



31 Introduction

32 One of the most promising developments from the field of basic human genetics research has been  
33 polygenic scores (PGS)<sup>1</sup>. Large scale genome-wide association studies (GWAS) have enabled the  
34 prediction of effect sizes for genetic variants across many different phenotypes. These effect sizes  
35 can then be used to calculate an individual's genetic propensity for a given trait or disease within a  
36 population, which is referred to as their PGS<sup>2, 3</sup>. At a population level, the overall prediction of a trait  
37 can be estimated using regression models to calculate the total phenotypic variance of a trait that  
38 can explained by the PGS in that population<sup>4, 5</sup>.

39 The theoretical upper bound for the prediction of a trait is its heritability calculated from genomic  
40 data<sup>6</sup>. However, the variance explained by PGS are often somewhat short of that upper bound. There  
41 are a number of potential reasons for this shortfall, including non-additive genetic effects<sup>7</sup>, causal  
42 variation not well captured by the available variants, differences in measurement of complex  
43 phenotypes<sup>8</sup>, differing genetic architectures<sup>9</sup> and allele frequencies<sup>10</sup> between populations, and  
44 reduced accuracy of effect size prediction from underpowered studies<sup>11</sup>. The differing genetic  
45 architecture between ancestries was highlighted by Martin et al.<sup>12</sup> as a key reason for lower  
46 predictive performance of PGS in non-European ancestry individuals. This lower predictive  
47 performance has the potential to further exacerbating global health inequalities due to the current  
48 reliance on European samples for conducting GWAS.

49 One potential solution for improving prediction of PGS across ancestries is to examine the genetic  
50 concordance between the GWAS discovery and test populations being examined. If the loci of  
51 interest contain a similar genetic architecture due to linkage disequilibrium (LD), then there can be  
52 greater confidence in the estimated effect sizes being consistent across ancestries at that position.  
53 However, among individuals that have a notably different genetic architecture, the confidence in the  
54 consistency of effect size is reduced and could be down weighted to reflect that uncertainty. This

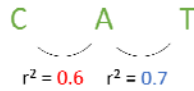
55 assumption led to the development of a software tool known as UP and Downstream Genetic  
56 scoring (UPDOG).

57 The predictive performance of UPDOG was assessed using the GWAS results of six phenotypes from  
58 primarily European ancestries and used to calculate and examine prediction in African, East Asian,  
59 and South American subsets of the UK Biobank. Genetic variant effect size estimates were obtained  
60 for each phenotype using five state-of-the-art PGS tools: Deterministic Bayesian Sparse Linear Mixed  
61 Model (DBSLMM)<sup>13</sup>, lassosum<sup>14</sup>, LDPred2<sup>15</sup>, MegaPRS<sup>16</sup> and PRS-CS<sup>17</sup>. Each tool therefore provided a  
62 benchmark for predictive performance that could then be used to examine whether UPDOG  
63 improved upon those benchmark predictions.

#### 64 Method

65 Three different models (see Figure 1) and a range of weightings were tested to examine how  
66 incorporating genetic data from both the summary statistics, a LD reference panel, and individuals in  
67 the test cohort could improve prediction accuracy of PGS. GWAS summary statistics and a LD  
68 reference panel were used to examine the effect sizes and LD structure around lead variants in those  
69 discovery cohorts. Lead variants are classified as those assigned an effect size by the state-of-the-art  
70 PGS tools. Where the LD structure is similar between the discovery cohort and an individual in the  
71 test cohort the expectation is that the lead variant's effect size will have greater accuracy compared  
72 to an individual in the test cohort where the LD structure is different. Model A applied adjustments  
73 to an individual's score based on what was carried at the up or downstream position regardless of  
74 what was carried at the lead position. Model B applied adjustments where the number of causal or  
75 protective alleles at the up or downstream position was different to what was carried at the lead  
76 position. Model C applied a haplotype-based approach where the minimum number of causal alleles  
77 across the downstream, lead, and upstream position was used. Model B with a scaling factor ( $\lambda$ ) of  
78 0.025 was identified as the optimum UK Biobank-tuned model (hereafter referred to as UPDOG) and  
79 is described below, with Model A and C covered in the supplementary material.

- Using the summary statistics, **A** was identified as the trait increasing allele at the lead position.



- A linkage disequilibrium reference panel identified the positions of the downstream and upstream variants, which for this example had an LD  $r^2$  of **0.6** and **0.7** with the lead position, respectively.

- Using the summary statistics, **C** and **T** were identified as the trait increasing alleles at the downstream and upstream positions, respectively.

Consider three individuals from the test cohort who carry the following number of **C**, **A**, and **T** alleles at the downstream, lead, and upstream positions, respectively:

Individual 1.	1	1	0
Individual 2.	1	1	2
Individual 3.	0	1	2

Assuming the **A** allele at the lead position has an effect size of **100**, under a typical risk score approach each individual would have a score of **100** as they each carry one copy of the trait increasing allele.

Three UPDOG models were assessed: **A**., **B**., and **C**.

**Model A.** Adjustment for trait increasing downstream and upstream alleles regardless of the lead allele

$$\text{score} + \text{weighting}((\text{alleles}_{\text{down}} * \text{lead effect size} * \text{LD}_{\text{down}} r^2) + (\text{alleles}_{\text{up}} * \text{lead effect size} * \text{LD}_{\text{up}} r^2))$$

Using a weighting of 0.025:

Individual 1.	$100 + 0.025((1 * 100 * 0.6) + (0 * 100 * 0.7)) = 100 + 0.025 (60 + 0)$	= 101.5
Individual 2.	$100 + 0.025((1 * 100 * 0.6) + (2 * 100 * 0.7)) = 100 + 0.025 (60 + 140)$	= 105
Individual 3.	$100 + 0.025((0 * 100 * 0.6) + (2 * 100 * 0.7)) = 100 + 0.025 (0 + 140)$	= 103.5

**Model B.** Adjustment for trait increasing downstream and upstream alleles where they differ from the lead allele

$$\text{score} + \text{weighting}(((\text{alleles}_{\text{down}} - \text{alleles}_{\text{lead}}) * \text{lead effect size} * \text{LD}_{\text{down}} r^2) + ((\text{alleles}_{\text{up}} - \text{alleles}_{\text{lead}}) * \text{lead effect size} * \text{LD}_{\text{up}} r^2))$$

Using a weighting of 0.025:

Individual 1.	$100 + 0.025(((1 - 1) * 100 * 0.6) + ((0 - 1) * 100 * 0.7)) = 100 + 0.025 (0 - 70)$	= 98.25
Individual 2.	$100 + 0.025(((1 - 1) * 100 * 0.6) + ((2 - 1) * 100 * 0.7)) = 100 + 0.025 (0 + 70)$	= 101.75
Individual 3.	$100 + 0.025(((0 - 1) * 100 * 0.6) + ((2 - 1) * 100 * 0.7)) = 100 + 0.025 (-60 + 70)$	= 100.25

**Model C.** Haplotype-based approach where the minimum number of trait increasing alleles was used

$$\min\{\text{alleles}_{\text{down}}, \text{alleles}_{\text{lead}}, \text{alleles}_{\text{up}}\} * \text{lead effect size}$$

Individual 1.	$\min\{1, 1, 0\} * 100 = 0$
Individual 2.	$\min\{1, 1, 2\} * 100 = 100$
Individual 3.	$\min\{0, 1, 2\} * 100 = 0$

80

81 Figure 1. Worked example of the three models that were examined to identify the optimum method  
 82 for obtaining the greatest number of improvements in trans-ancestral prediction

83

84 UPDOG

85 UPDOG requires four data inputs 1) the estimated effect sizes of the lead variants for a trait (these

86 can be obtained from  $P$ -value thresholding and clumping, although here the more advanced

87 methods that incorporate Bayesian or frequentist shrinkage methods are used); 2) the summary  
88 statistics providing the genome-wide results from the association analysis of that trait; 3) a LD  
89 reference panel matched to those summary statistics, for example, from the 1000 Genome Project<sup>18</sup>  
90 as used here; and 4) the test data for calculating the PGS for individuals in that dataset.

91 The first step undertaken by UPDOG is to split the genome up into chunks which are then  
92 parallelised to reduce the computational burden in terms of both memory and runtime. If estimated  
93 effect sizes for over 100,000 genome-wide lead variants are provided, then the genome is split into  
94 1,000 variant chunks, otherwise the genome is split into 30 Mb chunks. LD reference panel data and  
95 test data are then created for each chunk and extended by 250 Kb downstream and upstream.

96 Within each chunk, each lead variant is examined in turn and checked for a match to both the A1  
97 and A2 allele in the LD reference panel data and test data. A downstream variant is then identified  
98 using an iterative process moving one variant at a time away from the lead variant. The limits of the  
99 LD  $r^2$  of both the downstream and upstream variants with the lead variant was set between 0.5 and  
100 0.75 within a 250kb window. Therefore, the first variant identified with an LD  $r^2$  value  $> 0.5$  and  $<$   
101  $0.75$  and within 250kb of the lead variant is selected as the downstream variant. An upstream  
102 variant is identified in the same manner moving in the opposite direction away from the lead  
103 variant. An upstream variant is selected if it has an LD  $r^2$  value  $> 0.5$  and  $< 0.75$ , is within 250kb of the  
104 lead variant, and has an LD  $r^2$  value  $< 0.9$  with the selected downstream variant. If the  $r^2$  value is  $\geq$   
105  $0.9$  between the downstream and upstream variant, then the iterative process continues searching  
106 for an upstream variant that meets the criteria. The additional stipulation of an LD  $r^2$  value  $< 0.9$   
107 between the downstream and upstream variants ensures that they both positions provide additional  
108 information for the calculation of the PGS.

109 The calculation of a *score* for each lead genetic variant in the test data follows the standard  
110 approach, multiplying the *lead effect size* by the number of alleles carried by an individual at that  
111 position<sup>2</sup>. This *score* is then adjusted by examining the difference in the number of *alleles carried*

112 at the *down* and *up* stream positions compared to the number of *alleles carried* at the *lead*  
113 variant position with the same direction of effect as the lead variant to produce an UPDOG score  
114 such that:

$$UPDOG\ score = score + \lambda (downstream\ score + upstream\ score)$$

115 with

$$116 \quad downstream\ score = (alleles\ carried_{down} - alleles\ carried_{lead}) * lead\ effect\ size *$$

$$117 \quad LD_{Lead\_Down}$$

118 and

$$119 \quad upstream\ score = (alleles\ carried_{up} - alleles\ carried_{lead}) * lead\ effect\ size *$$

$$120 \quad LD_{Lead\_Up}$$

121 where  $LD_{Lead\_Down}$  is the LD  $r^2$  value between the lead variant and the downstream variant and  
122  $LD_{Lead\_Up}$  is the LD  $r^2$  value between the lead variant and the upstream variant. LD  $r^2$  value is  
123 calculated using the LD reference panel.  $\lambda$  is a scaling factor with values of 0.005, 0.0125, 0.025,  
124 0.05, 0.125, 0.25, and 0.5 examined. The *UPDOG scores* are then summed across all lead genetic  
125 variants to produce a PGS for each individual with the output from UPDOG being a vector of PGS for  
126 the individuals in the test data.

127 UPDOG is available for download from GitHub: <https://github.com/davemhoward/updog>

128 Evaluation of UPDOG

129 To evaluate the performance of UPDOG, predictive ability was compared against the benchmark  
130 predictions achieved from the five tools used to estimate variant effect sizes across six phenotypes  
131 and four different ancestries. The three models (A, B and C) and seven scaling factors ( $\lambda$ , as shown in  
132 the previous paragraph) were examined to identify the optimum UK Biobank-tuned model.

133 Phenotypes and variant effect size estimation

134 Large genome-wide association studies of six phenotypes were used to examine the predictive  
135 performance of UPDOG: body mass index<sup>19</sup>, coronary artery disease<sup>20</sup>, height<sup>21</sup>, major depression<sup>22</sup>,  
136 rheumatoid arthritis<sup>23</sup>, and type 2 diabetes<sup>24</sup>. The GWAS for the six phenotypes were selected due to  
137 UK Biobank not being included in each analysis (or a version excluding UK Biobank being available in  
138 the case of major depression), enabling UK Biobank to be used as the test cohort for PGS prediction.  
139 The estimated effect sizes of lead variants for each phenotype, the quality control applied, and the  
140 identification of the optimal shrinkage parameters to for obtain effect sizes were the same as those  
141 reported in Pain et al.<sup>25</sup>. Based on the results observed in that paper the estimated effects sizes  
142 from DBSLMM<sup>13</sup>, lassosum<sup>14</sup>, LDPred2<sup>15</sup>, MegaPRS<sup>16</sup>, and PRS-CS<sup>17</sup> were used.

#### 143 Testing of prediction in UK Biobank

144 The UK Biobank is a large health study of over a half a million individuals residing in the United  
145 Kingdom<sup>26</sup>. Quality control was applied to the entire UK Biobank study to exclude individuals that  
146 had a variant call rate <98%, where the reported sex mismatched the genotypic sex, or where there  
147 was relatedness up to the third degree based on a kinship coefficient >0.044 according to the KING  
148 toolset<sup>27</sup>. Genetic variants with a call rate <98%, a minor allele frequency <0.01, a deviation from  
149 Hardy–Weinberg equilibrium ( $P < 10^{-6}$ ), that were non-biallelic, or had an imputation accuracy  
150 (Info) score <0.7 were excluded. This left a total of 7,718,731 genetic variants for the PGS analysis.

151 The approach described by Privé<sup>28</sup> was used to identify 2,161 individuals of African ancestry, 441  
152 individuals of South American ancestry, 1,424 individuals of East Asian ancestry, and a European  
153 ancestry subset of 5,913 individuals. The European ancestry subset was restricted to a random  
154 sample of 2,000 individuals with ancestry from the United Kingdom and 2,000 individuals with Irish  
155 ancestry with the remaining 1,913 individuals with ancestry from Europe (South West), Europe  
156 (North East), Ashkenazi, and Finland.

157 Six phenotypes were assessed in the UK Biobank that were matched to the summary statistics  
158 previously described. Body mass index was obtained from Data-Field f.21001.0.0 with individuals

159 that were three standard deviations from the mean removed. Coronary artery disease was based on  
160 the definition used by Fürtjes et al.<sup>29</sup> incorporating both electronic health records and self-reporting  
161 of coronary artery disease to identify cases. Height was obtained from Data-Field f.50.0.0 with  
162 individuals that were three standard deviations from the mean removed. For depression, the broad  
163 depression phenotype based on help-seeking behaviour from Howard et al.<sup>30</sup> was used. Rheumatoid  
164 arthritis was based on the respective possible definition used by Glanville et al.<sup>31</sup> based on  
165 electronic health records and self-report, but without the assessment of medications. Finally, type 2  
166 diabetes was based on the definition used by Fürtjes et al.<sup>29</sup>, except that an exclusion criteria based  
167 on starting insulin within one year diagnosis was not applied (Data-Field f.2986.0.0).

#### 168 Assessment of UPDOG

169 The predictive performance of the respective PGS calculated by UPDOG for the two continuous  
170 phenotypes, body mass index and height, was assessed using Pearson's correlation coefficient. Sex  
171 and the first 16 genetic principal components were fitted as fixed effects and the PGS were  
172 standardised prior to the calculation of the correlation coefficient. A comparison of the correlation  
173 coefficients was made between the original PGS from each state-of-the-art tool and the PGS after  
174 applying UPDOG using the same effect sizes from each tool. Comparisons were made for each  
175 phenotype and within each ancestry group.

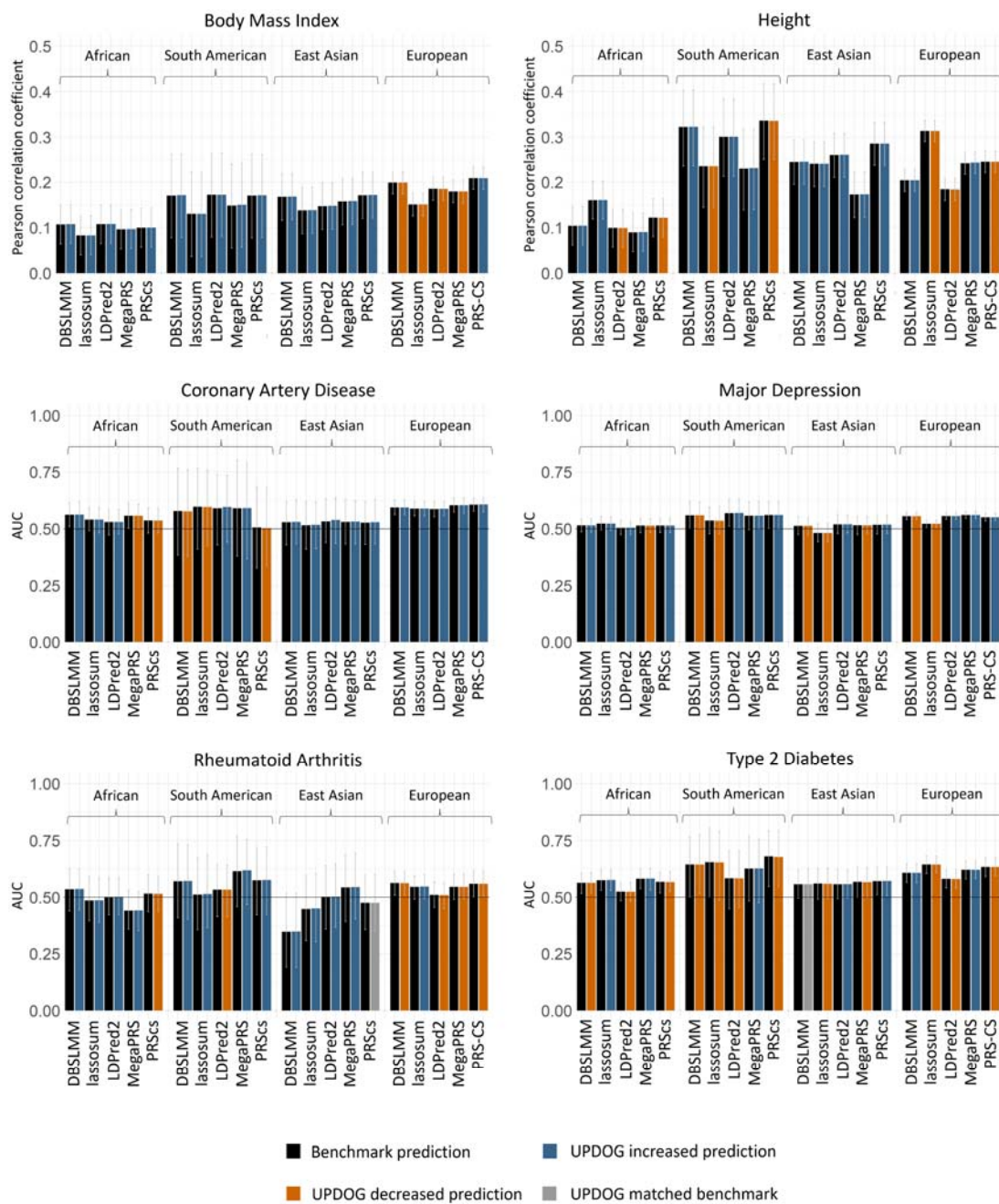
176 The predictive performance for the four binary phenotypes, coronary artery disease, major  
177 depression, rheumatoid arthritis, and type 2 diabetes, was assessed using a covariate-adjusted area  
178 under the receiver operating characteristic curve (AUC) and a semiparametric approach<sup>32</sup> using the  
179 AROC R-package<sup>33</sup>. The covariates adjusted were the same as for the continuous phenotypes: sex  
180 and the first 16 genetic principal components. A comparison of the AUC was made between the  
181 original PGS from each state-of-the-art tool and the PGS after from each state-of-the-art tool after  
182 applying UPDOG using the same effect sizes from each tool.

#### 183 Results



184 Summary statistics from six phenotypes (body mass index, coronary artery disease, height, major  
185 depression, rheumatoid arthritis, and type 2 diabetes) based on GWAS of European cohorts were  
186 analysed for their prediction into ancestral groups (African, South American, East Asian, and  
187 European) within the UK Biobank. Predictions were obtained using PGS derived from five state-of-  
188 the-art tools (DBSLMM, lassosum, LDPred2, MegaPRS, and PRS-CS) and three potential models (A, B,  
189 and C) compared. Model B (UPDOG) performed the best using  $\lambda = 0.025$  with the predictions for  
190 each phenotype, ancestral group, and tool compared to the benchmark prediction shown in Figure  
191 2.

192 In total, 62 out of 90 (68.9%) trans-ancestral predictions (based on 6 phenotypes x 5 state-of-the-art  
193 tools x 3 ancestries) were improved after applying UPDOG, two (2.2%) were unchanged, and 26  
194 (28.9%) were less predictive. Assuming a binomial distribution, the probability of observing at least  
195 62 improvements out of 90 predictions was  $2.2 \times 10^{-4}$ .



196

197 Figure 2. Predictive performance of UPDOG compared to benchmarks set by state-of-the-art tools.  
 198 Performance was judged across six phenotypes (body mass index, height, coronary artery disease, major  
 199 depression, rheumatoid arthritis, and type 2 diabetes) using the Pearson correlation coefficient for continuous  
 200 traits and the area under the receiver operator characteristic curve (AUC) for binary traits. Error bars indicate  
 201 the 95% confidence interval. Predictions were made in to four ancestries (African, South American, East Asian,  
 202 and European) using effect size estimates from five state of the art-the-art tools (DBSLMM, lassosum, LDPred2,  
 203 MegaPRS, and PRS-CS). The effect size estimates from the tools provide a benchmark (shown in black) with  
 204 which to assess the performance of UPDOG when using those same effect sizes. No change in prediction is  
 205 shown in grey, increased prediction in blue and decreased prediction in orange.

206

207 Trans-ancestral prediction within each phenotype

208 For body mass index, all 15 trans-ancestral predictions (3 ancestries and 5 tools) were improved  
209 after applying UPDOG using the effects size estimates from each state-of-the-art tool. Eleven out of  
210 15 trans-ancestral predictions for height were improved after applying UPDOG. The average change  
211 in the Pearson correlation coefficient for body mass index and height across the 15 predictions was  
212  $6.63 \times 10^{-4}$  and  $2.07 \times 10^{-4}$ , respectively. Coronary artery disease, major depression, rheumatoid  
213 arthritis, and type 2 diabetes had improvement in ten, nine, twelve, and five out of 15 trans-  
214 ancestral predictions after applying UPDOG, respectively. The average change in the AUC value for  
215 coronary artery disease, major depression, rheumatoid arthritis, and type 2 diabetes across the 15  
216 predictions was  $8.92 \times 10^{-4}$ ,  $1.35 \times 10^{-4}$ ,  $1.16 \times 10^{-3}$ , and  $-4.04 \times 10^{-4}$ , respectively.

217 Trans-ancestral prediction within each tool

218 After using UPDOG to generate polygenic scores from the calculated variant effect sizes from  
219 MegaPRS and LDpred2, 14 out of 18 (3 ancestries and 6 phenotypes) predictions improved based on  
220 an increased Pearson correlation coefficient or AUC value. The predictions from both DBSLMM and  
221 lassosum improved for twelve out of 18 predictions after using UPDOG. The predictions from PRS-CS  
222 improved for ten out of 18 predictions after using UPDOG.

223 Prediction within each ancestry

224 Prediction into an African subset from European-based summary statistics was improved for 21 out  
225 of 30 (6 phenotypes and 5 tools) predictions after using UPDOG. The prediction into a South  
226 American subset and an East Asian subset was improved in 18 and 23 cases out of 30 using UPDOG,  
227 respectively. UPDOG was developed to improve trans-ancestral prediction, the analysis within the  
228 same ancestry (i.e., prediction from European summary statistics into a European subset in UK  
229 Biobank) was also conducted and 14 out of 30 predictions improved after generating polygenic  
230 scores with UPDOG.

231 Discussion

232 Polygenic scores are one of the leading scientific advances from the genomic era as they translate  
233 genomic findings to individual-level measures of genetic loading<sup>34</sup>. Improving the accuracy and  
234 efficacy of PGS to enable their use within clinical settings to deliver personalised treatments is a vital  
235 area of research. Three key opportunities exist for improving predictions from PGS. Firstly, to  
236 improve the accuracy of the effect size estimates from genome-wide association studies. This can  
237 primarily be achieved through increased sample size or increased accuracy in phenotype  
238 ascertainment. Second, improving the process for incorporating the genome-wide effect sizes from  
239 multiple correlated genetic variants. Primarily this has been through the optimisation of shrinkage  
240 methods and lead to the development of state-of-the-art tools such as LDpred2<sup>15</sup>, MegaPRS<sup>16</sup>, and  
241 PRS-CS<sup>17</sup>. Finally, developing novel methods to use the calculated effect sizes to determine the PGS  
242 assigned to each individual in a test cohort. Where the genetic architecture is similar between the  
243 samples used in the genome-wide association study and the test cohort the standard approach of  
244 summing the risk alleles weighted by their assigned effect sizes<sup>2</sup> is likely to be most effective.  
245 However, where there are differing genetic architectures, potentially due to ancestry, between the  
246 samples used in the genome-wide association study and the test cohort then alternate approaches  
247 are required.

248 UPDOG represents a novel approach for calculating PGS when the effect size estimated for genetic  
249 variants are obtained using an ancestry that is different to the one in which the predictions are being  
250 made. By examining the genetic architecture surrounding the lead genetic variants based on the  
251 summary statistics and comparing those alleles with what each individual carries in the test cohort  
252 improvements in prediction should be possible. The performance of UPDOG was examined across  
253 four ancestries, six phenotypes, and benchmarked against five state-of-the-art tools for the  
254 calculation of variant effect sizes. The number of increases in prediction were above that expected

255 by random chance ( $P = 2.2 \times 10^{-4}$ ); however, the increases in either Pearson correlation coefficient or  
256 AUC value were modest.

257 The number of improved predictions from UPDOG was relatively consistent using the calculated  
258 effect sizes from the five state-of-the-art tools and across the three non-European ancestries.  
259 Among the six phenotypes tested, body mass index had the greatest number (15 out of 15) of  
260 improved trans-ancestral predictions when applying UPDOG. Rheumatoid arthritis, height, coronary  
261 artery disease, and major depression had between 9 and 12 increase trans-ancestral predictions.  
262 Type two diabetes was only improved for 5 out of 15 trans-ancestral predictions. It is unclear why  
263 there were differences between phenotypes for the number of improved predictions. The use of  
264 only UK Biobank participants should help reduce any bias in phenotypic ascertainment between the  
265 ancestries analysed compared to ancestries drawn from different cohorts. Power due to genome-  
266 wide association study sample size is unlikely to be the cause of these differences, with depression  
267 having the largest sample size and rheumatoid arthritis having the smallest. SNP-based heritability  
268 ( $h^2$ ) may have contributed to the performance of UPDOG with body mass index ( $h^2 = 0.49$ ) and  
269 height ( $h^2 = 0.25$ ) being more heritable than the other phenotypes ( $0.07 < h^2 < 0.16$ ) based on  
270 estimates from the UK Biobank SNP-Heritability Browser<sup>35</sup>. Body mass index and height were also  
271 continuous measures whereas the other phenotypes used binary measures of case control status.

272 An alternative tool for calculating PGS using data from more than one ancestry is PRS-CSx<sup>36</sup>. PRS-CSx  
273 enables the inclusion of results from multiple genome-wide association studies conducted using  
274 different ancestries and applies a shared continuous shrinkage prior and models population-specific  
275 allele frequencies and LD patterns to produce posterior variant effect size estimates. A multi-ethnic  
276 polygenic risk scores approach has also been suggested by Márquez-Luna et al.<sup>37</sup> and uses a sample  
277 size weighted average of estimated effect sizes to generate variant effect size estimates. Both these  
278 approaches seek to maximise the sample size and optimise the output of the calculated effect sizes.  
279 Currently UPDOG is the only tool that also considers the individuals within the test cohort. However,

280 UPDOG is not designed to consider either genome-wide association studies that have meta-analysed  
281 multiple ancestries or the inclusion of separate ancestry-specific association studies. The results  
282 presented here are based on prediction into individuals in the UK Biobank and may not be  
283 informative for individuals outside of this setting. The results presented here are potentially biased  
284 by the selection of the optimum model and weighting using the same sample for which the results  
285 are reported.

286 Strategies for increasing the predictive performance of PGS for disorders and diseases are crucial in  
287 ensuring that the right person gets the right support at the right time. There has very rightly been an  
288 extensive effort to widen genotyping to include more individuals from under-represented  
289 populations. Maximising the potential of this data is critical for overcoming global health  
290 inequalities. Therefore, UPDOG was developed to improve prediction of polygenic scores across  
291 different ancestries. UPDOG was able to provide an increase in trans-ancestral prediction in over  
292 two-thirds of instances examined using effect sizes estimated from state-of-the-art tools. The  
293 improvements in Pearson correlation coefficients or AUC values delivered by UPDOG were small.  
294 However, the development of tools that also consider the population being predicted into is an area  
295 of research that warrants further investigation.

296

#### 297 Acknowledgements

298 This research was conducted using the UK Biobank resource, application number 16577. We are  
299 grateful to the UK Biobank and all its voluntary participants. The UK Biobank study was conducted  
300 under generic approval from the NHS National Research Ethics Service (approval letter dated June  
301 17, 2011, Ref 11/NW/0382). All participants gave full informed written consent.

302 D.M.H is supported by a Sir Henry Wellcome Postdoctoral Fellowship (Reference 213674/Z/18/Z).

303 C.M.L acknowledges MRC grant MR/N015746/1. This investigation represents independent research

304 part-funded by the National Institute for Health Research (NIHR) Maudsley Biomedical Research  
305 Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views  
306 expressed are those of the authors and not necessarily those of the NHS, the NIHR or the  
307 Department of Health and Social Care.

308 This research was funded in whole, or in part, by the Wellcome Trust [Reference 213674/Z/18/Z].  
309 For the purpose of open access, the author has applied a CC BY public copyright licence to any  
310 Author Accepted Manuscript version arising from this submission.

#### 311 Authorship Contributions

312 D. M. H., O. P, A. C. G., E. V., and C. M. L. contributed to the concept and design of the work. D. M. H.  
313 developed the new software (UPDOG) used in the work and conducted the analysis. D. M. H., O. P,  
314 A. C. G., E. V., and C. M. L. contributed to the interpretation of the results. D. M. H. drafted the  
315 manuscript which was subsequently revised by O. P, A. C. G., E. V., and C. M. L.

#### 316 Declaration of competing financial interests

317 C.M.L is a member of the Myriad Neuroscience SAB, has received consultancy fees from UCB and  
318 speaker fees from SYNLAB.

#### 319 Data availability

320 According to Wellcome Trust's Policy on data, software and materials management and sharing, all  
321 data supporting this study will be openly available at <http://doi.org/doi:10.XXXXX/XXXXXXXX>

322

#### 323 References

- 324 1. Kullo IJ, Lewis CM, Inouye M, Martin AR, Ripatti S, Chatterjee N. Polygenic scores in  
325 biomedical research. *Nature Reviews Genetics* **23**, 524–532 (2022).
- 326 2. Choi SW, Mak TS-H, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses.  
327 *Nature Protocols* **15**, 2759-2772 (2020).
- 328

- 329  
330 3. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments.  
331 *Genome Medicine* **12**, 44 (2020).
- 332  
333 4. Daetwyler HD, Villanueva B, Woolliams JA. Accuracy of Predicting the Genetic Risk of Disease  
334 Using a Genome-Wide Approach. *PLoS One* **3**, e3395 (2008).
- 335  
336 5. Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLOS Genetics* **9**,  
337 e1003348 (2013).
- 338  
339 6. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from  
340 genome-wide association studies. *Genome Research* **17**, 1520-1528 (2007).
- 341  
342 7. Guindo-Martínez M, *et al.* The impact of non-additive genetic associations on age-related  
343 complex diseases. *Nature Communications* **12**, 2436 (2021).
- 344  
345 8. Duncan L, *et al.* Analysis of polygenic risk score usage and performance in diverse human  
346 populations. *Nature Communications* **10**, 3328 (2019).
- 347  
348 9. Rosenberg NA, Edge MD, Pritchard JK, Feldman MW. Interpreting polygenic scores,  
349 polygenic adaptation, and human phenotypic differences. *Evolution, Medicine, and Public*  
350 *Health* **2019**, 26-34 (2018).
- 351  
352 10. Kurniansyah N, *et al.* A multi-ethnic polygenic risk score is associated with hypertension  
353 prevalence and progression throughout adulthood. *Nature Communications* **13**, 3549 (2022).
- 354  
355 11. Lello L, Raben TG, Yong SY, Tellier LCAM, Hsu SDH. Genomic Prediction of 16 Complex  
356 Disease Risks Including Heart Attack, Diabetes, Breast and Prostate Cancer. *Scientific Reports*  
357 **9**, 15286 (2019).
- 358  
359 12. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current  
360 polygenic risk scores may exacerbate health disparities. *Nature Genetics* **51**, 584-591 (2019).
- 361  
362 13. Yang S, Zhou X. Accurate and scalable construction of polygenic scores in large biobank data  
363 sets. *The American Journal of Human Genetics* **106**, 679-693 (2020).
- 364  
365 14. Mak TSH, Porsch RM, Choi SW, Zhou X, Sham PC. Polygenic scores via penalized regression  
366 on summary statistics. *Genetic Epidemiology* **41**, 469-480 (2017).
- 367  
368 15. Privé F, Arbel J, Vilhjálmsson BJ. LDpred2: better, faster, stronger. *Bioinformatics* **36**, 5424-  
369 5431 (2020).
- 370  
371 16. Zhang Q, Privé F, Vilhjálmsson B, Speed D. Improved genetic prediction of complex traits  
372 from individual-level data or summary statistics. *Nature Communications* **12**, 4192 (2021).



- 373  
374 17. Ge T, Chen C-Y, Ni Y, Feng Y-CA, Smoller JW. Polygenic prediction via Bayesian regression and  
375 continuous shrinkage priors. *Nature Communications* **10**, 1776 (2019).
- 376  
377 18. The 1000 Genomes Project Consortium. A global reference for human genetic variation.  
378 *Nature* **526**, 68-74 (2015).
- 379  
380 19. Locke AE, *et al.* Genetic studies of body mass index yield new insights for obesity biology.  
381 *Nature* **518**, 197-206 (2015).
- 382  
383 20. Nikpay M, *et al.* A comprehensive 1000 Genomes-based genome-wide association meta-  
384 analysis of coronary artery disease. *Nature Genetics* **47**, 1121-1130 (2015).
- 385  
386 21. Wood AR, *et al.* Defining the role of common variation in the genomic and biological  
387 architecture of adult human height. *Nature Genetics* **46**, 1173-1186 (2014).
- 388  
389 22. Wray NR, *et al.* Genome-wide association analyses identify 44 risk variants and refine the  
390 genetic architecture of major depression. *Nature Genetics* **50**, 668-681 (2018).
- 391  
392 23. Okada Y, *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery.  
393 *Nature* **506**, 376-381 (2014).
- 394  
395 24. Scott RA, *et al.* An expanded genome-wide association study of type 2 diabetes in  
396 Europeans. *Diabetes* **66**, 2888-2902 (2017).
- 397  
398 25. Pain O, *et al.* Evaluation of polygenic prediction methodology within a reference-  
399 standardized framework. *PLOS Genetics* **17**, e1009021 (2021).
- 400  
401 26. Bycroft C, *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature*  
402 **562**, 203-209 (2018).
- 403  
404 27. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M. Robust relationship  
405 inference in genome-wide association studies. *Bioinformatics* **26**, 2867-2873 (2010).
- 406  
407 28. Privé F. Using the UK Biobank as a global reference of worldwide populations: application to  
408 measuring ancestry diversity from GWAS summary statistics. *Bioinformatics* **38**, 3477-3480  
409 (2022).
- 410  
411 29. Fürtjes AE, Coleman JRI, Tyrrell J, Lewis CM, Hagenaars SP. Associations and limited shared  
412 genetic aetiology between bipolar disorder and cardiometabolic traits in the UK Biobank.  
413 *Psychological Medicine*, 1-10 (2021).
- 414

- 415 30. Howard DM, *et al.* Genome-wide meta-analysis of depression identifies 102 independent  
416 variants and highlights the importance of the prefrontal brain regions. *Nature Neuroscience*  
417 **22**, 343-352 (2019).
- 418  
419 31. Glanville KP, Coleman JRI, O'Reilly PF, Galloway J, Lewis CM. Investigating pleiotropy  
420 between depression and autoimmune diseases using the UK Biobank. *Biological Psychiatry:*  
421 *Global Open Science* **1**, 48-58 (2021).
- 422  
423 32. Janes H, Pepe MS. Adjusting for covariate effects on classification accuracy using the  
424 covariate-adjusted receiver operating characteristic curve. *Biometrika* **96**, 371-382 (2009).
- 425  
426 33. Inacio de Carvalho V, Xose Rodriguez-Alvarez M. Bayesian nonparametric inference for the  
427 covariate-adjusted ROC curve. Preprint at  
428 <https://ui.adsabs.harvard.edu/abs/2018arXiv180600473> (2018).
- 429  
430 34. Lewis CM, Vassos E. Polygenic Scores in Psychiatry: On the Road From Discovery to  
431 Implementation. *American Journal of Psychiatry* **179**, 800-806 (2022).
- 432  
433 35. Walters R, Palmer D. UKB SNP-Heritability Browser.) (2022).
- 434  
435 36. Ruan Y, *et al.* Improving polygenic prediction in ancestrally diverse populations. *Nature*  
436 *Genetics* **54**, 573-580 (2022).
- 437  
438 37. Márquez-Luna C, Loh P-R, Consortium SATD, Consortium TSTD, Price AL. Multiethnic  
439 polygenic risk scores improve risk prediction in diverse populations. *Genetic Epidemiology*  
440 **41**, 811-823 (2017).
- 441  
442