0.1145	0.1128	0.1118
12 Eosinophil count	0.1335	0.1425	0.1422	0.1415	0.1406	0.1395
13 White blood cell count	0.1239	0.1278	0.1284	0.1276	0.1266	0.1261
14 Systolic Blood pressure	0.1114	0.1129	0.1119	0.1118	0.1108	0.1105
15 Tanning ability	0.1229	0.1716	0.1794	0.1818	0.1873	0.1892
16 Waist hip ratio	0.0793	0.0818	0.0810	0.0804	0.0798	0.0782
17 College Education	0.0716	0.0720	0.0731	0.0731	0.0748	0.0739
18 Hypertension	0.0523	0.0542	0.0541	0.0528	0.0521	0.0519
19 Cardiovascular Diseases	0.0423	0.0437	0.0433	0.0421	0.0410	0.0410
20 Morning Person	0.0372	0.0372	0.0366	0.0359	0.0349	0.0340
21 Eczema	0.0274	0.0278	0.0275	0.0271	0.0274	0.0258
22 Average across traits	0.1343	0.1432	0.1439	0.1442	0.1440	0.1436

**Table S18: Sensitivity of LDpred-funct results to number of bins used for regularization across 21 UK Biobank traits.** We report results with the number of posterior mean causal effect size bins used for regularization ( $K$ ) set to 10, 20, 50, 75 or 100. LDpred-funct- $K$  denotes each respective value of  $K$ . We also report results for LDpred-funct-inf, which is identical to LDpred-funct with  $K$  set to 1.

Trait	$h_g^2$	LDpred-funct-inf	Validation sample size				
			1000	2000	5000	10000	ALL
1 Height	0.57	0.4003	0.4105	0.4097	0.4100	0.4106	0.4128
2 Hair color	0.45	0.2624	0.2998	0.3005	0.3018	0.3076	0.3290
3 Platelet count	0.40	0.2315	0.2475	0.2443	0.2437	0.2435	0.2460
4 Bone mineral density	0.40	0.2137	0.2296	0.2260	0.2266	0.2256	0.2256
5 Balding type I	0.32	0.1075	0.1257	0.1227	0.1219	0.1237	0.1221
6 Red blood cell count	0.32	0.1571	0.1665	0.1657	0.1643	0.1638	0.1659
7 Age at menarche	0.31	0.1079	0.1161	0.1124	0.1115	0.1108	0.1122
8 FEV1 FVC ratio	0.31	0.1311	0.1408	0.1372	0.1345	0.1347	0.1330
9 Body mass index	0.31	0.1508	0.1549	0.1511	0.1512	0.1503	0.1499
10 RBC distribution width	0.29	0.1421	0.1550	0.1519	0.1515	0.1529	0.1533
11 Forced vital capacity	0.27	0.1145	0.1211	0.1172	0.1143	0.1136	0.1134
12 Eosinophil count	0.27	0.1335	0.1469	0.1434	0.1412	0.1414	0.1409
13 White blood cell count	0.27	0.1239	0.1315	0.1291	0.1278	0.1276	0.1270
14 Systolic Blood pressure	0.27	0.1114	0.1162	0.1138	0.1119	0.1107	0.1112
15 Tanning ability	0.23	0.1229	0.1524	0.1568	0.1769	0.1805	0.1842
16 Waist hip ratio	0.21	0.0793	0.0899	0.0841	0.0825	0.0812	0.0806
17 College Education	0.20	0.0716	0.0794	0.0751	0.0726	0.0727	0.0728
18 Hypertension	0.18	0.0523	0.0604	0.0557	0.0543	0.0537	0.0534
19 Cardiovascular Diseases	0.16	0.0423	0.0504	0.0459	0.0446	0.0435	0.0427
20 Morning Person	0.14	0.0372	0.0439	0.0407	0.0380	0.0369	0.0365
21 Eczema	0.12	0.0274	0.0354	0.0317	0.0288	0.0277	0.0272
Average across traits	0.29	0.1343	0.1464	0.1436	0.1433	0.1435	0.1447

**Table S19: Sensitivity of LDpred-funct results to number of validation samples across 21 UK Biobank traits.** We report results with the number of validation samples set to 1,000, 2,000, 5,000, 10,000 (the number of regularization bins is proportional to the number of validation samples; see Equation 6. Results are averaged across 100 random subsets of each size. ALL denotes results of LDpred-funct using the total number of validation samples (reported in Table S1). We also report results for LDpred-funct-inf, which is equivalent to LDpred-funct in the limit of a very small number of validation samples.



Data Set	Training $N$	P+T	LDpred	P+T-funct -LASSO	LDpred-funct-inf	LDpred-funct
UK Biobank in- terim release	113,660	0.2223	0.2276	0.2524	0.2777	0.2926
UK Biobank	408,092	0.3448	0.3860	0.3644	0.3995	0.4132
23andMe	698,430	0.2903	0.2919	0.2985	0.3148	0.3279
Meta-analysis of UK Biobank and 23andMe	1,107,430	0.3710	0.4004	0.3778	0.4193	0.4292
Fixed-effect meta-analysis	1,107,430	0.3687	0.3675	0.3663	0.3965	0.4051

**Table S20: Accuracy of 5 prediction methods in height meta-analysis of UK Biobank and 23andMe cohorts.** We report results for P+T, LDpred, P+T-funct-LASSO, LDpred-funct-inf and LDpred-funct, for each of 4 training data sets: UK Biobank interim release (113,660 training samples), UK Biobank (408,092 training samples), 23andMe (698,430 training samples) and meta-analysis of UK Biobank and 23andMe (1,107,430 training samples). We also report results for a fixed-effect meta-analysis of UK Biobank and 23andMe.

	Phenotype	$h_g^2$	LDpred-funct	AnnoPred	Difference
1	Height	0.57	0.4128 (0.0261)	0.4078 (0.0268)	-0.0046 (0.0186)
2	Hair color	0.45	0.3290 (0.1358)	0.2591 (0.1124)	-0.0683 (0.0284)
3	Platelet count	0.40	0.2460 (0.0269)	0.2351 (0.0221)	-0.0099 (0.0095)
4	Bone mineral density	0.40	0.2256 (0.025)	0.2316 (0.0211)	0.0062 (0.0042)
5	Balding type I	0.32	0.1221 (0.0235)	0.1452 (0.0207)	0.0230 (0.0131)
6	Red blood cell count	0.32	0.1659 (0.0202)	0.1680 (0.0155)	0.0018 (0.0034)
7	Age at menarche	0.31	0.1122 (0.0183)	0.1144 (0.0102)	0.003 (0.0028)
8	FEV1 FVC ratio	0.31	0.1330 (0.017)	0.1445 (0.0102)	0.0112 (0.0034)
9	Body mass index	0.31	0.1499 (0.0151)	0.1539 (0.0079)	0.0042 (0.0029)
10	RBC distribution width	0.29	0.1533 (0.0202)	0.1487 (0.0149)	-0.0046 (0.007)
11	Forced vital capacity	0.27	0.1134 (0.0148)	0.1190 (0.0071)	0.0056 (0.0021)
12	Eosinophil count	0.27	0.1409 (0.0191)	0.1386 (0.014)	-0.0025 (0.0108)
13	White blood cell count	0.27	0.1270 (0.0161)	0.1320 (0.0096)	0.0049 (0.0067)
14	Systolic Blood pressure	0.27	0.1112 (0.0133)	0.1173 (0.0069)	0.0067 (0.0019)
15	Tanning ability	0.23	0.1842 (0.0784)	0.1226 (0.0645)	-0.0616 (0.028)
16	Waist hip ratio	0.21	0.0806 (0.0116)	0.0853 (0.0071)	0.0047 (0.0039)
17	College Education	0.20	0.0728 (0.0109)	0.0707 (0.0066)	-0.0022 (0.0027)
18	Hypertension	0.18	0.0534 (0.0094)	0.0575 (0.0048)	0.0041 (0.0019)
19	Cardiovascular Diseases	0.16	0.0427 (0.0084)	0.0468 (0.004)	0.0040 (0.0012)
20	Morning Person	0.14	0.0365 (0.008)	0.0390 (0.0032)	0.0025 (0.0013)
21	Eczema	0.12	0.0272 (0.0064)	0.0306 (0.0034)	0.0044 (0.0014)
	Average across traits	0.29	0.1439 (0.0112)	0.1407 (0.0098)	-0.0032 (0.0034)

**Table S21: Accuracy of LDpred-funct and AnnoPred across 21 UK Biobank traits.** We report prediction  $R^2$  for LDpred-funct and AnnoPred, and difference in prediction  $R^2$  between AnnoPred and LDpred-funct. Block-jackknife standard errors are reported in parentheses. When running AnnoPred, we excluded SNPs from long-range LD regions (analogous to LDpred). We note that AnnoPred employs either (i) a prior in which the probability of being causal is the same for each SNP and the causal effect size variance varies across SNPs, or (ii) a prior in which the probability of being causal varies across SNPs and the causal effect size variance is the same for each SNPs. We considered only the first prior, as the second prior constructs categories of SNPs that share the same annotation values; in the case of continuous-valued annotations this would lead to an infinite number of categories.

	Trait	$h_g^2$	LDpred-funct-inf under different priors:		
			baselineLD (1000G)	baselineLD (UK10K)	baselineLD + LDAK (UK10K)
1	Eosinophil count	0.274	0.1335	0.1335	0.1342
2	Platelet count	0.401	0.2315	0.2327	0.2298
3	RBC distribution width	0.286	0.1421	0.1432	0.1451
4	Red blood cell count	0.319	0.1571	0.1566	0.1544
5	White blood cell count	0.273	0.1239	0.1246	0.1251
6	Bone mineral density	0.398	0.2137	0.2122	0.2117
7	Balding type I	0.323	0.1075	0.1040	0.1070
8	Body mass index	0.307	0.1508	0.1503	0.1502
9	Height	0.575	0.4003	0.4031	0.4033
10	Waist hip ratio	0.210	0.0793	0.0793	0.0785
11	Systolic Blood pressure	0.267	0.1114	0.1113	0.1136
12	College Education	0.198	0.0716	0.0788	0.0790
13	Eczema	0.118	0.0274	0.0283	0.0277
14	Cardiovascular Diseases	0.160	0.0423	0.0446	0.0449
15	Hypertension	0.179	0.0523	0.0548	0.0555
16	FEV1 FVC ratio	0.309	0.1311	0.1309	0.1323
17	Forced vital capacity	0.274	0.1145	0.1147	0.1140
18	Morning Person	0.137	0.0372	0.0404	0.0404
19	Hair color	0.446	0.2624	0.2749	0.2723
20	Tanning ability	0.235	0.1229	0.1254	0.1232
21	Age at menarche	0.313	0.1079	0.0995	0.0930

**Table S22: Accuracy of LDpred-funct-inf(1000G), LDpred-funct-inf(UK10K) and LDpred-funct-inf(UK10K, baseline-LD+LDAK) across 21 UK Biobank traits.** We report results for each trait. Results for Average across traits are reported in Table S14.