

1 **Metabolic Polygenic Risk Scores for Prediction of Obesity, Type 2 Diabetes, and Related**
2 **Morbidities**

3
4 Min Seo Kim^{1,2,*}, Qiuli Chen^{3,4,5,*}, Yang Sui^{1,2,11,12*}, Xiong Yang^{3,4}, Shaoqi Wang^{3,4,5}, Lu-Chen
5 Weng^{1,2}, So Mi Jemma Cho^{1,2}, Satoshi Koyama^{1,2}, Xinyu Zhu^{3,4,6}, Kang Yu^{3,4,5}, Xingyu Chen^{3,4,5},
6 Rufan Zhang^{3,4}, Wanqing Yin^{3,4,5}, Shuangqiao Liao^{3,4,5}, Zhaoqi Liu^{3,4,5,6}, Fowzan S Alkuraya^{7,8,9},
7 Pradeep Natarajan^{1,2,10,11}, Patrick T. Ellinor^{1,2,11,12**}, Akl C. Fahed^{1,2,11,12**}, Minxian
8 Wang^{3,4,5,6,13,14,15**}

9
10 * co-first authors

11 ** co-senior authors

12
13 1. Medical and Population Genetics and Cardiovascular Disease Initiative, Broad Institute of MIT
14 and Harvard, Cambridge, MA, USA

15 2. Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA

16 3. National Genomics Data Center, China National Center for Bioinformation, Beijing, China

17 4. Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China

18 5. University of Chinese Academy of Sciences, Beijing, China

19 6. College of Future Technology, Sino-Danish College, University of Chinese Academy of
20 Sciences, Beijing, China

21 7. Department of Translational Genomics, Center for Genomic Medicine, King Faisal Specialist
22 Hospital and Research Center, Riyadh, Saudi Arabia

23 8. Lifera Omics, Riyadh, Saudi Arabia

24 9. College of Medicine, Alfaisal University, Riyadh, Saudi Arabia

25 10. Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA

26 11. Harvard Medical School, Boston, MA, USA

27 12. Cardiology Division, Massachusetts General Hospital, Boston, MA, USA

28 13. Department of Cardiovascular Surgery, Zhongnan Hospital of Wuhan University, Wuhan,
29 China

30 14. Hubei Provincial Engineering Research Center of Minimally Invasive Cardiovascular Surgery,
31 Wuhan, China

32 15. Wuhan Clinical Research Center for Minimally Invasive Treatment of Structural Heart Disease,
33 Wuhan, China.

34

35

36 **Please address correspondence to:**

37 Akl C. Fahed, MD, MPH

38 Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA

39 E-mail: afahed@mgh.harvard.edu

40

41

42 Minxian Wang, PhD

43 National Genomics Data Center, China National Center for Bioinformation, Beijing, China

44 E-mail: wangmx@big.ac.cn

45

46

47

48 Word count: 6414

49

50

51

52

53

54

55

56

57

58 **Abstract**

59 Obesity and type 2 diabetes (T2D) are metabolic diseases with shared pathophysiology.
60 Traditional polygenic risk scores (PRS) have focused on these conditions individually, yet the
61 single disease approach falls short in capturing the full dimension of metabolic dysfunction. We
62 derived biologically enriched metabolic PRS (MetPRS), a composite score that uses multi-
63 ancestry genome-wide association studies of 22 metabolic traits from over 10 million people.
64 MetPRS, optimized to predict obesity (O-MetPRS) and T2D (D-MetPRS), was validated in the UK
65 Biobank (UKB, n=15,000), and tested in UKB hold-out set (n=49,377), then externally tested in 3
66 cohorts – All of Us (n=245,394), Mass General Brigham (MGB) Biobank (n=53,306), and a King
67 Faisal Specialist Hospital and Research Center cohort (n=6,416). O-MetPRS and D-MetPRS
68 outperformed existing PRSs in predicting obesity and T2D across 6 ancestries (European, African,
69 East Asian, South Asian, Latino/admixed American, and Middle Eastern). O-MetPRS and D-
70 MetPRS also predicted morbidities and downstream complications of obesity and T2D, as well as
71 the use of GLP-1 receptor agonists in contemporary practice. Among 37,329 MGB participants
72 free of T2D and obesity at baseline, those in the top decile of O-MetPRS had a 103% relatively
73 higher chance, and those in the top decile of D-MetPRS had an 80% relatively higher chance of
74 receiving a GLP-1 receptor agonist prescription compared to individuals at the population median
75 of MetPRS. The biologically enriched MetPRS is poised to have an impact across all layers of
76 clinical utility, from predicting morbidities to informing management decisions.

77

78

79

80

81

82

83

84

85

86

87

88

89 Introduction

90 Obesity and type 2 diabetes (T2D) are closely interconnected through a shared
91 pathophysiology, often characterized by insulin resistance, and together constitute a major portion
92 of metabolic diseases. This relationship, frequently referred to as "diabesity",¹ reflects the
93 overlapping genetic architecture of both conditions, emphasizing the need to study their combined
94 effects on metabolic health and related morbidities.^{2,3} While research into the shared and distinct
95 genetic bases of metabolic traits has provided valuable biological insights,⁴ this multivariable
96 knowledge has not been fully translated into clinical utility via genetic risk assessment. The
97 polygenic risk score (PRS) has emerged as a powerful tool for assessing composite genetic risks,
98 gaining traction for its potential clinical applications.^{5,6}

99 The PRSs for obesity and T2D have been developed individually based on a single
100 disease approach thus far,⁷ without accounting for the multivariable nature of these conditions.
101 We and others previously demonstrated that incorporating genetic information for relevant
102 diseases enabled a significant boost in disease risk predictive performance.⁸⁻¹⁰ Building on this
103 multi-disease approach, we have now further incorporated a broader spectrum of measures for
104 diabesity, acknowledging that different measures of obesity (e.g., body mass index[BMI], waist
105 circumference[WC], waist-to-hip ratio[WHR]) and diabetes (e.g., HbA1c, fasting glucose[FG],
106 insulin sensitivity) capture intersecting but also complementary biology of these conditions. For
107 instance, BMI typically measures overall body mass, whereas WC provides more specific
108 information about abdominal or central adiposity.¹¹ Leveraging multiple indices that measure
109 various dimensions of metabolic dysfunction may effectively capture the underlying metabolic
110 complexity and potentially translate it into a biologically enriched and clinically meaningful PRS.

111 By integrating genetic information for related diseases and measures, we developed a
112 metabolic PRS (MetPRS) optimized for obesity (O-MetPRS) and T2D (D-MetPRS). We
113 incorporated genetic data from 22 metabolic traits, utilizing multi-ancestry genome-wide
114 association study (GWAS) data from over 10 million participants. The model was validated and
115 tested in the UK Biobank (UKB), and externally tested in three multiethnic cohorts comprising up
116 to 300,000 participants. This multi-disease and multi-measure approach has resulted in a
117 biologically enriched and cross-ancestry transferable PRS capable of predicting a broad spectrum
118 of metabolic health outcomes with high accuracy and relevance to clinical care.

119 Results

120 We integrated GWAS summary statistics from 22 obesity and diabetes related traits
121 including, but not limited to, BMI, body weight, body fat percentage, WC, WHR, hip circumference,
122 abdominal subcutaneous adipose tissue (ASAT), visceral adipose tissue (VAT), gluteofemoral
123 adipose tissue (GFAT), HbA1c, fasting glucose, fasting insulin, 2-hr glucose, and insulin
124 sensitivity index (**Supplementary Table 1**). The GWAS summary statistics were derived from the
125 largest contemporary data consortia, including Million Veteran Program (MVP), FinnGen, Biobank
126 Japan (BBJ), UK Biobank (UKB), and Taiwan Biobank, encompassing over 10 million participants.
127 Participants in the UK Biobank were divided into three distinct groups: the dataset for GWAS, the

128 training set, and the hold-out test set, with no overlap between samples across these groups
129 (**Extended Data Fig.1**). The obesity and T2D models were validated on 15,000 individuals of
130 European ancestry, including 3,814 individuals with obesity (BMI ≥ 30 Kg/m² for European) and
131 1,196 with T2D. UKB hold-out test set consisted of 49,377 individuals, with 70.1% of European
132 ancestry, 13.9% of African ancestry, 2.6% of East Asian ancestry, and 13.4% of South Asian
133 ancestry (**Extended Data Fig.1 and Supplementary Table 2**). The remaining European-ancestry
134 samples (N=378,921) were used for *de novo* GWAS analysis, and incorporated in the PRS score
135 development (**Supplementary Table 1**). The model development process is illustrated in **Fig. 1**.

136 Genetic Correlations Among Metabolic Traits

137 We selected 22 traits associated with diabetes and insulin resistance for which GWAS
138 summary statistics are currently available and have no sample overlap with UK biobank participants
139 (**Supplementary Table 1**). We assessed the genetic correlations among 22 metabolic traits using
140 cross-trait LD-score regression (**Fig. 2**). Significant correlations were observed between traits
141 associated with obesity, such as BMI, WC, body weight, VAT, reflecting their shared genetic
142 underpinnings. For example, the genetic correlation between BMI and WC was 0.90, while VAT
143 showed strong correlations with both body fat (0.71) and WC (0.79), which remained significant
144 after multiple testing corrections (**Fig. 2 and Supplementary Table 3**). In contrast, the Modified
145 Stumvoll insulin sensitivity index (ISI)¹² displayed negative genetic correlations with most traits,
146 as the index is inversely related to insulin resistance.

147

148 Model Development and Internal Validation of MetPRS in the UK Biobank

149 We used a previously described multi-ancestry and multi-trait GWAS integration
150 framework⁸ for developing high-performance and cross-ancestry transferable metabolic PRS
151 (MetPRS), with the purpose specifically optimized to predict obesity (O-MetPRS) and T2D (D-
152 MetPRS), **Fig 1**. The predictive performance of O-MetPRS and D-MetPRS was first evaluated in
153 the hold-out test set from the UK Biobank (n = 49,377) (**Fig. 3 and Extended Data Fig.1**). For
154 obesity among European participants, O-MetPRS achieved an odds ratio per standard deviation
155 (OR/SD) of 2.22 (95% CI: 2.16–2.28), outperforming all previously published PRSs for obesity –
156 the second best performing PRS for obesity (PGS catalog: PGS000027) had an OR/SD of 1.79
157 (95% CI: 1.75–1.84) (**Fig. 3A and Supplementary Table 4**). Moreover, the range of obesity
158 prevalence stratified by O-MetPRS was significantly broader, spanning from 2.9% to 70.3%
159 across percentiles, compared to 6.3% to 58.2% for the second best performing PRS, PGS000027
160 (**Fig. 3B**). For T2D, D-MetPRS showed an OR/SD of 2.08 (95% CI: 2.00–2.18), also exceeding
161 the performance of all previously published PRSs (**Fig. 3C and Supplementary Table 5**).
162 Similarly, D-MetPRS stratified a notably wider range of T2D prevalence, from 0.9% to 28.8%
163 across percentiles, compared to 2.6% to 23.9% for the second best PGS001357 (**Fig. 3D**).
164 Likewise, O-MetPRS and D-MetPRS consistently ranked as the top-performing PRSs for obesity
165 and T2D among African and East Asian participants (**Fig. 3A and Fig. 3C**). These findings
166 demonstrate that MetPRS provides superior risk stratification for both obesity and T2D in diverse
167 ethnic groups compared to existing PRSs.

168 We further tested the associations of O-MetPRS and D-MetPRS with various metabolic
169 traits (**Extended Data Fig.2 and Supplementary Table 6**). A consistent significance across all
170 traits ($P < 0.0001$) suggested that both O-MetPRS and D-MetPRS effectively capture the genetic
171 predispositions related to obesity and T2D, respectively. O-MetPRS demonstrated large
172 association coefficients observed for weight ($\beta = 4.75$ Kg per SD, $SE = 0.07$, $P < 0.0001$), hip
173 circumference ($\beta = 2.95$ cm per SD, $SE = 0.047$, $P < 0.0001$), and BMI ($\beta = 1.86$ Kg/m² per SD,
174 $SE = 0.024$, $P < 0.0001$). D-MetPRS showed strong associations with HbA1c ($\beta = 1.434$ mmol/mol
175 per SD, $SE = 0.034$, $P < 0.0001$), triglycerides (TG) ($\beta = 6.53$ mg/dl per SD, $SE = 0.003$, $P <$
176 0.0001), and glucose ($\beta = 3.098$ mg/dl per SD, $SE = 0.007$, $P < 0.0001$). Notably, O-MetPRS has
177 a larger effect size on obesity-related traits and D-MetPRS has larger effect size on diabetes-
178 related traits, suggesting that each score captured its relevant biological insights.

179 **External validation of MetPRS in multiple cohorts**

180 We validated the MetPRS in three external multiethnic cohorts: All of Us ($n = 245,394$),
181 Mass General Brigham (MGB) Biobank ($n = 53,306$), and the King Faisal Specialist Hospital and
182 Research Center [KFSH&RC] cohort ($n = 6,416$) (**Fig. 4; Supplementary Tables 7-12**). MetPRS
183 was compared against the top 5 obesity and T2D PRSs from the PGS Catalog.¹³ We selected the
184 top 5 PRSs that consistently ranked highly across all external datasets as comparators, excluding
185 any scores that incorporated each test cohort (i.e., All of Us) in their score development. The
186 OR/SD with 95% CI for prevalent obesity and T2D was assessed for O-MetPRS and D-MetPRS,
187 respectively, in a logistic regression model adjusted for age, sex, genotyping array, and the first
188 10 principal components of ancestry. MetPRS was the best-performing score across diverse
189 ancestries, except for O-MetPRS in the East/South Asian population of MGB Biobank, D-MetPRS
190 in Middle Eastern population of All of Us, and D-MetPRS in East Asian population of MGB Biobank
191 (**Fig.4**). In addition to predicting binary obesity, we also evaluated O-MetPRS for its ability to
192 predict continuous BMI (**Extended Data Fig.3**). O-MetPRS outperformed the top five comparator
193 scores across diverse ancestries in predicting BMI, with the exception of East/South Asian and
194 African populations in the MGB Biobank (**Extended Data Fig.3 and Supplementary Table 13**).
195 The consistently high performance across multi-ancestry external cohorts underscored the
196 robustness and generalizability of MetPRS.

197 Although MetPRS consistently performed better than pre-existing scores across
198 ancestries in relative space, its performance was suboptimal in individuals of African ancestry
199 (**Fig. 4**). To address this gap, we developed an African-optimized MetPRS by training the model
200 in UK Biobank African participants, which led to a further performance improvement tested in the
201 All of Us and MGB Biobank (**Extended Data Fig.4**).

202 **Prediction of Morbidities Related to Obesity and Type 2 Diabetes**

203 Obesity and T2D are not only major health challenges on their own but also serve as
204 strong risk factors and precursors for a wide array of downstream diseases and complications.
205 We assessed the ability of MetPRS to predict morbidities related to obesity and T2D using a Cox
206 proportional hazards model for individuals in the MetPRS top decile compared to those of the rest

207 cohort; the risk of all tested morbidities was significantly higher in top decile group for O-MetPRS
208 and D-MetPRS (**Extended Data fig.5; Supplementary Table 14**). When compared to the top 5
209 comparator scores, O-MetPRS had the best prediction of obesity-related morbidities, ranking first
210 for predicting 12 out of 14 obesity-related morbidities such as hypertension (hazard ratio[HR] 1.39;
211 95% CI: 1.30-1.49), venous thromboembolism (HR1.47; 95% CI: 1.21-1.80), and Nonalcoholic
212 fatty liver disease (HR 1.83; 95% CI: 1.41-2.37) (**Fig. 5A and 5B; Extended Data Fig. 6**). Similarly,
213 D-MetPRS most accurately predicted diabetes-related morbidities, ranking first for predicting 11
214 out of 13 diabetes-related morbidities such as neuropathy (HR 1.87; 95% CI: 1.35-2.59),
215 nephropathy (HR 1.89; 95% CI: 1.65-2.16), and retinopathy (HR 1.87; 95% CI: 1.35-2.59) (**Fig.**
216 **5C and 5D; Extended Data Fig. 6**).

217 **MetPRS and Risk of GLP-1 Receptor Agonist Prescription and Metabolic-bariatric Surgery**

218 The MGB Biobank consisted of 51,779 eligible participants, of whom 4,490 (8.7%) were
219 prescribed GLP-1 receptor agonists during a mean follow-up of 5.53 (SD 2.96) years, and 1,487
220 (2.9%) underwent metabolic-bariatric surgery during a mean follow-up of 5.53 (SD 3.32) years.
221 Among 51,779 participants, 37,329 were free of obesity and T2D at baseline. To evaluate the
222 relationship between MetPRS and clinical interventions in contemporary practice, we employed
223 Cox proportional hazards regression models, adjusting for age, sex, and top 10 ancestry principal
224 components. Participants were categorized into three risk groups based on their MetPRS – the
225 top 10%, middle quintile (40–60%), and bottom 10%.

226 Among 37,329 MGB individuals free of obesity and T2D, those in the top decile compared
227 to the population median quintile of O-MetPRS had a markedly increased risk of receiving a GLP-
228 1 agonist prescription (HR 2.03; 95% CI: 1.65-2.50) and metabolic-bariatric surgery (HR 3.61; 95%
229 CI: 2.06-6.33). The cumulative incidence of GLP-1 receptor agonist prescriptions and metabolic-
230 bariatric surgery showed significant variation across different O-MetPRS risk groups at 10-year
231 (**Fig. 6A and 6B; Extended Data Fig. 7A and 7B**). Among the group free of obesity and T2D at
232 baseline, participants in the top 10% risk group had a cumulative incidence of 12.6% for GLP-1
233 receptor agonist prescriptions and 0.4% for metabolic-bariatric surgery. For the middle quintile
234 group, the cumulative incidences were 6.5% and 0.1%, while the bottom 10% group had
235 incidences of 3.4% and 0.0%, respectively (**Fig.6A and 6B**). These trends were consistent in the
236 overall population, which included some participants with obesity and T2D at baseline who had
237 not been treated with these interventions. The cumulative incidence of GLP-1 receptor agonist
238 prescriptions and metabolic-bariatric surgery was generally higher in this population (**Extended**
239 **Data Fig. 7A and 7B**).

240 Similarly, those in the top decile compared to the population median quintile of D-MetPRS
241 had a substantially increased risk of receiving a GLP-1 agonist prescription (HR 1.80; 95% CI:
242 1.47-2.21) and metabolic-bariatric surgery (HR 2.29; 95% CI: 1.29-4.06). The cumulative
243 incidence of GLP-1 receptor agonist prescriptions and metabolic-bariatric surgery also varied
244 notably across the D-MetPRS risk groups at the 10-year mark (**Fig. 6C and 6D; Extended Data**
245 **Fig. 7**). In the population, those in the top 10% risk category had cumulative incidences of 12.1%
246 for GLP-1 receptor agonist prescriptions and 0.4% for metabolic-bariatric surgery. In contrast, the
247 middle quintile group had cumulative incidences of 7.0% and 0.2%, and the bottom 10% group

248 had 3.9% and 0.0%, respectively (**Fig.6C and 6D**). These patterns persisted in the overall
249 population with higher incidence rates (**Extended Data Fig. 7C and 7D**).
250

251 Discussion

252 This study introduces MetPRS for diabetes based on a novel approach that captures the
253 genetic architecture of metabolic dysfunction by synthesizing genetic association data from
254 related measures and indices. This method leveraged biological insights to optimize genetic risk
255 prediction for obesity, T2D, their related morbidities, and even clinical interventions in
256 contemporary practice. We demonstrate that MetPRS outperformed all previous PRSs in
257 predicting obesity and T2D across multiple ancestries, including European, African, East Asian,
258 South Asian, Latino/admixed American, and Arab populations. Moreover, MetPRS precisely
259 predicted downstream morbidities and clinical interventions associated with diabetes, such as
260 GLP-1 receptor agonist prescriptions and metabolic-bariatric surgery, underscoring its potential
261 clinical utility. The biologically enriched MetPRS is well-positioned to penetrate various aspects
262 of clinical practice, from predicting morbidities to guiding management decisions.

263 Metabolic disorders are inherently multifactorial and involve complex biology. As such,
264 integrating GWAS data for various diabetes-related measures resulted in a biologically-enriched
265 PRS. No single metric can fully capture the etiology of metabolic disorders, necessitating the
266 complementary use of multiple measures in clinical practice. For instance, HbA1c and fasting
267 glucose levels offer complementary insights – HbA1c reflects long-term glycemic control, while
268 fasting glucose levels indicate immediate glucose control.¹⁴ Likewise, multiple obesity measures,
269 such as BMI, WC, WHR, VAT, ASAT, and GFAT provide both converging and diverging insights
270 into different obesity subtypes and risk trajectories for obesity related morbidities.¹⁵ This growing
271 body of evidence supported our strategy to develop a biologically-enriched PRS by leveraging
272 genetic information across multiple diabetes-related measures. Emdin et al. previously
273 demonstrated that a PRS for WHR adjusted for BMI (WHRadjBMI) could capture a genetic
274 predisposition to adiposity that was not evident when using the genetic risk of WHR or BMI
275 alone.¹⁶ Meanwhile, their study was limited to combining only two metrics and was not specifically
276 aimed at developing PRS that captures whole dimensions of underlying etiologies. We integrated
277 a wider range of measures (22 measures), creating a metabolic PRS that is more comprehensive
278 and better equipped to reflect the complex and heterogeneous nature of metabolic disease
279 manifested by obesity and T2D.

280 O-MetPRS and D-MetPRS demonstrated superior predictive accuracy compared to
281 previously developed PRSs for obesity and T2D in the PGS catalog in both internal and external
282 validations (**Fig. 3 and 4**). The genome-wide PRS for obesity (GPS_{BMI}) introduced by Khera et
283 al.(PGS000027),⁷ which utilized 2.1 million genetic variants, had previously set the benchmark for
284 obesity PRS performance. However, our O-MetPRS showed a significant improvement, with an
285 over 20% increase in OR/SD compared to GPS_{BMI} (GPS_{BMI} OR/SD 1.79 vs. O-MetPRS OR/SD
286 2.22 in European population) (**Fig. 3**). This substantial enhancement could be attributed to the
287 difference in genetic inputs and reflecting their biological relevance (**Extended Data Fig.8**); O-
288 MetPRS incorporated multiple ancestry GWAS from 22 traits, whereas GPS_{BMI} relied on only BMI
289 and European GWAS data. Similarly, D-MetPRS demonstrated over 20% increase in OR/SD
290 compared to the second highest T2D PRS—AnnoPred PRS (PGS001357) (AnnoPred OR/SD

291 1.69 vs. O-MetPRS OR/SD 2.08 in European population).¹⁷ Such improvement over previous top
292 scores for both obesity and T2D underscores the value of our multiplicative approach.

293 Obesity and T2D are not only major health concerns in their own right but also strong risk
294 factors for numerous morbidities,^{18–22} making it important to predict their related comorbidities for
295 better health management. Our study demonstrates that MetPRS had a substantial improvement
296 in predicting downstream morbidities and complications of diabetes compared to existing PRSs
297 (**Fig. 5 and Extended Data Fig.6**), likely due to its biologically enriched composition. Moreover,
298 MetPRS effectively predicted cardiovascular mortality, likely indicating its ability to identify
299 individuals at risk for poorer outcomes or prognosis, beyond merely predicting disease incidence.
300 Few AI-based tools and clinical scores have been developed to predict obesity-related or diabetic
301 complications, yet they require clinical factors that usually become evident in middle age as
302 inputs.^{23–25} In contrast, MetPRS can predict morbidities long before the clinical factors manifest,
303 potentially as early as birth.⁷ As sequencing technologies advance and become more cost-
304 effective and affordable,^{26,27} the use of MetPRS can shift morbidity care earlier in life and support
305 primordial prevention.

306 The MetPRS was able to predict the likelihood of requiring GLP-1 receptor agonist
307 prescription in a population free of obesity and T2D at baseline, over a period of following up to
308 13 years. The GLP-1 receptor agonists are indicated for treating obesity and T2D, yet their
309 indications are expanding as mounting evidence shows their efficacy in reducing cardiovascular
310 diseases,²⁸ steatohepatitis,²⁹ chronic kidney disease,³⁰ and other conditions.^{31,32} Despite its
311 growing status as a widely used treatment, the challenges in the pharmacological intervention lie
312 in evaluating the risk-benefit ratio for each individual, as GLP-1 receptor agonists are not without
313 adverse effects.^{33,34} An even greater concern is the significant economic burden these
314 medications impose on healthcare systems, particularly in countries like the US, where up to 40%
315 of the population may be eligible for treatment.^{35,36} Therefore, the ability to extend genetic
316 prediction to GLP-1 receptor agonist prescription holds clinical implications, adding a data point
317 for risk-benefit assessment, patient-physician communication, and cost-effective rollout. Currently,
318 GLP-1 receptor agonists are typically reserved for individuals with obesity or diabetes who have
319 comorbidities or have not achieved sufficient control through lifestyle modifications.^{37,38} The ability
320 to genetically predict the eventual need for this drug could potentially move the needle in
321 management plans and health system strategies.

322 MetPRS is a multi-ancestry PRS that can be readily applied to diverse populations, as it
323 showed robust prediction performance across different ancestries in our external validation data
324 sets (**Fig. 4**). The Eurocentric bias inherent in many genetic studies has significantly limited the
325 generalizability of PRSs, presenting a major barrier to their implementation in real-world settings
326 where populations are inherently diverse.^{39–42} Our study addressed this challenge by
327 incorporating GWAS data from multiple ancestries when developing the model, ensuring that
328 genetic information from non-European populations is represented. We then validated the
329 MetPRS across multiple ancestries, demonstrating consistent best performances across these
330 populations. Notably, even though some existing PRSs were optimized for their specific non-
331 European ancestries, our multi-ancestry MetPRS still outperformed them in respective non-

332 European populations (**Fig. 3 and Fig. 4**). Given its reliable performance among various
333 ancestries, MetPRS can be readily implemented in ethnically diverse settings and contexts. Our
334 findings contribute to the growing body of evidence that multi-ancestry genetic studies can
335 facilitate more inclusive and widely applicable health services and alleviate the risk of propagating
336 potential healthcare disparities in genomic medicine.^{8,43}

337 There are several limitations that must be acknowledged. First, while we incorporated
338 multi-ancestry and non-European GWAS data to train MetPRS, the majority of the data used for
339 model development and validation is still derived from GWAS of European descent, which may
340 affect the score's performance in non-European populations. Reassuringly, MetPRS consistently
341 performed better than pre-existing scores across ancestries on a relative scale. However, its
342 performance was suboptimal in individuals of African ancestry, and thus we developed an African-
343 optimized MetPRS, which led to a modest improvement in performance. Second, our study
344 centered on obesity and T2D among metabolic traits. Future research could explore the inclusion
345 of additional metabolic traits or different combinations based on biological commonality. However,
346 there are several justifications for our focus. We specifically targeted obesity and T2D due to their
347 closely related pathophysiology^{1,2} and shared treatment options, including GLP-1 receptor
348 agonists and metabolic-bariatric surgery.⁴⁴⁻⁴⁶ Convergence in clinical management was a
349 particularly important consideration, as our aim for MetPRS was to design a PRS that can predict
350 across various layers of clinical aspects. The intersecting foundation in biology, etiology, and
351 clinical workflow pertinent to both conditions enabled MetPRS to be a clinically relevant PRS.
352 Finally, the clinical implementation of MetPRS requires further consideration of how to interpret
353 and communicate PRS results to healthcare providers and patients, as well as how to measure
354 its impact on patient outcomes and healthcare decision-making.⁴⁷

355 **Conclusion**

356 This study presented a novel approach to genetic risk prediction of diabetes by introducing
357 the MetPRS, a composite metabolic PRS that integrates multi-ancestry, multi-disease, and multi-
358 measure genetic data. Our findings have significant implications for precision medicine, offering
359 more comprehensive scores to predict risks for obesity and T2D, as well as their complications
360 and clinical management. The MetPRS outperformed traditional single-trait PRS models,
361 demonstrating the value of a biologically-enriched approach to genetic risk prediction. By
362 leveraging multi-ancestry GWAS data, MetPRS represents a step forward in creating more
363 equitable and inclusive tools for genomic medicine.

364

365

366

367

368 **Methods**

369 **Study cohorts**

370 *UK Biobank*

371 The UK Biobank (UKB) is an observational study that enrolled over 500,000 individuals
372 aged 40 to 69 across the United Kingdom between 2006 and 2010. At the time of recruitment,
373 participants provided electronically-signed consent, completed questionnaires on socio-
374 demographic, lifestyle, and health-related factors, and underwent various physical
375 assessments.⁴⁸ Anthropometric measurements, including body mass index, were measured at
376 the initial enrollment visit. Among participants, 43,521 underwent MRI imaging between 2014 and
377 2020.^{49,50} Participants were genotyped using the UK BiLEVE Axiom Array or the UK Biobank
378 Axiom Array, each containing over 800,000 genome-wide variants. Imputation was carried out
379 using reference panel from the Haplotype Reference Consortium (HRC),⁵¹ UK10K,⁵² and the 1000
380 Genome Phase 3 data.⁵³ The HRC was used as the primary imputation reference panel due to
381 its large sample size (64,976 broadly European haplotypes). The 1000 Genomes phase 3 dataset
382 was employed to assist with the phasing of samples from non-European ancestries. Imputation
383 results were also combined from the merged UK10K and 1000 Genomes phase 3 reference
384 panels and the HRC panel, with HRC imputation being prioritized when a SNP was present in
385 both panels. Imputation was conducted using IMPUTE 4 (<https://jmarchini.org/software/>).

386

387 *All of Us*

388 The All of Us (AoU) Research Program is a longitudinal cohort study that has enrolled a
389 diverse group of traditionally underrepresented individuals aged 18 and older from more than 730
390 sites across the United States.⁵⁴ Since its inception in 2018, the AoU program has consented to
391 over 800,000 participants, with more than 560,000 completing the initial enrollment process,
392 including health questionnaires and biospecimen collection. For these participants, ongoing
393 linkage to electronic health record (EHR) data, including International Classification of Diseases
394 (ICD)-9/ICD-10, Systematized Nomenclature of Medicine (SNOMED), and Current Procedural
395 Terminology (CPT) codes, is maintained. This study utilized whole-genome sequencing (WGS)
396 data from over 245,000 participants available in the Controlled Tier Dataset, version 7 release.
397 Data analysis was conducted on the All of Us Researcher Workbench under the guidelines
398 defined by the All of Us Ethical Conduct of Research Policy.

399

400 *Mass General Brigham Biobank*

401 The Mass General Brigham Biobank (MGBB) is a large-scale, healthcare-affiliated
402 biobank based in the Greater Boston area of Massachusetts.^{55,56} Established in 2010, the MGBB
403 has enrolled approximately 140,000 diverse individuals within the Mass General Brigham (MGBB)
404 network, the largest healthcare system in Massachusetts. The biobank aims to uncover the
405 complex relationships among genomic profiles, environmental factors, and disease
406 manifestations in clinical practice. MGBB provides biomedical samples, such as plasma, serum,
407 DNA, buffy coats, collected from patients and linked to clinical data from the EHR, quantitative

408 data derived from medical images, and survey data on lifestyle, environment, and family history.⁵⁷
409 Genome-wide genotyping arrays of 53,306 MGBB participants were used in this study.

410

411 *King Faisal Specialist Hospital and Research Center cohort*

412 The King Faisal Specialist Hospital and Research Center (KFSH&RC) cohort consists of
413 indigenous Arabs from Saudi Arabia referred for cardiology care at KFSH&RC, a tertiary care
414 hospital recognized throughout Saudi Arabia and the Middle East, for work-up of cardiac disease
415 from all over the country⁵⁸. In addition to blood samples for DNA extraction and genotyping array
416 analyses, participants provided access to their electronic health records for phenotype data and
417 shared socio-demographic information through a clinical research coordinator. All participants
418 gave informed consent to participate in the study.

419

420 **Phenotype definitions**

421

422 For the UK Biobank participants, weight and height were collected at baseline when
423 participants attended the initial assessment center. BMI was constructed from height and weight
424 measured during the initial Assessment Centre visit. BMI is not present if either of these readings
425 were omitted. Obesity was defined as a BMI ≥ 25 kg/m² for individuals of East Asian ancestry and
426 a BMI ≥ 30 kg/m² for individuals of other ancestries,^{59,60} which was used also for AoU, MGBB and
427 KFSH&RC cohorts. T2D was identified either through self-reported diagnosis during an interview
428 with a trained nurse or via the ICD-10 code E11.X in hospitalization records. For comorbidities,
429 detailed definitions of diseases included in the comorbidities analysis are provided in
430 **Supplementary Table 17**. For example, coronary artery disease cases were defined centrally
431 based on self-report at enrollment, hospitalization records, or death registry records; retinopathy
432 case ascertainment was based on hospitalization records and death registry records; sleep apnea
433 ascertainment was based on self-report at the time of enrollment, hospitalization records, and
434 death registry records.

435 For All of Us, BMI was taken from physical measurements of EHR data. T2D was identified
436 as EHR records containing the term “Type 2 diabetes mellitus” in Controlled Tier Dataset v7
437 Conditions. For MGBB, BMI was obtained from the electronic health records T2D was identified
438 as glycated hemoglobin $\geq 6.5\%$ or diagnosis records from the electronic health records. For
439 the KFSH&RC cohort, disease status and metabolic traits were collected during routine clinical
440 care and manually extracted from the electronic health record by trained medical personnel. The
441 diagnosis of T2D was curated from the electronic health records. Height and weight
442 measurements obtained in the context of care and the first visit measured on referral to the King
443 Faisal Specialist Hospital and Research Center was used.

444

445 **Define polygenic risk score percentiles**

446

447 Residual scores for MetPRS, and PRS scores calculated by the variant weight from the
448 PGS Catalog were extracted after adjustment for the first ten principal components of genetic
449 ancestry, then normalized to a mean of 0 with a standard deviation of 1. The percentile distribution

450 of the polygenic risk scores was analyzed to assess the stratification of obesity (BMI \geq 30 kg/m²)
451 and T2D.

452

453 **Quality control and sample selection**

454

455 *UK Biobank*

456 Participants were excluded from the analysis if they met any of the following criteria: (1) a
457 mismatch between self-reported sex and genetically inferred sex, (2) sex chromosome aneuploidy,
458 (3) a genotyping call rate below 0.95, (4) being outliers for heterozygosity, (5) having second-
459 degree or closer relatives, (6) with unknown self-reported ancestry, or (7) without BMI or T2D
460 phenotypes. After applying these sample quality controls, 444,345 participants remained
461 available (**Extended Data Fig.1**).

462 To further boost the power for constructing PRS scores, we conducted de novo GWAS for
463 ten traits (BMI, T2D, body fat percentage, height, body weight, waist circumference, hip
464 circumference, HbA1c, Glucose, cT1) traits utilizing data from 378,921 participants randomly
465 selected from UKB European ancestry participants (self-reported ancestry, Field ID 21000). For
466 the remaining participants of European ancestry, 15,000 and 34,626 individuals were assigned to
467 the training and testing sets for PRS score development, respectively. A total of 1,304 individuals
468 who self-identified as Chinese were included for East Asian ancestry validation, 6,842 individuals
469 who self-identified as Black were included for African ancestry validation, and 6,605 individuals
470 of South Asian descent (self-reported as Indian, Pakistani, or Bangladeshi) were included for
471 South Asian ancestry validation. Individuals with body (abdominal) magnetic resonance imaging
472 (MRI) data (UK Biobank Field ID 12224) were excluded from the training and validation datasets,
473 since GWAS for MRI-derived visceral (VAT), abdominal subcutaneous (ASAT), and gluteofemoral
474 (GFAT) adipose tissue volumes was conducted on UKB MRI-tested participants;⁶¹ we additionally
475 excluded individuals with whole body DXA scan (UK Biobank Field ID 20201), (**Extended Data**
476 **Fig.1**).

477 Genetic variants were included in the GWAS if they met the following quality control criteria:
478 (1) imputation quality score (INFO) > 0.3, (2) minor allele frequency (MAF) > 0.001, (3) minor
479 allele count (MAC) > 100, and (4) SNP genotyping missing rate < 10%. These criteria resulted in
480 a total of 15,848,715 imputed variants available for analysis. Common variant association studies
481 for ten traits (BMI, T2D, body fat percentage, height, body weight, waist circumference, hip
482 circumference, HbA1c, Glucose, cT1) were conducted using REGENIE (v3.2.7), genotype
483 variants were centrally imputed from UK Biobank.⁴⁸

484

485 *All of Us*

486 Within All of Us, details of whole genome sequencing and quality control are described
487 extensively in Jurgens et al.⁶² and the All of Us Genomic Research Data Quality Report
488 C2022Q4R9 CDR v7, available at [https://support.researchallofus.org/hc/en-](https://support.researchallofus.org/hc/en-us/articles/4617899955092-All-of-Us-Genomic-Quality-Report)
489 [us/articles/4617899955092-All-of-Us-Genomic-Quality-Report](https://support.researchallofus.org/hc/en-us/articles/4617899955092-All-of-Us-Genomic-Quality-Report). Briefly, 229,517 out of total
490 245,394 samples remained after filtering for sex concordance, a cross individual contamination
491 rate below 3%, a call rate above 98% and related samples by AoU Central quality control. We
492 used the Allele Count/Alele Frequency (ACAF) threshold SNP callset curated by AoU, which

493 includes SNPs of population-specific allele frequency > 1% or allele counts over 100 for each
494 ancestral subpopulation. The inferred ancestry information is obtained from the “ancestry_pred”
495 column, which is available as part of the genetically predicted ancestry TSV file. The quality
496 control metrics were centrally provided by the AoU project.⁵⁴ SNPs deviated from Hardy-Weinberg
497 equilibrium ($P < 1 \times 10^{-6}$) in each genetic ancestry were removed. Principal components of genetic
498 ancestry used for the correction of population structure were calculated in each ancestry group
499 separately using SNPs present in the 1000 Genomes project phase 3 release.

500

501 *Mass General Brigham Biobank*

502 Within MGB Biobank, 53,306 individuals were genotyped by Illumina Global Screening
503 Array (Illumina, CA) and imputed to the TOPMed multi-ancestry imputation reference panel
504 (TOPMed r2 panel).⁵⁶ Variants with high missingness (> 2%) and low MAF (< 1%) and variants
505 that failed the Hardy–Weinberg test ($P < 1 \times 10^{-6}$) were removed. After excluding individuals due
506 to discrepancies between reported and genotypic sex, sex chromosome aneuploidy, genotyping
507 call rate < 0.95, heterozygosity outliers, and excess relatedness (second-degree relatives or
508 closer), 49,825 individuals remained for validation. We inferred genetic ancestry using a public
509 diverse population, including 3,380 unrelated individuals from the 1000 Genomes Project (1KG)
510 and the Human Genome Diversity Project (HGDP), called 1KG_HGDP.⁶³ We extracted common,
511 high-quality SNPs (missingness < 1%, MAF > 1%) across MGBB and the 1KG_HGDP dataset.
512 After pruning SNPs, we computed SNP weights for the genetic principal component using the
513 1KG_HGDP dataset. Then, we projected MGBB participants into the same principal component
514 space using the top 20 PCs. Using genetic PCs in the 1KG_HGDP dataset as a feature matrix,
515 we trained a K-nearest neighbor model (k=1) for 1KG_HGDP reference populations to assign
516 population labels to MGBB participants.

517

518 *King Faisal Specialist Hospital and Research Center cohort*

519 All participants from the King Faisal Specialist Hospital and Research Center cohort were
520 of Middle Eastern ancestry and variants and sample quality control were previously described.⁵⁸
521 Briefly, we excluded samples due to variant calling missingness > 5%, heterozygosity rate > 5
522 standard deviations above the mean, and mismatch between genotypically-determined and self-
523 reported sex. Variant-level quality control was conducted to remove variants with call rate < 98%,
524 MAF < 0.01, or Hardy–Weinberg test ($P < 1 \times 10^{-6}$). After imputation by TOPMed r2 panel datasets,
525 the variants with INFO < 0.3 and MAF < 0.01 were removed.

526

527 **Genetic correlation estimation between traits**

528

529 We selected 22 traits that directly or indirectly serve as proxies for the status of obesity
530 and T2D (**Supplementary Table 1**). Our focus on obesity and T2D among various metabolic
531 diseases is due to their closely related pathophysiology and shared treatment options, including
532 GLP-1 receptor agonists and metabolic-bariatric surgery. This common biological, etiological, and
533 clinical foundation justified the development of a composite metabolic PRS for diabetes. We
534 defined a ‘trait’ as a composite of binary disease status (e.g., T2D) and continuous measures for
535 metabolic status (e.g., BMI, HbA1c).

536 Genetic correlations between 22 traits were estimated using cross-trait LD-score
537 regression (LDSC v1.0.1). Precomputed LD scores were obtained from ~1.2 million common
538 SNVs in the well-imputed HapMap3 variants panel.⁶⁴ For each trait, only GWAS with the largest
539 sample size of European individuals was used to perform cross-trait LD-score regression analysis.
540 A Bonferroni corrected P value threshold of 0.00227 (0.05/22) was used to define statistical
541 significance.

542

543 **Polygenic risk score derivation**

544

545 Summary statistics from recent obesity and T2D-related GWAS studies (**Supplementary**
546 **Table 1**) conducted in diverse ancestries were used to determine the primary weights for obesity
547 and T2D PRSs. To ensure an independent holdout dataset for training and validating the obesity
548 and T2D metabolic polygenic risk scores (O-MetPRS and D-MetPRS), UK Biobank participants
549 for de novo GWAS were excluded for PRS score development (**Extended Data Fig.1**). Ancestry-
550 specific linkage disequilibrium (LD) reference panels were derived from the 1000 Genomes
551 Project phase 3 data to match the ancestry of the discovery GWAS.

552 The construction of O-MetPRS and D-MetPRS involved a two-layer process (**Fig.1B**). In
553 layer 1, multiple PRS derived from ancestry-specific GWAS data were combined to yield a multi-
554 ancestry PRS for each trait. In layer 2, the multi-ancestry PRSs for 22 traits were further combined
555 to generate MetPRS.

556 In layer 1, separate PRSs were constructed based on each ancestry-stratified GWAS
557 using the LDpred2 method, a Bayesian approach that calculates posterior mean effects for all
558 variants, considering prior GWAS effect sizes and subsequent shrinkage based on LD.⁶⁵ Only
559 HapMap3 variants—a set of 1,296,172 variants representing common genetic variation across
560 diverse populations—were included in the score calculation.

561 The LDpred2 method used default parameters, including proportion of causal variant
562 (values of $P = 1.0 \times 10^{-4}$, 1.8×10^{-4} , 3.2×10^{-4} , 5.6×10^{-4} , 1.0×10^{-3} , 1.8×10^{-3} , 3.2×10^{-3} , 5.6
563 $\times 10^{-3}$, 1.0×10^{-2} , 1.8×10^{-2} , 3.2×10^{-2} , 5.6×10^{-2} , 1.0×10^{-1} , 1.8×10^{-1} , 3.2×10^{-1} , 5.6×10^{-1}
564 and 1), heritability scales ($s = 0.7$, 1, and 1.4), and whether a sparse LD matrix was applied.
565 These parameter combinations yielded 102 candidate PRSs for each GWAS summary statistic.
566 Genotypes were extracted from a centrally imputed data repository by the UK Biobank, processed
567 using bgenix,⁶⁶ and PRSs were calculated for each individual in the UK Biobank using PLINK
568 2.0.⁶⁷ The best PRS was selected based on its performance in predicting obesity or T2D in an
569 independent sample of 15,000 White British individuals from UK Biobank. The obesity model used
570 log(BMI) as the outcome, while the diabetes model used T2D. The best score for each ancestry
571 was selected based on predictive performance measured by odds ratio (binary traits) or
572 incremental R^2 (continuous traits). The discriminative capacities (AIC) were then used to
573 determine the optimal combination of different ancestry scores. A regression model estimated the
574 mixing weights for ancestry-specific PRSs, which were then linearly combined into a single multi-
575 ancestry score. Similar procedures were followed for other diabetes-related traits.

576 In layer 2, these multi-ancestry trait-specific PRSs were linearly combined with the multi-
577 ancestry obesity/T2D PRSs from layer 1 to generate the final O-MetPRS and D-MetPRS. Same
578 as the layer 1, stepAIC was used for the feature selection for predicting log(BMI) or T2D to identify

579 the best combination of trait-level scores for mixing. Then, a regression model was used to
580 estimate the mixing weights for each individual trait-specific PRS as described above, which were
581 then linearly combined into a single MetPRS (O-MetPRS or D-MetPRS). Among the 93 ancestry-
582 and trait-specific scores analyzed in the MetPRS training, 31 and 37 scores significantly
583 contributed to the overall prediction in O-MetPRS and D-MetPRS, respectively, after optimization
584 with stepAIC and weighting via regressions across both layers. The final mixing weights for each
585 layer were summarized in **Extended Data fig 8**, and were listed in **Supplementary Table 15 and**
586 **16**.

587

588 **Polygenic risk score validation**

589

590 MetPRS underwent internal hold-out validation within UK Biobank. To evaluate its
591 performance, we compared it with previously published polygenic risk scores for BMI and T2D
592 from the PGS Catalog (<https://www.pgscatalog.org/>). The variant effect sizes for these scores
593 were obtained from the PGS Catalog and calculated within the same UK Biobank testing dataset,
594 which included 34,626 European ancestry samples, 6,842 of African ancestry, 1,304 of East Asian
595 ancestry, and 6,605 South Asian ancestry samples (**Extended Data Fig.1**). External validation
596 was also performed in the All of Us (n=245,394), MGB Biobank (n=53,306), and a KFSH&RC
597 cohort (n=6,416).

598

599 **GLP-1 agonist prescription and metabolic-bariatric surgery**

600

601 We conducted a comprehensive analysis using data from the MGB Biobank to evaluate
602 the predictive utility of MetPRS for incident GLP-1 receptor agonist prescriptions and metabolic-
603 bariatric surgeries. GLP-1 receptor agonist prescription and metabolic-bariatric surgery were
604 selected as clinical endpoints as they are interventions indicated for obesity and/or T2D. Initial
605 datasets were filtered to include only individuals with complete genotypic and phenotypic data,
606 excluding those censored before the biobank enrollment date, leaving 51,779 participants for
607 analysis. For each participant, O-MetPRS and D-MetPRS were calculated using ancestry-
608 adjusted models. The first ten principal components were regressed out of the raw O-MetPRS
609 and D-MetPRS using linear regression to account for population stratification. The residuals from
610 this regression, which represent the MetPRS adjusted for ancestry, were then standardized to
611 have a mean of 0 and a standard deviation of 1.

612 To assess the association of MetPRS with clinical interventions, we utilized Cox
613 proportional hazards regression models adjusted for age, sex, and ancestry PCs. We stratified
614 participants into three risk groups based on their MetPRS values: top 10%, middle quintile (40–
615 60%), and bottom 10%. The primary outcomes were the cumulative incidence of GLP-1 receptor
616 agonist prescriptions and metabolic-bariatric surgery over the follow-up period. Incidence was
617 tracked from baseline until the first event or censoring. Competing risk models were applied to
618 calculate cumulative incidence functions, and Kaplan-Meier curves were generated to visualize
619 time-to-event data.

620

621 **Statistical analysis**

622 For score training and validations, prediction for a continuous trait was calculated using
623 linear regression models, while logistic regression models were employed for predicting the risk
624 of a binary trait. All the models included baseline covariates: enrollment age, sex, genotyping
625 array, and the first ten principal components of genetic ancestry. For the comorbidities analysis,
626 the top decile of the PRS distribution was labeled as "carriers," while the remaining individuals
627 were labeled as "non-carriers". The top five obesity and T2D scores in the score comparisons
628 among 34,626 European individuals from the UK Biobank were compared with the O-MetPRS
629 and D-MetPRS, respectively. Cox proportional-hazards models were used to estimate hazard
630 ratios (HRs) for incident diseases in the UK Biobank, with enrollment age, genetically inferred sex,
631 genotyping array, and the first ten principal components as covariates. The frequency of the
632 highest prediction for each score was then calculated. All statistical analyses were performed
633 using R software (version 4.2.2, R Project for Statistical Computing), and figures were generated
634 using the ggplot2 R package (version 3.4.2).

635

636 **Data availability**

637

638 All data are made available from the UK Biobank ([https://www.ukbiobank.ac.uk/enable-](https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access)
639 [your-research/apply-for-access](https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access)) to researchers from universities and other institutions with
640 genuine research inquiries following institutional review board and UK Biobank approval. This
641 research was conducted using the UK Biobank resource under application number 89885. All of
642 Us data are made available from the *All of Us* Research Study to researchers from universities
643 and other institutions with genuine research inquiries following institutional review board and *All*
644 *of Us* approval. KFSH&RC and MGBB data are governed by local laws, so relevant data could be
645 made available by contacting the investigators. The genome-wide association data supporting the
646 findings of this study are publicly available in KoGES (<https://koges.leelabsg.org/>), Biobank Japan
647 (<http://jenger.riken.jp/en/result>), FinnGen (https://www.finnngen.fi/en/access_results), AGEN T2D
648 (<https://kp4cd.org/index.php/node/309>), GIANT
649 (<https://portals.broadinstitute.org/collaboration/giant/>) and Million Veteran Program (via dbGaP at
650 <https://ftp.ncbi.nlm.nih.gov/dbgap/studies/>, under accession number phs001672). The research
651 was approved by the Beijing Institute of Genomics (Chinese Academy of Sciences) and the China
652 National Center for Bioinformatics institutional review board, and the Mass General Brigham
653 institutional review board.

654

655 **Acknowledgments**

656

657 P.N. is supported by grants from the National Institutes of Health (R01HL127564, U01HG011719),
658 P.T.E. is supported by grants from the National Institutes of Health (1R01HL092577,
659 1R01HL157635, 5R01HL139731), from the American Heart Association Strategically Focused
660 Research Networks (18SFRN34110082), and from the European Union (MAESTRIA 965286).
661 P.N. is supported by grant from National Human Genome Research Institute (U01HG011719).
662 A.C.F. is supported by grants from the National Heart Lung and Blood Institute (K08HL161448
663 and R01HL164629). Dr. Wang is supported by the Pioneering Action Grants of the Chinese
664 Academy of Sciences. We gratefully acknowledge all participants of UKB, AoU, MGBB and

665 KFSH&RC for their contributions, without whom this research would not have been possible. We
666 also thank the National Institutes of Health's AoU Research Program, the UKB resource, the MGB
667 team and KFSH&RC team, for making available the participant data examined in this study.
668 Further, we acknowledge the dedication and expertise of the research coordinators and the
669 comprehensive Biobank team, whose efforts in participant enrollment and data collection were
670 indispensable.

671
672 **Author contributions**
673

674 Concept and design: M.S.K, Q.C., P.T.E., A.C.F. and M.W.. Acquisition, analysis or interpretation
675 of data: M.S.K, Q.C., Y.S., X.Y., S.W., L.C.W., S.M.J.C., S.K., X.Y., K.Y., X.C., R.Z., W.Y., S.L.,
676 Z.L., A.C.F. and M.W. Drafting of the manuscript: M.S.K, Q.C., Y.S., A.C.F. and M.W. Critical
677 revision of the manuscript for important content: Z.L., F.A., P.N.
678

679 **Declaration of interests**
680

681 P.T.E. has received sponsored research support from Bayer AG, IBM Health, Bristol Myers
682 Squibb and Pfizer and Novo Nordisk. P.T.E. has also served on advisory boards or consulted for
683 Bayer AG, MyoKardia, and Novartis. P.N. reports research grants from Allelica, Amgen, Apple,
684 Boston Scientific, Genentech / Roche, and Novartis, personal fees from Allelica, Apple,
685 AstraZeneca, Blackstone Life Sciences, Creative Education Concepts, CRISPR Therapeutics, Eli
686 Lilly & Co, Esperion Therapeutics, Foresite Labs, Genentech / Roche, GV, HeartFlow, Magnet
687 Biomedicine, Merck, Novartis, TenSixteen Bio, and Tourmaline Bio, equity in Bolt, Candela,
688 Mercury, MyOme, Parameter Health, Preciseli, and TenSixteen Bio, and spousal employment at
689 Vertex Pharmaceuticals, all unrelated to the present work. A.C.F. reports being co-founder of
690 Goodpath, serving as scientific advisor to MyOme and HeartFlow, and receiving a research grant
691 from Foresite Labs. The remaining authors declare no competing interests.

692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707

708 **References:**

- 709
- 710 1. Ng, A. C., Delgado, V., Borlaug, B. A. & Bax, J. J. Diabesity: the combined burden of obesity
711 and diabetes on heart disease and the role of imaging. *Nat. Rev. Cardiol.* **18**, 291–304
712 (2021).
- 713 2. Chadt, A., Scherneck, S., Joost, H.-G. & Al-Hasani, H. Molecular links between obesity and
714 diabetes: “diabesity”. (2015).
- 715 3. Ingelsson, E. & McCarthy, M. I. Human genetics of obesity and type 2 diabetes mellitus: past,
716 present, and future. *Circ. Genomic Precis. Med.* **11**, e002090 (2018).
- 717 4. Coral, D. E. *et al.* A phenome-wide comparative analysis of genetic discordance between
718 obesity and type 2 diabetes. *Nat. Metab.* **5**, 237–247 (2023).
- 719 5. Fahed, A. C., Philippakis, A. A. & Khera, A. V. The potential of polygenic scores to improve
720 cost and efficiency of clinical trials. *Nat. Commun.* **13**, 2922 (2022).
- 721 6. Hao, L. *et al.* Development of a clinical polygenic risk score assay and reporting workflow.
722 *Nat. Med.* **28**, 1006–1013 (2022).
- 723 7. Khera, A. V. *et al.* Polygenic prediction of weight and obesity trajectories from birth to
724 adulthood. *Cell* **177**, 587–596. e9 (2019).
- 725 8. Patel, A. P. *et al.* A multi-ancestry polygenic risk score improves risk prediction for coronary
726 artery disease. *Nat. Med.* **29**, 1793–1803 (2023).
- 727 9. Maier, R. M. *et al.* Improving genetic prediction by leveraging genetic correlations among
728 human diseases and traits. *Nat. Commun.* **9**, 989 (2018).
- 729 10. Abraham, G. *et al.* Genomic risk score offers predictive performance comparable to clinical
730 risk factors for ischaemic stroke. *Nat. Commun.* **10**, 5819 (2019).
- 731 11. Ross, R. *et al.* Waist circumference as a vital sign in clinical practice: a Consensus
732 Statement from the IAS and ICCR Working Group on Visceral Obesity. *Nat. Rev. Endocrinol.*
733 **16**, 177–189 (2020).

- 734 12. Williamson, A. *et al.* Genome-wide association study and functional characterization
735 identifies candidate genes for insulin-stimulated glucose uptake. *Nat. Genet.* **55**, 973–983
736 (2023).
- 737 13. Lambert, S. A. *et al.* The Polygenic Score Catalog as an open database for reproducibility
738 and systematic evaluation. *Nat. Genet.* **53**, 420–425 (2021).
- 739 14. Committee, A. D. A. P. P. & Committee, A. D. A. P. P. 2. Classification and diagnosis of
740 diabetes: Standards of Medical Care in Diabetes—2022. *Diabetes Care* **45**, S17–S38 (2022).
- 741 15. Lotta, L. A. *et al.* Association of genetic variants related to gluteofemoral vs abdominal fat
742 distribution with type 2 diabetes, coronary disease, and cardiovascular risk factors. *Jama*
743 **320**, 2553–2563 (2018).
- 744 16. Emdin, C. A. *et al.* Genetic association of waist-to-hip ratio with cardiometabolic traits, type 2
745 diabetes, and coronary heart disease. *Jama* **317**, 626–634 (2017).
- 746 17. Ye, Y. *et al.* Interactions between enhanced polygenic risk scores and lifestyle for
747 cardiovascular disease, diabetes, and lipid levels. *Circ. Genomic Precis. Med.* **14**, e003128
748 (2021).
- 749 18. Kim, M. S. *et al.* Causal effect of adiposity on the risk of 19 gastrointestinal diseases: a
750 Mendelian randomization study. *Obesity* **31**, 1436–1444 (2023).
- 751 19. Kim, M. S. *et al.* Association between adiposity and cardiovascular outcomes: an umbrella
752 review and meta-analysis of observational and Mendelian randomization studies. *Eur. Heart*
753 *J.* **42**, 3388–3403 (2021).
- 754 20. Kim, M. S. *et al.* Association of genetic risk, lifestyle, and their interaction with obesity and
755 obesity-related morbidities. *Cell Metab.* **36**, 1494-1503. e3 (2024).
- 756 21. Kim, M. S. *et al.* Integration of observational and causal evidence for the association
757 between adiposity and 17 gastrointestinal outcomes: An umbrella review and meta-analysis.
758 *Obes. Rev.* e13823 (2024).

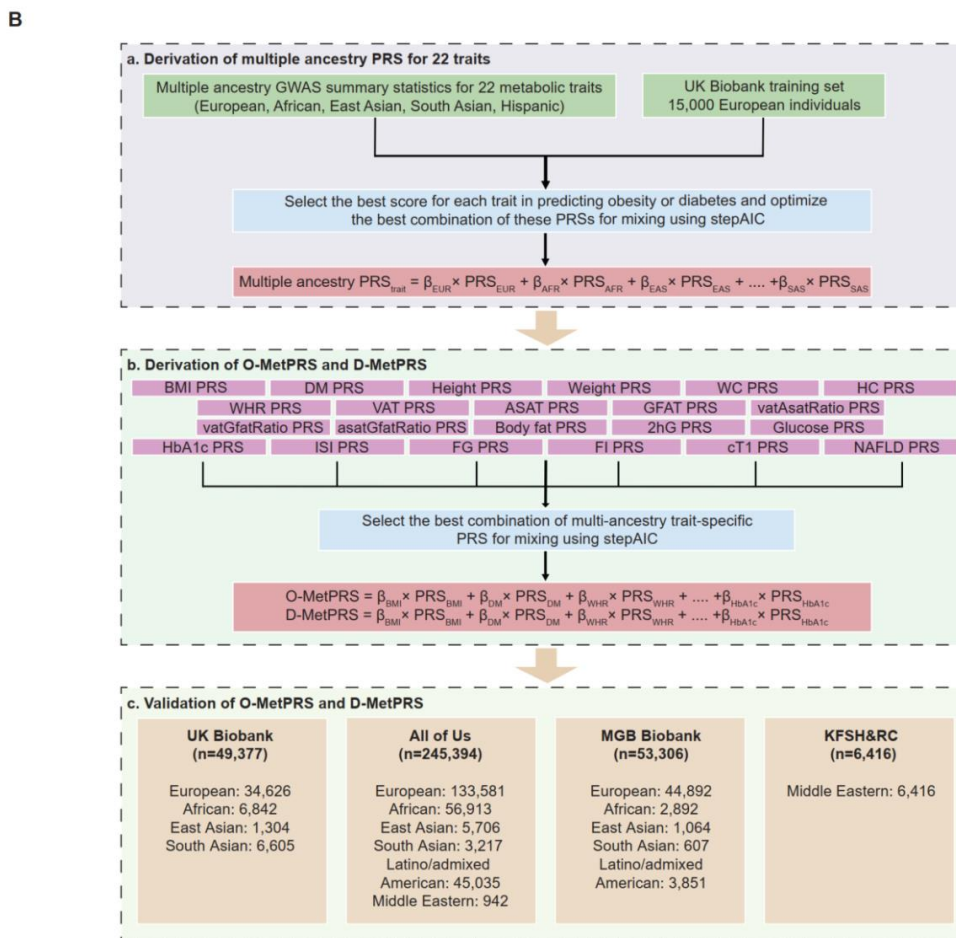
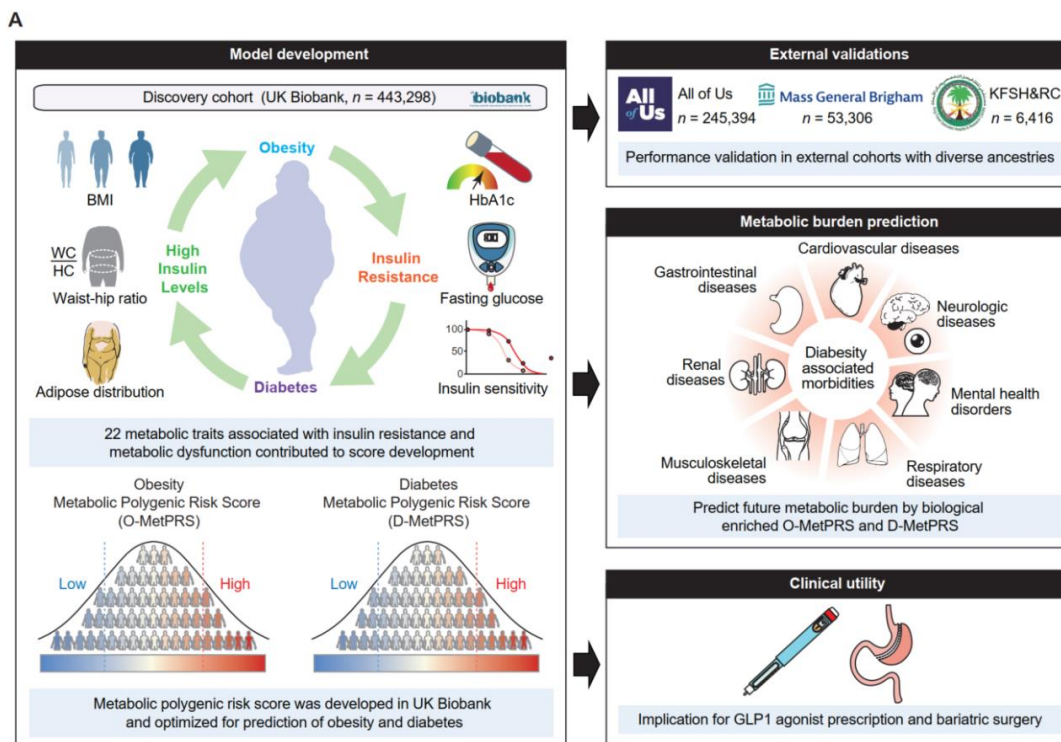
- 759 22. Tsilidis, K. K., Kasimis, J. C., Lopez, D. S., Ntzani, E. E. & Ioannidis, J. P. Type 2 diabetes
760 and cancer: umbrella review of meta-analyses of observational studies. *Bmj* **350**, (2015).
- 761 23. Al-Sari, N. *et al.* Precision diagnostic approach to predict 5-year risk for microvascular
762 complications in type 1 diabetes. *EBioMedicine* **80**, (2022).
- 763 24. Aminian, A. *et al.* Predicting 10-year risk of end-organ complications of type 2 diabetes with
764 and without metabolic surgery: a machine learning approach. *Diabetes Care* **43**, 852–859
765 (2020).
- 766 25. Turchin, A. *et al.* EXIST: EXamining risk of excess adiposity—Machine learning to predict
767 obesity-related complications. *Obes. Sci. Pract.* **10**, e707 (2024).
- 768 26. Satam, H. *et al.* Next-generation sequencing technology: current trends and advancements.
769 *Biology* **12**, 997 (2023).
- 770 27. Kullo, I. J. *et al.* Polygenic scores in biomedical research. *Nat. Rev. Genet.* **23**, 524–532
771 (2022).
- 772 28. Lincoff, A. M. *et al.* Semaglutide and cardiovascular outcomes in obesity without diabetes.
773 *N. Engl. J. Med.* **389**, 2221–2232 (2023).
- 774 29. Sanyal, A. J. *et al.* A Phase 2 Randomized Trial of Survodutide in MASH and Fibrosis. *N.*
775 *Engl. J. Med.* **391**, 311–319 (2024).
- 776 30. Perkovic, V. *et al.* Effects of semaglutide on chronic kidney disease in patients with type 2
777 diabetes. *N. Engl. J. Med.* **391**, 109–121 (2024).
- 778 31. De Giorgi, R. *et al.* 12-month neurological and psychiatric outcomes of semaglutide use for
779 type 2 diabetes: a propensity-score matched cohort study. *EClinicalMedicine* **74**, (2024).
- 780 32. Kosiborod, M. N. *et al.* Semaglutide in patients with heart failure with preserved ejection
781 fraction and obesity. *N. Engl. J. Med.* **389**, 1069–1084 (2023).
- 782 33. He, L. *et al.* Association of glucagon-like peptide-1 receptor agonist use with risk of
783 gallbladder and biliary diseases: a systematic review and meta-analysis of randomized

- 784 clinical trials. *JAMA Intern. Med.* **182**, 513–519 (2022).
- 785 34. Sodhi, M., Rezaeianzadeh, R., Kezouh, A. & Etmnan, M. Risk of gastrointestinal adverse
786 events associated with glucagon-like peptide-1 receptor agonists for weight loss. *Jama* **330**,
787 1795–1797 (2023).
- 788 35. Baig, K., Dusetzina, S. B., Kim, D. D. & Leech, A. A. Medicare Part D Coverage of
789 Antiobesity Medications — Challenges and Uncertainty Ahead. *N. Engl. J. Med.* **388**, 961–
790 963 (2023).
- 791 36. Wright, D. R., Guo, J. & Hernandez, I. A Prescription for Achieving Equitable Access to
792 Antiobesity Medications. *JAMA Health Forum* **4**, e230493 (2023).
- 793 37. Committee, A. D. A. P. P. & Committee, A. D. A. P. P. 9. Pharmacologic approaches to
794 glycemic treatment: Standards of Medical Care in Diabetes—2022. *Diabetes Care* **45**, S125–
795 S143 (2022).
- 796 38. Melson, E., Ashraf, U., Papamargaritis, D. & Davies, M. J. What is the pipeline for future
797 medications for obesity? *Int. J. Obes.* 1–19 (2024).
- 798 39. Kachuri, L. *et al.* Principles and methods for transferring polygenic risk scores across global
799 populations. *Nat. Rev. Genet.* **25**, 8–25 (2024).
- 800 40. Ruan, Y. *et al.* Improving polygenic prediction in ancestrally diverse populations. *Nat. Genet.*
801 **54**, 573–580 (2022).
- 802 41. Ding, Y. *et al.* Polygenic scoring accuracy varies across the genetic ancestry continuum.
803 *Nature* **618**, 774–781 (2023).
- 804 42. Lennon, N. J. *et al.* Selection, optimization and validation of ten chronic disease polygenic
805 risk scores for clinical implementation in diverse US populations. *Nat. Med.* **30**, 480–487
806 (2024).
- 807 43. Koyama, S. *et al.* Population-specific and trans-ancestry genome-wide analyses identify
808 distinct and shared genetic risk loci for coronary artery disease. *Nat. Genet.* **52**, 1169–1177

- 809 (2020).
- 810 44. Kim, M. S. *et al.* Association of bariatric surgery with indicated and unintended outcomes:
811 An umbrella review and meta-analysis for risk–benefit assessment. *Obes. Rev.* **25**, e13670
812 (2024).
- 813 45. Mingrone, G. *et al.* Bariatric–metabolic surgery versus conventional medical treatment in
814 obese patients with type 2 diabetes: 5 year follow-up of an open-label, single-centre,
815 randomised controlled trial. *The Lancet* **386**, 964–973 (2015).
- 816 46. Brown, E., Heerspink, H. J., Cuthbertson, D. J. & Wilding, J. P. SGLT2 inhibitors and GLP-1
817 receptor agonists: established and emerging indications. *The Lancet* **398**, 262–276 (2021).
- 818 47. Vogan, K. Implementing polygenic risk scores in the clinic. *Nat. Genet.* **56**, 557–557 (2024).
- 819 48. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data.
820 *Nature* **562**, 203–209 (2018).
- 821 49. Littlejohns, T. J. *et al.* The UK Biobank imaging enhancement of 100,000 participants:
822 rationale, data collection, management and future directions. *Nat. Commun.* **11**, 2624 (2020).
- 823 50. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide
824 range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
- 825 51. the Haplotype Reference Consortium. A reference panel of 64,976 haplotypes for genotype
826 imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
- 827 52. The UK10K Consortium *et al.* The UK10K project identifies rare variants in health and
828 disease. *Nature* **526**, 82–90 (2015).
- 829 53. The 1000 Genomes Project Consortium *et al.* A global reference for human genetic
830 variation. *Nature* **526**, 68–74 (2015).
- 831 54. All of Us Research Program Genomics Investigators. Genomic data in the All of Us
832 Research Program. *Nature*, 627(8003):340-346 (2024)
- 833 55. Boutin, N. T. *et al.* The evolution of a large biobank at Mass General Brigham. *J. Pers. Med.*

- 834 **12**, 1323 (2022).
- 835 56. Koyama, S. *et al.* Decoding Genetics, Ancestry, and Geospatial Context for Precision
836 Health. *medRxiv* 2023.10.24.23297096 (2023).
- 837 57. Bhattacharya, R. *et al.* Clonal Hematopoiesis Is Associated With Higher Risk of Stroke.
838 *Stroke* **53**, 788–797 (2022).
- 839 58. Shim, I. *et al.* Clinical utility of polygenic scores for cardiometabolic disease in Arabs. *Nat.*
840 *Commun.* **14**, 6535 (2023).
- 841 59. Haam, J.-H. *et al.* Diagnosis of Obesity: 2022 Update of Clinical Practice Guidelines for
842 Obesity by the Korean Society for the Study of Obesity. *J. Obes. Metab. Syndr.* **32**, 121–129
843 (2023).
- 844 60. Hassapidou, M. *et al.* European Association for the Study of Obesity Position Statement on
845 Medical Nutrition Therapy for the Management of Overweight and Obesity in Adults
846 Developed in Collaboration with the European Federation of the Associations of Dietitians.
847 *Obes. Facts* **16**, 11–28 (2023).
- 848 61. Agrawal, S. *et al.* Inherited basis of visceral, abdominal subcutaneous and gluteofemoral fat
849 depots. *Nat. Commun.* **13**, 3771 (2022).
- 850 62. Jurgens, S. J. *et al.* Rare coding variant analysis for human diseases across biobanks and
851 ancestries. *Nat. Genet.* **56**, 1811–1820 (2024).
- 852 63. Koenig, Z. *et al.* A harmonized public resource of deeply sequenced diverse human
853 genomes. *Genome Res.* **34**, 796–809 (2024).
- 854 64. Schizophrenia Working Group of the Psychiatric Genomics Consortium *et al.* LD Score
855 regression distinguishes confounding from polygenicity in genome-wide association studies.
856 *Nat. Genet.* **47**, 291–295 (2015).
- 857 65. Privé, F., Arbel, J. & Vilhjálmsson, B. J. LDpred2: better, faster, stronger. *Bioinformatics* vol.
858 36 5424–5431 (2020).

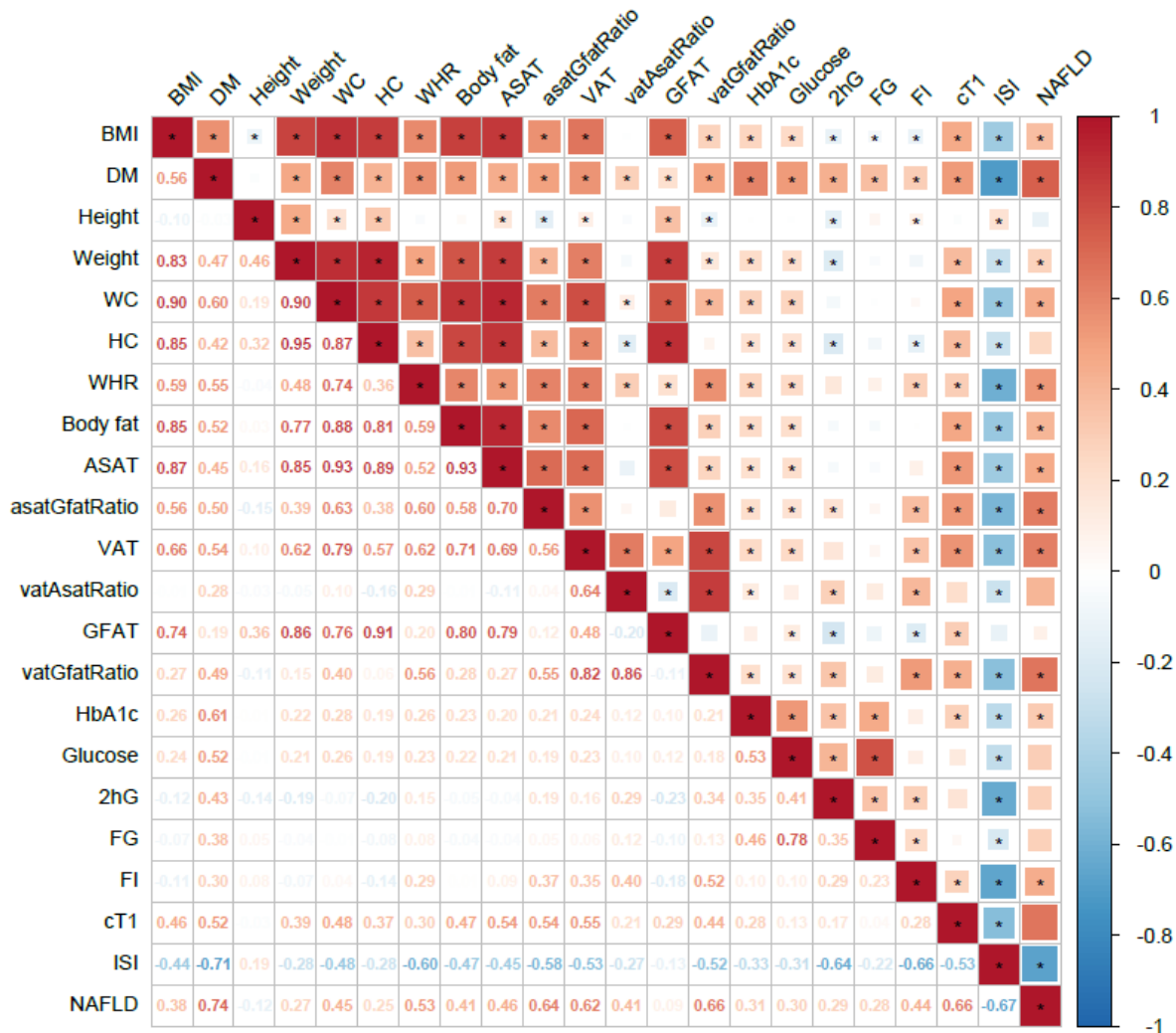
- 859 66. Band, G. & Marchini, J. BGEN: a binary file format for imputed genotype and haplotype
860 data. Preprint at <https://doi.org/10.1101/308296> (2018).
- 861 67. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer
862 datasets. *GigaScience* **4**, 7 (2015).



864 **Fig.1 | Study overview.**

865 Overview of study **(A)** and MetPRS development **(B)**. Polygenic risk scores were developed using
866 cohort-specific, ancestry-stratified summary statistics for 22 metabolic traits. (a) For each trait,
867 such as BMI, the optimal combination of cohort-specific, ancestry-stratified PRSs was identified
868 using stepAIC, the selected PRSs were then combined linearly to create multi-ancestry scores
869 for predicting target trait (Obesity or Type 2 diabetes). The best mixing weights (β) were calculated
870 using regression on 15,000 European participants from the UK Biobank training set (layer 1).(b)
871 The most effective combination of multi-ancestry, trait-specific PRSs was identified using stepAIC
872 for predicting target trait, and their optimal mixing weights (β) for selected PRSs were determined
873 through regression on the same dataset. The chosen PRSs were then linearly combined to
874 produce O-MetPRS and D-MetPRS (layer 2). The participants in the validation and testing set are
875 non-overlapping individuals and independent with input GWAS samples. (c) These MetPRS were
876 tested on internal (UK Biobank hold out test set) and external (All of Us, MGB Biobank, KFSH&RC)
877 datasets. The genetic predispositions to traits indicative of obesity (i.e., BMI, waist circumference,
878 and visceral adipose tissue) or T2D (i.e., fasting glucose, HbA1c, and insulin sensitivity) can
879 contribute unique insights into the multifaceted nature of metabolic dysfunction. Leveraging
880 genetic signature of related traits might capture a more comprehensive underlying metabolism
881 and enhance the overall prediction of obesity and T2D. BMI: body mass index; DM: type 2
882 diabetes; WC: waist circumference; HC: hip circumference; WHR: waist-to-hip ratio; VAT: visceral
883 adipose tissue; ASAT: abdominal subcutaneous adipose tissue; GFAT: gluteofemoral adipose
884 tissue; vatGfatRatio: VAT/GFAT ratio; asatGfatRatio: ASAT/GFAT ratio; vatAsatRatio: VAT/ASAT
885 ratio; 2hG: 2-hour glucose; ISI: insulin sensitivity index; FG: fasting glucose; FI: fasting insulin;
886 cT1: MRI-based Liver iron corrected T1; NAFLD: non-alcoholic fatty liver disease.

887
888
889
890
891
892
893
894
895
896
897



898

899

Fig.2 | Genetic correlations across 22 metabolic traits and measures associated with insulin resistance.

900

901

Genetic correlations were obtained from cross-trait LD-score regression using sex-combined summary statistics. Stumvoll insulin sensitivity index (ISI) is inversely proportional to insulin resistance as it measures insulin sensitivity. Genetic correlation heatmap matrix remained significant after multiple testing corrections are indicated by asterisks. BMI: body mass index; DM: type 2 diabetes; WC: waist circumference; HC: hip circumference; WHR: waist-to-hip ratio; ASAT: abdominal subcutaneous adipose tissue; VAT: visceral adipose tissue; GFAT: gluteofemoral adipose tissue; asatGfatRatio: ASAT/GFAT ratio; vatAsatRatio: VAT/ASAT ratio; vatGfatRatio: VAT/GFAT ratio; 2hG: 2-hour glucose; FG: fasting glucose; FI: fasting insulin; cT1: MRI-based Liver iron corrected T1; ISI: insulin sensitivity index; NAFLD: non-alcoholic fatty liver disease.

902

903

904

905

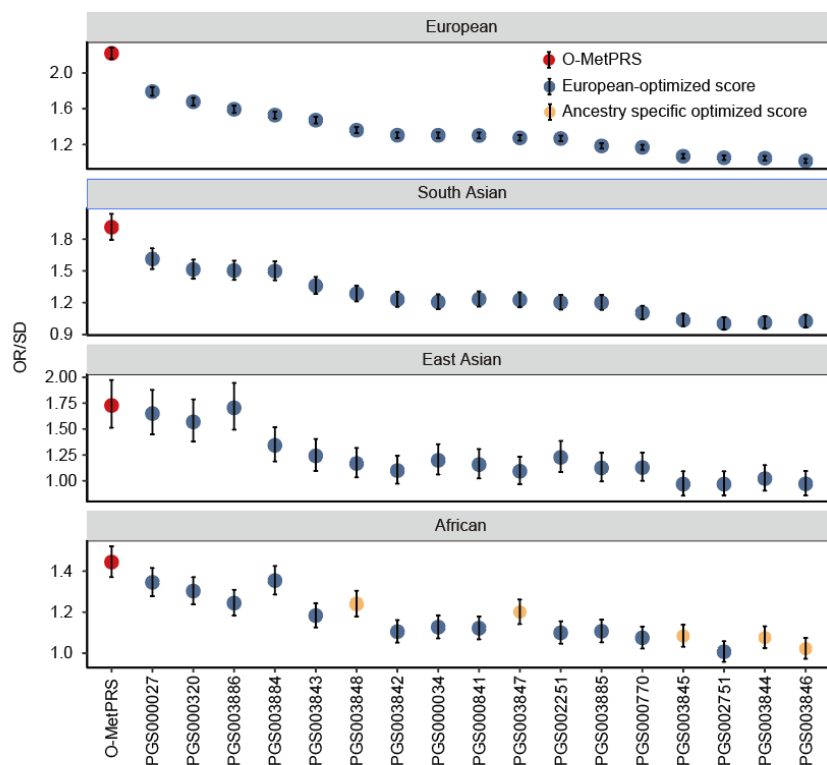
906

907

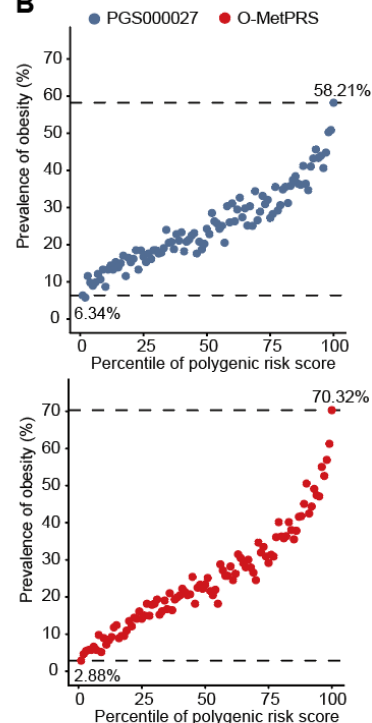
908

909

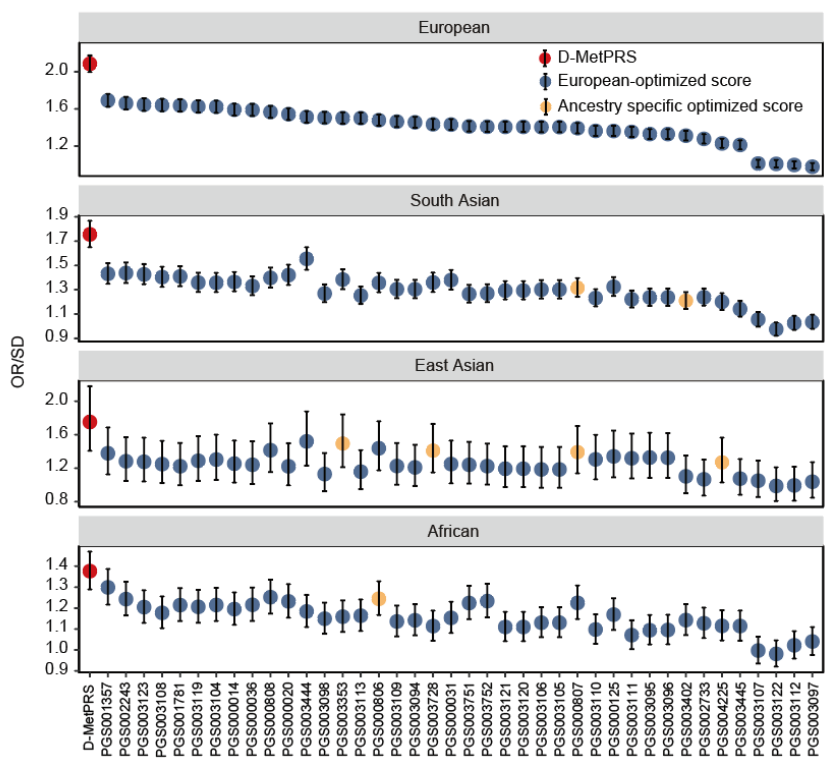
A



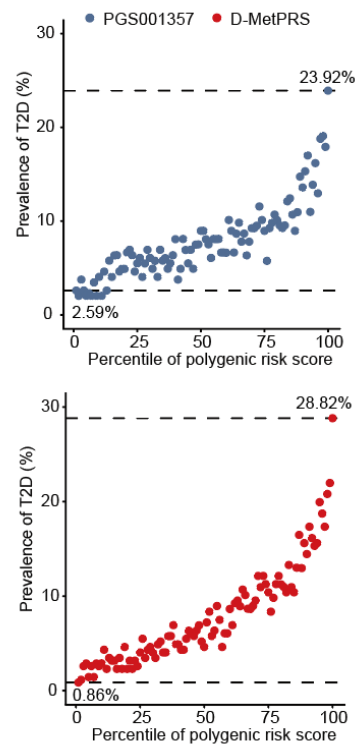
B



C



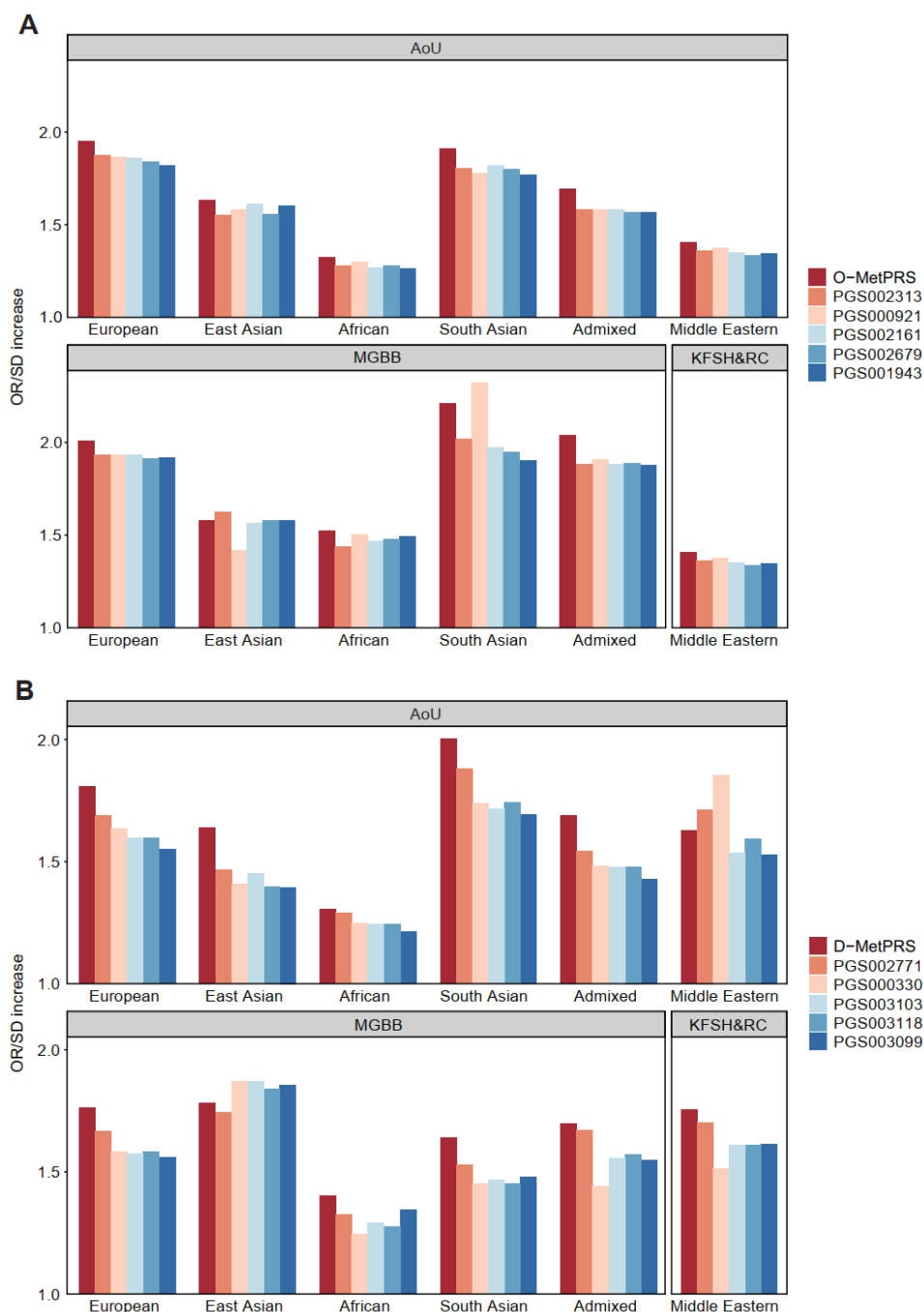
D



911 **Fig.3 | Internal validation of MetPRS performance in UK Biobank hold-out test set**

912 The OR/SD with 95% CI and prevalence (%) across percentile bins were assessed for obesity
913 **(A,B)** and T2D **(C,D)** in UK Biobank hold out test set (n = 49,377). MetPRS was compared with
914 all obesity and T2D PRSs published in PGS Catalog (<https://www.pgscatalog.org/>) that had not
915 included UK Biobank in score development. We then selected highest-performing score among
916 published studies for obesity (PGS000027 - GPS_BMI - Khera AV et al) and T2D (PGS001357 -
917 T2D_AnnoPred_PRS - Ye et al), and compared them with MetPRS by grouping according to
918 polygenic risk score percentiles. Obesity was defined as a BMI of 25 Kg/m² or above for East
919 Asians, and 30 Kg/m² or above for all other populations. O-MetPRS: obesity MetPRS; D-MetPRS:
920 T2D MetPRS; OR/SD: odds ratio per increase in 1 standard deviation.

921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953



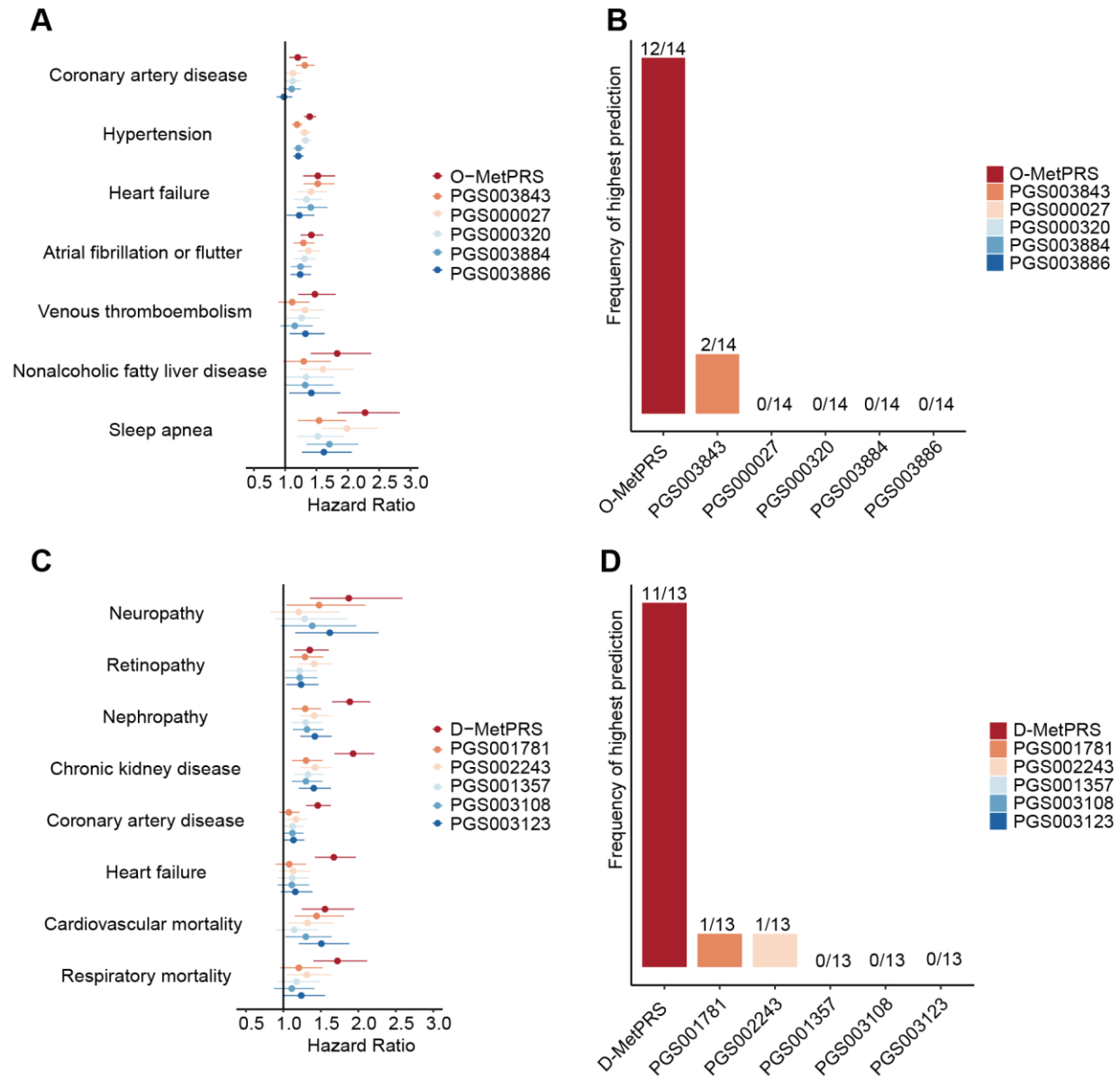
954
955

956 **Fig.4 | External validation of MetPRS in multiple ancestries.**

957 The OR/SD for prevalent obesity **(A)** and T2D **(B)** was assessed for each polygenic risk score in
 958 a logistic regression model adjusted for age, sex, genotyping array and the first ten principal
 959 components of ancestry in external datasets: All of Us (n = 245,394), MGB Biobank (n = 53,306),
 960 and KFSH&RC cohort (n = 6,416). MetPRS was compared with top 5 obesity and T2D scores in
 961 PGS Catalog. Polygenic risk scores that consistently ranked highly across external datasets were
 962 selected as comparators (top 5), excluding any scores that incorporated any test cohort (i.e., All

963 of Us) in their score development. Obesity was defined as a BMI of 25 Kg/m² or above for East
964 Asians, and 30 Kg/m² or above for all other populations. AoU: All of Us; MGBB: Massachusetts
965 General Brigham Biobank; Admixed: Latino/admixed American; O-MetPRS: obesity MetPRS; D-
966 MetPRS: T2D MetPRS; OR/SD: odds ratio per increase in 1 standard deviation of PRS score.

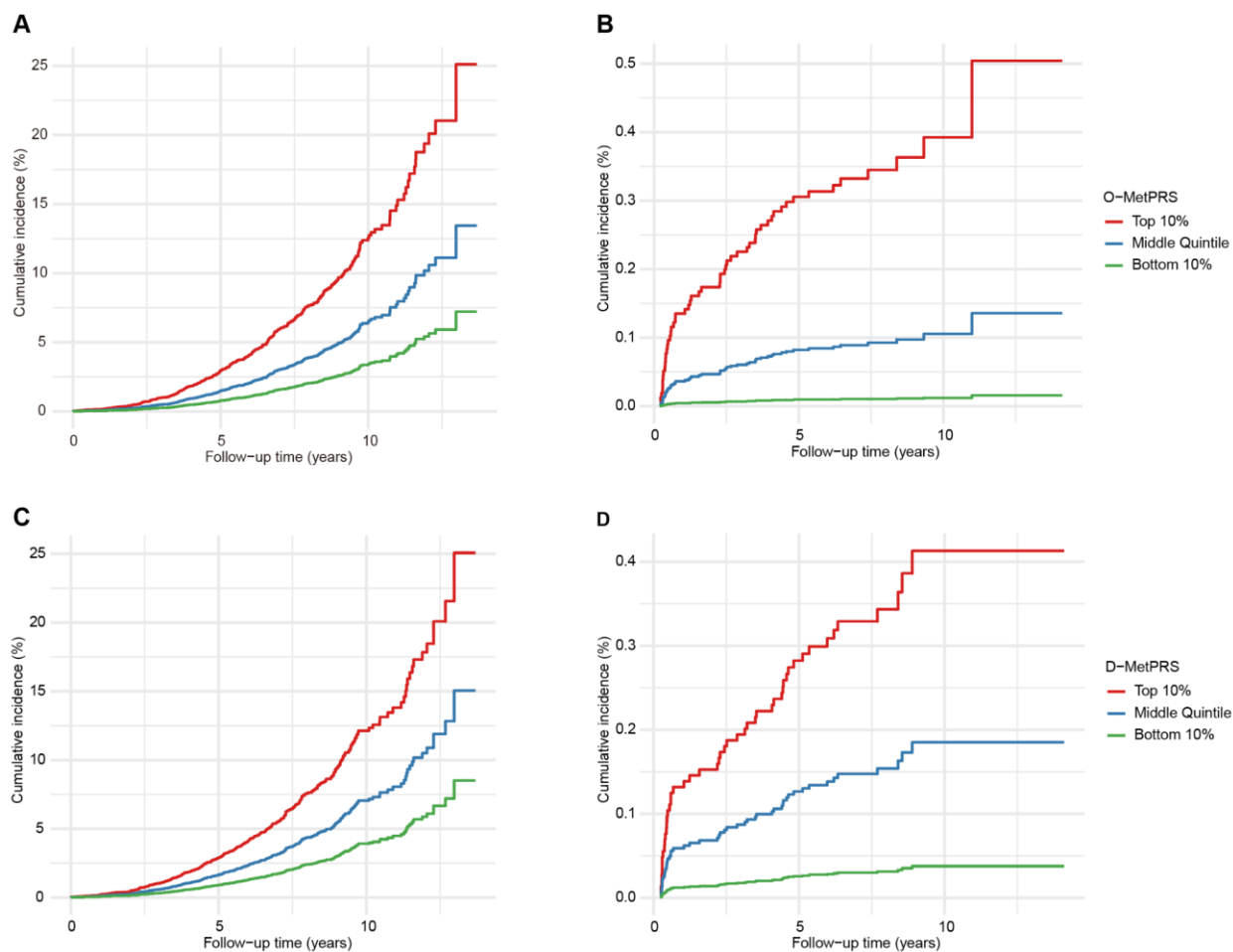
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004



1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015

Fig.5 | Prediction of obesity and diabetes-related morbidities.

The top decile of the MetPRS distribution was identified as "carriers," and their risk was assessed against the rest of the cohort. Risk was estimated using a Cox proportional hazards regression model, adjusted for age, sex, genotyping array and the first ten principal components of ancestry. The performance of O-MetPRS and the other top scores in predicting obesity-related morbidities was evaluated (A), and their corresponding rank frequencies across morbidities were summarized (B). The performance of D-MetPRS and other top scores in predicting diabetes-related morbidities and complications was evaluated (C), and their corresponding rank frequencies across morbidities and complications were summarized (D). The forest plot for the other conditions refers to Extended Data Fig.5. Full list of morbidities is provided in Supplementary Table 14.



1016
1017

1018 **Fig.6 | Risk for incident GLP-1 receptor agonist prescription and metabolic-bariatric**
1019 **surgery according to MetPRS strata in MGB Biobank participants free of obesity and T2D**
1020 **at the baseline**

1021 The 37,329 participants without baseline obesity and T2D were included for analysis. Cumulative
1022 incidence of GLP-1 receptor agonist prescription (A) and metabolic-bariatric surgery (B)
1023 stratified by O-MetPRS; cumulative incidence of GLP-1 receptor agonist prescription (C)
1024 and metabolic-bariatric surgery (D) stratified by D-MetPRS. The risks were estimated using Cox proportional
1025 hazards regression model adjusted for age, sex, and the first ten principal components of ancestry
1026 in a 37,329 population without obesity and T2D at the baseline in MGB Biobank. GLP-1 receptor
1027 agonist prescription and metabolic-bariatric surgery were selected as clinical endpoints as they
1028 are interventions indicated for obesity or diabetes or both.

1029