# **AORTA Gene: Polygenic prediction** improves detection of thoracic aortic 2 aneurysm 3

4

1

James P. Pirruccello MD<sup>1,2,3</sup>, Shaan Khurshid MD, MPH<sup>4,5,6,7</sup>, Honghuang Lin PhD<sup>8,9</sup>, 5 Lu-Chen Weng PhD<sup>5,6</sup>, Siavash Zamirpour MS<sup>10</sup>, Shinwan Kany MD<sup>6,11</sup>, Avanthi Raghavan MD<sup>4,6</sup>, Satoshi Koyama MD, PhD<sup>5,6,12</sup>, Ramachandran S. Vasan MD<sup>8,13,14</sup>, Emelia J. Benjamin MD, ScM<sup>8,13,14</sup>, Mark E. Lindsay MD, PhD<sup>4,5,6,7,15</sup>, Patrick T. Ellinor 6 7 8 MD, PhD<sup>4,5,6,7</sup> 9

10

11 1 Division of Cardiology, University of California San Francisco, San Francisco, California, USA, 12 2 Institute for Human Genetics, University of California San Francisco, San Francisco, 13 California, USA, 3 Bakar Computational Health Sciences Institute, University of California San 14 Francisco, San Francisco, California, USA, 4 Cardiology Division, Massachusetts General 15 Hospital, Boston, Massachusetts, USA, 5 Cardiovascular Research Center, Massachusetts 16 General Hospital, Boston, Massachusetts, USA, 6 Cardiovascular Disease Initiative, Broad 17 Institute of MIT and Harvard, Cambridge, Massachusetts, USA, 7 Harvard Medical School, 18 Boston, Massachusetts, USA, 8 Framingham Heart Study, Boston University and National 19 Heart, Lung, and Blood Institute, Framingham, Massachusetts, USA, 9 Department of Medicine, 20 University of Massachusetts Chan Medical School, Worcester, Massachusetts, USA, 10 School 21 of Medicine, University of California San Francisco, San Francisco, California, USA, 11 22 Department of Cardiology, University Heart and Vascular Center Hamburg-Eppendorf, 23 Hamburg, Germany, 12 Laboratory for Cardiovascular Genomics and Informatics, RIKEN 24 Center for Integrative Medical Sciences, Kanagawa, Japan, 13 Department of Medicine, 25 Cardiology and Preventive Medicine Sections, Boston Medical Center, Boston University 26 Chobanian and Avedisian School of Medicine, Boston, Massachusetts, USA, 14 Epidemiology 27 Department, Boston University School of Public Health, Boston, Massachusetts, USA, 15 28 Thoracic Aortic Center, Massachusetts General Hospital, Boston, Massachusetts, USA 29 30 Running Title: AORTA Gene 31 **Word Count:** 4,053 32 Keywords: Ascending aorta, polygenic score, UK Biobank, All of Us, Framingham

- 33 Heart Study, Mass General Brigham
- 34
- 35 Correspondence:
- 36 James P. Pirruccello, MD
- 37 University of California San Francisco
- 38 Division of Cardiology
- 39 555 Mission Bay Blvd South #3118
- 40 San Francisco, CA 94158
- 41 james.pirruccello@ucsf.edu

Patrick T. Ellinor, MD. PhD Massachusetts General Hospital Cardiology Division 55 Fruit Street Boston, MA 02114 ellinor@mgh.harvard.edu

## 42 Abstract

43 Background: Thoracic aortic disease is an important cause of morbidity and mortality in 44 the US, and aortic diameter is a heritable contributor to risk. Could a polygenic 45 prediction of ascending aortic diameter improve detection of aortic aneurysm? 46 47 **Methods**: Deep learning was used to measure ascending thoracic aortic diameter in 48 49,939 UK Biobank participants. A genome-wide association study (GWAS) was 49 conducted in 39,524 participants and leveraged to build a 1.1 million-variant polygenic 50 score with *PRScs-auto*. Aortic diameter prediction models were built with the polygenic 51 score ("AORTA Gene") and without it. The models were tested in a held-out set of 4,962 52 UK Biobank participants and externally validated in 5,469 participants from Mass 53 General Brigham Biobank (MGB), 1,298 from the Framingham Heart Study (FHS), and 54 610 participants from All of Us. 55 56 **Results:** In each test set, the AORTA Gene model explained more of the variance in 57 thoracic aortic diameter compared to clinical factors alone: 39.9% (95% CI 37.8-42.0%) 58 vs 29.2% (95% CI 27.1-31.4%) in UK Biobank, 36.5% (95% CI 34.4-38.5%) vs 32.5% 59 (95% CI 30.4-34.5%) in MGB, 41.8% (95% CI 37.7-45.9%) vs 33.0% (95% CI 28.9-60 37.2%) in FHS, and 34.9% (95% CI 28.8-41.0%) vs 28.9% (95% CI 22.9-35.0%) in All 61 of Us. AORTA Gene had a greater AUROC for identifying diameter  $\geq$ 4cm in each test 62 set: 0.834 vs 0.765 (P=7.3E-10) in UK Biobank, 0.808 vs 0.767 in MGB (P=4.5E-12), 63 0.856 vs 0.818 in FHS (P=8.5E-05), and 0.827 vs 0.791 (P=7.8E-03) in All of Us. 64

2/35

#### AORTA Gene

- 65 **Conclusions**: Genetic information improved estimation of thoracic aortic diameter when
- 66 added to clinical risk factors. Larger and more diverse cohorts will be needed to develop
- 67 more powerful and equitable scores.
- 68
- 69

## 70 Introduction

Thoracic aortic disease is an important cause of morbidity and mortality<sup>1,2</sup>. Ascending
thoracic aortic enlargement is a well-established risk factor for ascending aortic
dissection<sup>3,4</sup>. Indeed, the majority of dissections occur in individuals with aortic diameter
≥4cm, including more than 90% in the International Registry of Aortic Dissection and
100% in Kaiser<sup>4,5</sup>. Contemporary guidelines do not describe a role for population
screening of thoracic aortic aneurysm<sup>6</sup>, and universal imaging would likely be
impractical.

78

79 These limitations have spurred the development of clinical risk scores to identify individuals who are likely to have a rtic enlargement on confirmatory imaging<sup>7–9</sup>. These 80 81 scores have been shown to explain approximately 30% of the variance in ascending 82 aortic diameter, with an area under the receiver operator characteristic curve (AUROC) 83 near 0.78 for detecting individuals with diameter  $\geq$ 4cm. Population genetics studies 84 suggest that ascending aortic diameter is highly heritable, with a proportion of variance attributable to common genetic factors of over 60%<sup>10</sup>. We therefore sought to assess 85 86 whether incorporating polygenic risk might improve estimation of ascending aortic 87 diameter over clinical factors alone.

88

## 89 Methods

### 90 Study design

91 Model development and internal validation were conducted in UK Biobank, with external 92 validation within the *All of Us* cohort, the Mass General Brigham Biobank (MGB), and 93 the Framingham Heart Study (FHS). This study was reported in accordance with the 94 transparent reporting of a multivariable prediction model for individual prognosis or 95 diagnosis (TRIPOD) statement<sup>11</sup>.

96

97 Study protocols complied with the tenets of the Declaration of Helsinki. All UK Biobank participants provided written informed consent<sup>12</sup>, and only those participants who had 98 not withdrawn consent as of March 7, 2023 were analyzed. The UK Biobank analyses 99 100 were considered exempt by the UCSF Institutional Review Board (IRB), #22-37715. UK 101 Biobank analyses were conducted under application #41664. Each All of Us biobank participant provided written informed consent<sup>13</sup>. The All of Us analyses were considered 102 103 exempt by the UCSF IRB, #22-37715. The MGB study protocols were approved with a 104 waiver of informed consent by the Mass General Brigham IRB. All FHS participants 105 provided written informed consent, and the FHS imaging analyses were previously approved by the IRB of the Boston University Medical Center<sup>14</sup>. The FHS analyses were 106 107 also approved by the MGB IRB.

#### 108 Study populations

In UK Biobank participants, ascending aortic diameter was measured from magnetic
 resonance imaging (MRI) using a previously described deep learning model<sup>15</sup>. Model
 5/35 AORTA Gene

111 development and internal validation were conducted in UK Biobank. External validation 112 of the models was pursued in MGB participants with ascending aortic diameter measured from TTE for clinical indications<sup>16</sup>; in FHS participants with ascending aortic 113 114 diameter measured for research purposes from non-contrast computed tomography (CT)<sup>14</sup>: and in *All of Us* biobank participants who had clinical aortic diameter 115 measurements from transthoracic echocardiography (TTE)<sup>13</sup>. In MGB, race or ethnicity 116 was extracted from the electronic health record<sup>16</sup> and the validation analyses were 117 118 repeated in individuals identified as Black.

### 119 Modality-specific differences

The clinical approach to measuring ascending aortic diameter differs by modality. UK Biobank MRI-based aortic diameter measurements were derived from the blood pool diameter only, incorporating neither aortic wall. For TTE, used in the MGB and *All of Us* cohorts, standard measurements use the "leading edge to leading edge" approach, incorporating one wall. For non-contrast CT measurements, used in FHS, both walls are incorporated into the diameter.

### 126 Splitting the UK Biobank for training and validation

49,939 UK Biobank participants with MRI measurements of aortic diameter were
randomly assigned a number from 0-999. Those from 0-799 contributed to the genomewide association study (GWAS) used to produce the polygenic score (N=39,524). Those
from 800-999 related within 3 degrees of kinship to the 0-799 group were excluded to
avoid polygenic score overfitting (N=557). Those remaining from 0-899 contributed to
the development of non-genetic models (N=44,420). Those from 800-899 were included

6/35

- in the derivation of weights for the polygenic score component of the AORTA Gene
- 134 combined clinical and genetic model (N=4,896). Those ≥900 were analyzed for
- 135 validation (N=4,962; **Figure 1**).

#### 136 Thoracic aortic aneurysm in All of Us

- 137 Within All of Us, thoracic aortic aneurysm was defined as the first occurrence of
- 138 SNOMED codes 433068007, 74883004, or 426948001; ICD10CM codes of I71.01,
- 139 I71.1, I71.2, or I77.810; ICD9CM codes of 441.01, 441.1, 441.2, or 447.71; ICD10PCS
- 140 codes of 02RX0JZ or 02RW0JZ; the ICD9Proc code 39.73; or CPT4 codes 33863,
- 141 75957, 33880, 75956, or 33881. Prevalent analyses were conducted based on the
- 142 presence of disease labels prior to the time of enrollment. Time-to-event analyses were
- 143 conducted beginning at the time of enrollment and censored on July 1, 2022 using the R
- 144 *survival* package, excluding participants whose diagnoses occurred prior to enrollment.
- 145 Any tabular data with an All of Us sample count <20 was modified in accordance with All
- 146 of Us policies.

#### 147 Genome-wide association study and polygenic scores

- 148 Common genetic contributions to ascending aortic diameter in the 39,524 GWAS
- 149 participants were discovered using REGENIE v3.2.7<sup>17</sup> with imputed variants provided by
- 150 UK Biobank<sup>18,19</sup>. GWAS covariates included age, age<sup>2</sup>, sex, the MRI serial number, the
- 151 genotyping array, and the first ten principal components of ancestry. The summary
- 152 statistics were trimmed to the ~1.1 million UK Biobank-specific variants within HapMap3
- 153 identified by the *PRScs* authors, and underwent Bayesian weighting by *PRScs-auto* on
- 154 its default settings to produce polygenic score weights<sup>20</sup>.

155	
-----	--

- 156 The polygenic score was applied to all UK Biobank participants with imputed genetic
- 157 data<sup>19</sup>. Variants with imputation quality <0.3 were removed.

158

- 159 The polygenic score was applied to all MGB participants with imputed genetic data<sup>16,21</sup>
- 160 with imputation into the TOPMed panel<sup>22</sup>.

161

- 162 The polygenic score was applied to FHS participants with CT and genetic data
- 163 available. For FHS, genotyping was done on the Affymetrix GeneChip 500K Array Set &

164 50K Human Gene Focused Panel (Affymetrix, Santa Clara, CA, USA). Variants with call

rate <97% or Hardy–Weinberg equilibrium P<1E-06 were excluded. The remaining

- 166 variants were then imputed to the 1000 Genomes Project phase I release 3 panel by
- 167 MACH version  $1.0^{23,24}$ . Variants with imputation quality <0.3 were removed.

168

169 The polygenic score was applied to All of Us participants with whole genome

170 sequencing data available in release "R7"<sup>25</sup>.

171

Within each set of participants, the polygenic score was residualized for the first 20 principal components of ancestry (except for FHS, where 10 were used due to availability). The polygenic score in each group was zero-centered and scaled by the standard deviation from the UK Biobank training set: for studies that applied the average polygenic score (UK Biobank, *All of Us*) the standard deviation was 3.4E-07,

and for those that applied the summed polygenic score (MGB, FHS) the standarddeviation was 0.38.

### 179 Ascending aortic diameter model training

180 Hierarchical group least absolute shrinkage and selection operator models were built to estimate ascending aortic diameter using the R package *alinternet*<sup>26</sup>. Three models 181 182 were constructed with *glinternet* v1.0.11 using the following independent variables: (a) age,  $age^2$ , and sex; (b) age,  $age^2$ , sex, and polygenic score; (c) age,  $age^2$ , body mass 183 184 index (BMI), heart rate, systolic and diastolic blood pressure, height, weight, sex, and a 185 history of diabetes, hypertension, or hyperlipidemia. A fourth model (d) was constructed 186 by taking a linear combination of the model (c) and the polygenic score. The weighting 187 for model (c) was estimated in all participants randomly assigned to 0-899, while the 188 weighting for model (b) and the linear combination for model (d) were estimated in those 189 assigned to 800-899 to avoid overlap with the GWAS cohort. For ease of reference, we 190 label model (c) "AORTA Score" (AORTA: optimized regression for thoracic aneurysm) 191 and model (d) "AORTA Gene".

### 192 Statistical analysis

All statistical significance tests were two-sided, with significance defined as P < 0.05,</li>
and performed in R 4.2.3 unless otherwise stated. For visualization, scores were plotted
against measured thoracic aortic diameter.

#### 196 Outcomes

197	The primary outcome was correlation between the scores and ascending aortic
198	diameter using linear models, expressed as R <sup>2</sup> (variance explained). Secondary
199	outcomes included tests of calibration, and the performance of the scores for identifying
200	ascending aortic diameter ≥4.0cm; the clinical diagnosis of aneurysm was also tested in
201	All of Us.
202	
203	To assess calibration, each score was used as an independent variable in a linear
204	model predicting aortic diameter that also had an intercept, permitting significance
205	testing of both the slope and the intercept. Mean absolute error was also computed.
206	
207	To assess the performance for identifying diameter ≥4cm, the AUROC was computed.
208	Confusion matrices and their derived statistical measures were produced based on the
209	presence or absence of aortic diameter ≥4cm and the presence or absence of predicted
210	diameters above a fixed threshold determined in the training set for each score. The
211	thresholds for the confusion matrix analysis were defined in the 4,896 participants in the
212	training set.
213	
214	AUROCs were compared using the DeLong test <sup>27</sup> . The continuous net reclassification

215 index (NRI) was calculated following the formula of Pencina, *et al*<sup>28</sup>.

## 216 Data and code availability

217 Computed scores are returned to UK Biobank, where data are made available by UK

- 218 Biobank to researchers from research institutions with genuine research inquiries,
  - 10/35

#### **AORTA Gene**

- 219 following IRB and UK Biobank approval. FHS data are made available to researchers
- 220 with approved research applications. The dbGAP study accession number used for FHS
- validation was #phv00076329.v1.p5 for ascending aortic diameter. MGB data are
- available to MGB investigators; external collaboration requests can be initiated through
- 223 https://biobank.massgeneralbrigham.org/for-researchers. The complete set of AORTA
- 224 Score covariates and their weights are available as R programs at
- github.com/carbocation/genomisc and the polygenic score weights will be available on
- the Polygenic Score Catalog (pgscatalog.org).

## 227 Results

### 228 Clinical characteristics

229 Among the 4,896 UK Biobank participants whose data contributed to training the 230 genetic models, 2,489 were women and 2,407 were men. The ascending aortic 231 diameter was 3.05 ± 0.31cm for women and 3.34 ± 0.34cm for men. The 4,962 internal 232 validation set participants had similar measurements (Table 1; Figure 1). The MGB 233 validation set consisted of 5,469 participants with genetic data and aortic diameter 234 measured from TTE. The FHS validation set consisted of 1,298 participants with genetic 235 data and aortic diameter measured from non-contrast CT. The All of Us validation set 236 consisted of 610 participants 40 years and older with genetic data and aortic diameter 237 measured from TTE; 60% identified as female and 88% as having non-Hispanic white 238 ethnicity.

### 239 Model validation for aortic diameter

Testing calibration, the slope of the AORTA Gene model was statistically consistent with
one in UK Biobank (P=0.2), MGB (P=1.0), FHS (P=0.6), and *All of Us* (P=0.4) (Figure
2). The intercept was consistent with zero in UK Biobank (P=0.2), MGB (P=0.4), and *All of Us* (P=0.8), but not FHS (intercept estimate 0.26cm, P=0.01). The same pattern was
observed for the clinical AORTA Score. In contrast, the model incorporating only age,
sex, and the polygenic score had a calibration slope inconsistent with one in most
cohorts outside of UK Biobank (eTable AA).

247

248	The AORTA Gene model's mean absolute error (MAE) for ascending aortic diameter
249	was 0.212cm (95% CI 0.207-0.217cm) in UK Biobank, 0.282cm (95% CI 0.275-
250	0.289cm) in MGB, 0.344cm (95% CI 0.331-0.358cm) in FHS, and 0.293cm (95% CI
251	0.273-0.313cm) in All of Us. The clinical AORTA Score had an inferior MAE in all
252	cohorts: 0.232cm (95% CI 0.227-0.237cm) in UK Biobank, 0.291cm (95% CI 0.284-
253	0.298cm) in MGB, 0.351cm (95% CI 0.337-0.365cm) in FHS, and 0.306cm (95% CI
254	0.285-0.327cm) in All of Us. The age/sex/genetics model had an MAE of 0.226cm (95%
255	CI 0.221-0.231cm) in UK Biobank, 0.293cm (95% CI 0.286-0.300cm) in MGB, 0.346cm
256	(95% CI 0.332-0.360cm) in FHS, and 0.318cm (95% CI 0.296-0.340) in All of Us.
257	
258	The AORTA Gene model explained 39.9% of the variance in ascending aortic diameter
259	(99% CI 37.8-42.0%; P=2.6E-551) in the UK Biobank validation set, 36.5% (95% CI
259 260	(99% CI 37.8-42.0%; P=2.6E-551) in the UK Biobank validation set, 36.5% (95% CI 34.4-38.5%, P=2.4E-541) in MGB, 41.8% (95% CI 37.7-45.9%; P=3.5E-154) in FHS,
259 260 261	(99% CI 37.8-42.0%; P=2.6E-551) in the UK Biobank validation set, 36.5% (95% CI 34.4-38.5%, P=2.4E-541) in MGB, 41.8% (95% CI 37.7-45.9%; P=3.5E-154) in FHS, and 34.9% (95% CI 28.8-41.0%; P=1E-58) in <i>All of Us</i> ( <b>Figure 3</b> ). In comparison, the
259 260 261 262	(99% CI 37.8-42.0%; P=2.6E-551) in the UK Biobank validation set, 36.5% (95% CI 34.4-38.5%, P=2.4E-541) in MGB, 41.8% (95% CI 37.7-45.9%; P=3.5E-154) in FHS, and 34.9% (95% CI 28.8-41.0%; P=1E-58) in <i>All of Us</i> ( <b>Figure 3</b> ). In comparison, the clinical AORTA Score explained 29.3% of the variance in UK Biobank (99% CI 27.1-
259 260 261 262 263	(99% CI 37.8-42.0%; P=2.6E-551) in the UK Biobank validation set, 36.5% (95% CI 34.4-38.5%, P=2.4E-541) in MGB, 41.8% (95% CI 37.7-45.9%; P=3.5E-154) in FHS, and 34.9% (95% CI 28.8-41.0%; P=1E-58) in <i>All of Us</i> ( <b>Figure 3</b> ). In comparison, the clinical AORTA Score explained 29.3% of the variance in UK Biobank (99% CI 27.1-31.4%), 32.5% in MGB (95% CI 30.4-34.5%), 33.0% (95% CI 28.9-37.2%) in FHS, and
259 260 261 262 263 264	(99% CI 37.8-42.0%; P=2.6E-551) in the UK Biobank validation set, 36.5% (95% CI 34.4-38.5%, P=2.4E-541) in MGB, 41.8% (95% CI 37.7-45.9%; P=3.5E-154) in FHS, and 34.9% (95% CI 28.8-41.0%; P=1E-58) in <i>All of Us</i> ( <b>Figure 3</b> ). In comparison, the clinical AORTA Score explained 29.3% of the variance in UK Biobank (99% CI 27.1-31.4%), 32.5% in MGB (95% CI 30.4-34.5%), 33.0% (95% CI 28.9-37.2%) in FHS, and 28.9% (95% CI 22.9-35.0%) in <i>All of Us</i> . The age/sex/genetics model explained 31.8%
259 260 261 262 263 264 265	(99% CI 37.8-42.0%; P=2.6E-551) in the UK Biobank validation set, 36.5% (95% CI 34.4-38.5%, P=2.4E-541) in MGB, 41.8% (95% CI 37.7-45.9%; P=3.5E-154) in FHS, and 34.9% (95% CI 28.8-41.0%; P=1E-58) in <i>All of Us</i> ( <b>Figure 3</b> ). In comparison, the clinical AORTA Score explained 29.3% of the variance in UK Biobank (99% CI 27.1-31.4%), 32.5% in MGB (95% CI 30.4-34.5%), 33.0% (95% CI 28.9-37.2%) in FHS, and 28.9% (95% CI 22.9-35.0%) in <i>All of Us</i> . The age/sex/genetics model explained 31.8% (95% CI 29.6-33.9%) of variance in UK Biobank, 33.2% (95% CI 31.2-35.3%) in MGB,
259 260 261 262 263 264 265 266	(99% CI 37.8-42.0%; P=2.6E-551) in the UK Biobank validation set, 36.5% (95% CI 34.4-38.5%, P=2.4E-541) in MGB, 41.8% (95% CI 37.7-45.9%; P=3.5E-154) in FHS, and 34.9% (95% CI 28.8-41.0%; P=1E-58) in <i>All of Us</i> ( <b>Figure 3</b> ). In comparison, the clinical AORTA Score explained 29.3% of the variance in UK Biobank (99% CI 27.1- 31.4%), 32.5% in MGB (95% CI 30.4-34.5%), 33.0% (95% CI 28.9-37.2%) in FHS, and 28.9% (95% CI 22.9-35.0%) in <i>All of Us</i> . The age/sex/genetics model explained 31.8% (95% CI 29.6-33.9%) of variance in UK Biobank, 33.2% (95% CI 31.2-35.3%) in MGB, 33.3% (95% CI 29.1-37.5%) in FHS, and 31.0% (95% CI 24.9-37.1%) in <i>All of Us</i> .

### 268 Model validation for a 4cm aortic diameter threshold

- 269 113 of the 4,962 internal validation set participants in UK Biobank had aortic diameter
- 270 ≥4cm (2.3%), as did 335 of 5,469 MGB participants (6.1%), 116 of 1,298 FHS
- participants (8.9%), and 65 of 610 *All of Us* participants (10.7%).
- 272
- The AUROC for detecting diameter ≥4cm for the AORTA Gene model was 0.834 (95%
- 274 CI 0.798-0.869) in UK Biobank, 0.808 (95% CI 0.783-0.832) in MGB, 0.856 (95% CI
- 275 0.821-0.891) in FHS, and 0.827 (95% CI 0.776-0.878) in *All of Us* (**Figure 4**; **eTable**
- 276 **BB**).
- 277
- 278 For the clinical AORTA Score, the AUROC was 0.765 (95% CI 0.723-0.807) in UK
- 279 Biobank, 0.767 (95% CI 0.741-0.792) in MGB, 0.818 (95% CI 0.776-0.859) in FHS, and
- 280 0.791 (95% CI 0.735-0.846) for All of Us. The AUROC of AORTA Gene was
- significantly greater in all cohorts: P=7.3E-10 against a null hypothesis of no difference
- 282 in UK Biobank, P=4.5E-12 in MGB, P=8.5E-05 in FHS, and P=7.8E-03 in *All of Us*
- 283 (eTable CC).
- 284

The age/sex/genetics model had an AUROC of 0.805 (95% CI 0.767-0.843) in UK
Biobank, 0.805 (95% CI 0.781-0.829) in MGB, 0.819 (95% CI 0.781-0.856) in FHS, and
0.813 (95% CI 0.757-0.870) in *All of Us*. The AORTA Gene AUROC was significantly
greater in UK Biobank (P=0.03) and FHS (P=3.9E-03), but not in MGB (P=0.7) or *All of Us* (P=0.5) (**eTable CC**).

290

The continuous NRI was positive for AORTA Gene over the clinical AORTA Score and the age/sex/genetics model in all cohorts (**eTable CC**). Decomposed into its NRI(event) and NRI(nonevent) components, AORTA Gene had a greater NRI(event) in all cohorts compared to the AORTA Score and the age/sex/genetics model. However, its NRI(nonevent) value was lower than that of the AORTA Score in MGB and UK Biobank, and lower than that of the age/sex/genetics model in MGB and *All of Us*.

#### 297 Score threshold demonstrations

To evaluate the consequences of various score thresholds whereby a portion of the
population could be brought forward for confirmatory thoracic imaging, two thresholds
were defined within the UK Biobank training set and applied in the validation sets: one
targeted at the top 10% of the population, and the other designed to have a sensitivity of
90% for aortic diameter ≥4cm (eTable BB).

303

For the target top 10% AORTA Gene cutoff, 485 (9.8%) of 4,962 participants were

above this threshold in UK Biobank, 640 (11.7%) of 5,469 in MGB, 99 (7.6%) of 1,295 in

306 FHS, and 97 (15.9%) of 610 in *All of Us*. 59 of 113 UK Biobank participants with

diameter ≥4cm were correctly classified as enlarged (sensitivity 52.2%), as were 162 of

308 335 in MGB (48.4%), 47 of 116 in FHS (40.5%), and 35 of 65 in *All of Us* (53.8%). The

precision was 0.122 in UK Biobank, 0.253 in MGB, 0.475 in FHS, and 0.361 in All of Us,

310 yielding respective F1 scores of 0.197, 0.332, 0.437, and 0.432 (**eTable BB**). A

311 comparable threshold for the clinical AORTA Score flagged 484 (9.8%) of 4,962 UK

312 Biobank participants, 722 (13.2%) of 5,469 in MGB, 92 (7.1%) of 1,295 in FHS, and 110

313 (18.0%) of 610 in All of Us. 41 of 113 UK Biobank participants with diameter ≥4cm were

correctly classified as enlarged (sensitivity 36.3%) as were 143 of 335 in MGB (42.7%),
38 of 116 in FHS (32.8%), and 33 of 65 in *All of Us* (50.8%). The precision was 0.85 in
UK Biobank, 0.198 in MGB, 0.413 in FHS, and 0.300 in *All of Us*, yielding respective F1
scores of 0.137, 0.271, 0.365, and 0.377 (eTable BB).

318

319 For the target 90% sensitive AORTA Gene cutoff, 1,737 (35.0%) of 4,962 UK Biobank 320 participants were above this threshold, 1,943 (35.5%) of 5,469 in MGB, 336 (25.9%) of 321 1,295 in FHS, and 247 (40.5%) of 610 in All of Us. 92 of 113 UK Biobank participants 322 with diameter  $\geq$ 4cm were correctly classified as enlarged (sensitivity 81.4%), as were 323 258 of 335 in MGB (77.0%), 88 of 116 in FHS (75.9%), and 56 of 65 in All of Us 324 (86.2%). The precision was 0.053 in UK Biobank, 0.133 in MGB, 0.262 in FHS, and 325 0.227 in All of Us, yielding respective F1 scores of 0.099, 0.227, 0.389, and 0.359 326 (eTable DD). A comparable threshold for the clinical AORTA Score flagged 2,006 327 (40.4%) of 4,962 UK Biobank participants, 2,194 (40.1%) of 5,469 in MGB, 405 (31.3%) 328 of 1,295 in FHS, and 266 (43.6%) of 610 in All of Us. 87 of 113 UK Biobank participants 329 with diameter  $\geq$ 4cm were correctly classified as enlarged (sensitivity 77.0%), as were 330 264 of 335 in MGB (78.8%), 86 of 116 in FHS (74.1%), and 52 of 65 in All of Us 331 (80.0%). The precision was 0.043 in UK Biobank, 0.120 in MGB, 0.212 in FHS, and 332 0.195 in All of Us, yielding respective F1 scores of 0.082, 0.209, 0.330, and 0.314 333 (eTable DD).

### 334 **Thoracic aortic aneurysm diagnosis in** *All of Us*

In the 164,789 *All of Us* participants aged 40 or older with genetic data, 1,904 had an

electronic health record (EHR)-based diagnosis of thoracic aortic aneurysm prior to

- enrollment. AORTA Gene's AUROC was 0.760 (95% CI 0.750-0.771) compared to
- 338 0.739 (95% CI 0.728-0.750) for the AORTA Score (eTable EE), P=2.4E-16 against a
- null hypothesis of no difference (**eTable FF**). For incident disease (1,632 cases
- diagnosed after enrollment), AORTA Gene's AUROC was 0.748 (95% CI 0.737-0.759)
- 341 compared to 0.729 (95% CI 0.718-0.741) for the AORTA Score, P=9.5E-10 against a
- 342 null hypothesis of no difference (**eTables EE-FF**).

### 343 Analysis in Black individuals in MGB

- In MGB, 340 individuals were identified as Black through the electronic health record<sup>16</sup>.
- 345 In these individuals, AORTA Gene had an intercept consistent with zero (P=0.5) and a
- 346 slope consistent with one (P=0.5). The MAE was 0.297cm (95% CI 0.272-0.321cm),
- and the model explained 34.8% of variance in ascending aortic diameter (95% CI 26.7-
- 43.0%). In comparison, for the clinical AORTA Score these values were 0.294cm (95%
- 349 CI 0.269-0.319cm) and 34.3% (95% CI 26.2-42.5%) (**eTable AA**). The AUROC for
- detecting diameter ≥4cm for the AORTA Gene model was 0.858 (eTable BB), nominally
- 351 greater than that of the clinical AORTA Score (AUROC 0.820, P=0.04 for the difference,
- **eTable CC**), with a positive continuous NRI (0.473).
- 353

## 354 Discussion

Ascending thoracic aortic diameter has a SNP heritability estimated to be greater than 355 356 60%<sup>10</sup>, which makes the incorporation of common-variant genetic data for the 357 presymptomatic detection of individuals with aortic enlargement or aneurysm 358 conceptually appealing. But whether aortic polygenic scores encode information not 359 already latently captured through correlation with clinical covariates was unknown. 360 Here, we observed that AORTA Gene, which incorporated both an aortic polygenic 361 score and clinical covariates, improved estimation of ascending aortic diameter and 362 identification of individuals with diameter  $\geq$ 4cm in four cohorts—UK Biobank, MGB. 363 FHS, and All of Us—when compared to a previously validated model for aortic diameter 364 built from the same clinical covariates.

365

366 For the estimation of aortic diameter as a quantitative measure, a clinical model built from the same covariates as the AORTA Score<sup>9</sup> explained 29.3% of the variance in 367 368 ascending aortic diameter in UK Biobank, 28.9% in All of Us, 32.5% in MGB, and 33.0% 369 in FHS. The AORTA Gene model that additionally incorporated the aortic diameter 370 polygenic score increased those values to 39.9% in UK Biobank, 34.9% in All of Us, 371 36.5% in MGB, and 41.8% in FHS. These represent a 10.6% absolute improvement in 372 explained variance (36% relative improvement) in UK Biobank, 4% (12.3% relative 373 improvement) in MGB, 8.8% (27% relative improvement) in FHS, and 6% (21% relative 374 improvement) in All of Us. These substantive improvements from the polygenic score 375 over a clinical model are comparable in magnitude to the addition of key clinical predictors over age and sex alone<sup>9</sup>. 376

377

378 Incorporation of the polygenic score also improved identification of individuals with 379 clinically relevant aortic enlargement (i.e., diameter  $\geq 4$  cm). The clinical model had an 380 AUROC of 0.765 in UK Biobank, 0.791 in All of Us, 0.767 in MGB, and 0.818 in FHS; 381 these respectively improved to 0.834, 0.827, 0.808, and 0.856 for AORTA Gene. The 382 AUROC is independent of the choice of score threshold, so we also demonstrated two 383 example thresholds: one based on the goal of being 90% sensitive for diameter  $\geq$ 4cm, 384 and another based on a scenario where there are resources to image up to 10% of the 385 population. In both scenarios, for all four cohorts, sensitivity and specificity were 386 modestly improved with the incorporation of the polygenic score. Interestingly, we also 387 observed that, compared to the comprehensive clinical score, a simpler model 388 consisting only of age, sex, and the polygenic score performed similarly. Because such 389 a model can be estimated for any future time from birth, this may have implications for 390 attempts at early risk stratification before the onset of clinical risk factors. We also noted 391 that, for the task of identifying All of Us participants with an EHR-based diagnosis of 392 aneurysm, all models had attenuated performance—likely in part due to limited 393 ascertainment of thoracic aortic disease in current clinical practice. Nevertheless, even 394 for this task, the AORTA Gene model remained superior to the clinical score (P=9.5E-395 10).

396

397 We also note two limitations of the current model that point to future directions: first,

398 despite a heritability of nearly 60%, the current polygenic score for thoracic aortic

diameter derived from a GWAS of 39,524 participants added approximately 10% to the

400 variance already explained by a clinical model, suggesting that it accounted for 401 approximately 17% of the heritability of aortic diameter. In contrast, polygenic scores for height—built from GWAS of millions of participants—are nearly saturated<sup>29</sup>. Larger 402 403 GWAS sample sizes are required to increase the variance explained by polygenic 404 estimates of ascending aortic diameter. Second, polygenic scores derived in largely 405 European-ancestry populations do not capture the full range of human genetic diversity, limiting their utility in individuals with more diverse genetic identities<sup>30</sup>. To examine this, 406 407 we evaluated 340 individuals in MGB identified as Black in the electronic health record<sup>16</sup>. In these individuals, the addition of the polygenic score did not improve the 408 continuous estimation of aortic diameter compared to the clinical score (R<sup>2</sup> 34.8% for 409 410 AORTA Gene vs 34.3% for the clinical AORTA Score), although it did nominally 411 improve identification of individuals with diameter  $\geq 4$  cm (AUROC 0.858 vs 0.820). 412 P=0.04). We expect a key problem from incorporating genetic data to be one of inequity 413 due to differential accrual of benefits to individuals with European genetic identities. 414 Larger and more diverse GWAS are required to improve genetic predictions for people 415 across diverse genetic ancestries.

416

Anticipation of the relevant clinical scenario for the application of polygenic scores is important for placing the current findings into context. The cost of acquiring a targeted set of images of the ascending thoracic aorta may be comparable that of obtaining genetic information. Therefore, we expect that the value of polygenic scores for primary prevention will be realized within the context of a healthcare system that incorporates genetic information as part of the prior probability across many diseases and risk 423 factors, where the cost of one-time genotyping would be amortized over all downstream 424 uses. In such a system, might there be a role, in principle, for the incorporation of a 425 genetic estimator of thoracic aortic diameter into a comprehensive model to identify 426 people who are likely to have ascending aortic enlargement? The results from the 427 present analysis suggest so.

#### 428 Limitations

429 While all participants were incorporated into analyses without exclusion for race,

430 ethnicity, or ancestry, the participants in UK Biobank, MGB, FHS, and All of Us with

thoracic imaging predominantly had genetic identities similar to that of Europeans. None

432 of the models described in this manuscript are anticipated to be effective for identifying

433 individuals with aortic enlargement driven by rare pathogenic variants, which are not

434 captured in polygenic scores constructed from common variants. The models were

435 derived and tested in individuals over the age of 40 years. The attenuation in variance

436 explained in the external cohorts compared to UK Biobank may be, in part, a

437 consequence of cryptic relatedness causing overfitting to the UK Biobank cohort;

438 reducing this discrepancy is an area of interest for future efforts. Future efforts are also

439 needed to understand whether aortic prediction models can improve outcomes.

#### 440 **Conclusion**

Integrating genetic information into a validated clinical model improved estimation of
ascending aortic diameter and the identification of individuals with thoracic aortic
aneurysm.

## 444 Acknowledgments

445	JPP had full access to all the data in the study and takes responsibility for the integrity
446	of the data and the accuracy of the data analysis. JPP and PTE conceived of the study.
447	JPP conducted the genetic analyses, developed the models, validated the models, and
448	conducted bioinformatic analyses in the UK Biobank, MGB, and All of Us. HL validated
449	the models in FHS. LCW applied the polygenic score in MGB. S Koyama performed
450	PCA in MGB. JPP and PTE wrote the paper. All other authors contributed to the
451	analysis plan or provided critical revisions.

## 452 Sources of funding

453 Organizations that provided financial support had no role in the design and conduct of

454 the study; collection, management, analysis, and interpretation of the data; preparation,

455 review, or approval of the manuscript; or decision to submit the manuscript for

456 publication.

457

458 Dr. Pirruccello is supported by National Institutes of Health (NIH) K08HL159346. Dr. Lin is supported by the NIH grant U01AG068221. Dr. Kany is supported by the Walter-459 460 Benjamin Fellowship from the Deutsche Forschungsgemeinschaft (521832260). Dr. 461 Raghavan is supported by the John S. LaDue Memorial Fellowship in Cardiovascular 462 Medicine from Harvard Medical School. Dr. Benjamin is supported by NIH grants 463 R01HL128914, R01HL092577, R01HL141434, and U54HL120163; and American Heart 464 Association (AHA) grant 18SFRN34110082. Dr. Lindsay is supported by the Fredman 465 Fellowship for Aortic Disease and the Toomey Fund for Aortic Dissection Research and 22/35 AORTA Gene

466	has received salary support from Bayer AG. Dr. Ellinor is supported by grants from the
467	NIH (R01HL092577, 1R01HL157635), from the AHA Strategically Focused Research
468	Networks (18SFRN34110082), and from the European Union (MAESTRIA 965286).
469	
470	The All of Us Research Program is supported by the National Institutes of Health, Office
471	of the Director: Regional Medical Centers: 1 OT2 OD026549; 1 OT2 OD026554; 1 OT2
472	OD026557; 1 OT2 OD026556; 1 OT2 OD026550; 1 OT2 OD 026552; 1 OT2
473	OD026553; 1 OT2 OD026548; 1 OT2 OD026551; 1 OT2 OD026555; IAA #: AOD
474	16037; Federally Qualified Health Centers: HHSN 263201600085U; Data and Research
475	Center: 5 U2C OD023196; Biobank: 1 U24 OD023121; The Participant Center: U24
476	OD023176; Participant Technology Systems Center: 1 U24 OD023163;
477	Communications and Engagement: 3 OT2 OD023205; 3 OT2 OD023206; and
478	Community Partners: 1 OT2 OD025277; 3 OT2 OD025315; 1 OT2 OD025337; 1 OT2
479	OD025276. In addition, the All of Us Research Program would not be possible without
480	the partnership of its participants.
481	
482	From the Framingham Heart Study of the National Heart Lung and Blood Institute of the
483	National Institutes of Health and Boston University School of Medicine, this project has

- 484 been funded in whole or in part with Federal funds from the National Heart Lung and
- 485 Blood Institute, Department of Health and Human Services, under Contract No.

486 75N92019D00031.

## 487 Disclosures

- 488 Dr. Ellinor has received sponsored research support from Bayer AG, IBM Health, Bristol
- 489 Myers Squibb and Pfizer. Dr. Ellinor has also served on advisory boards or consulted
- 490 for Bayer AG, MyoKardia and Novartis. The Broad Institute has filed for a patent on an
- 491 invention from Drs. Ellinor, Lindsay, and Pirruccello related to a previous genetic risk
- 492 predictor for aortic disease. Remaining authors report no disclosures.

## 493 References

- Howard DPJ, Banerjee A, Fairhead JF, Perkins J, Silver LE, Rothwell PM, Oxford
   Vascular Study. Population-based study of incidence and outcome of acute aortic
   dissection and premorbid risk factor control: 10-year results from the Oxford
   Vascular Study. *Circulation*. 2013;127:2031–2037.
- Yamaguchi T, Nakai M, Yano T, Matsuyama M, Yoshino H, Miyamoto Y, Sumita Y, Matsuda H, Inoue Y, Okita Y, Minatoya K, Ueda Y, Ogino H. Population-based incidence and outcomes of acute aortic dissection in Japan. *European Heart Journal Acute Cardiovascular Care*. 2021;10:701–709.
- 502 3. Kim JB, Spotnitz M, Lindsay ME, MacGillivray TE, Isselbacher EM, Sundt TM. Risk
  503 of Aortic Dissection in the Moderately Dilated Ascending Aorta. *J Am Coll Cardiol.*504 2016;68:1209–1219.
- 505 4. Solomon MD, Leong T, Sung SH, Lee C, Allen JG, Huh J, LaPunzina P, Lee H, 506 Mason D, Melikian V, Pellegrini D, Scoville D, Sheikh AY, Mendoza D, Naderi S, 507 Sheridan A, Hu X, Cirimele W, Gisslow A, Leung S, Padilla K, Bloom M, Chung J, 508 Topic A, Vafaei P, Chang R, Miller DC, Liang DH, Go AS, Kaiser Permanente 509 Northern California Center for Thoracic Aortic Disease. Association of Thoracic Aortic Aneurysm Size With Long-term Patient Outcomes: The KP-TAA Study. 510 511 JAMA Cardiology [Internet]. 2022 [cited 2022 Oct 11]; Available from: 512 https://doi.org/10.1001/jamacardio.2022.3305
- 5. Pape LA, Tsai TT, Isselbacher EM, Oh JK, O'gara PT, Evangelista A, Fattori R, Meinhardt G, Trimarchi S, Bossone E, Suzuki T, Cooper JV, Froehlich JB, Nienaber CA, Eagle KA, International Registry of Acute Aortic Dissection (IRAD)
   Investigators. Aortic diameter >or = 5.5 cm is not a good predictor of type A aortic dissection: observations from the International Registry of Acute Aortic Dissection (IRAD). *Circulation*. 2007;116:1120–1127.
- 519 Isselbacher EM, Preventza O, Hamilton Black J, Augoustides JG, Beck AW, Bolen 6. 520 MA, Braverman AC, Bray BE, Brown-Zimmerman MM, Chen EP, Collins TJ, 521 DeAnda A, Fanola CL, Girardi LN, Hicks CW, Hui DS, Schuyler Jones W, Kalahasti 522 V, Kim KM, Milewicz DM, Oderich GS, Ogbechie L, Promes SB, Gyang Ross E, 523 Schermerhorn ML, Singleton Times S, Tseng EE, Wang GJ, Woo YJ. 2022 524 ACC/AHA Guideline for the Diagnosis and Management of Aortic Disease: A 525 Report of the American Heart Association/American College of Cardiology Joint 526 Committee on Clinical Practice Guidelines. Circulation. 2022;146:e334-e482.
- 527 7. Mori M, Gan G, Deng Y, Yousef S, Weininger G, Daggula KR, Agarwal R, Shang
  528 M, Assi R, Geirsson A, Vallabhajosyula P. Development and Validation of a
  529 Predictive Model to Identify Patients With an Ascending Thoracic Aortic Aneurysm.
  530 *Journal of the American Heart Association*. 2021;10:e022102.
- 8. Obel LM, Diederichsen AC, Steffensen FH, Frost L, Lambrechtsen J, Busk M,
   Urbonaviciene G, Egstrup K, Karon M, Rasmussen LM, Gerke O, Bovling AS,
   Lindholt JS. Population-Based Risk Factors for Ascending, Arch, Descending, and
   Abdominal Aortic Dilations for 60-74–Year-Old Individuals. *Journal of the American College of Cardiology*. 2021;78:201–211.
- 536 9. Pirruccello JP, Lin H, Khurshid S, Nekoui M, Weng L-C, Ramachandran VS,
  537 Isselbacher EM, Benjamin EJ, Lubitz SA, Lindsay ME, Ellinor PT. Development of
  538 a Prediction Model for Ascending Aortic Diameter Among Asymptomatic

- 539 Individuals. *JAMA*. 2022;328:1935–1944.
- Pirruccello JP, Chaffin MD, Chou EL, Fleming SJ, Lin H, Nekoui M, Khurshid S,
   Friedman SF, Bick AG, Arduini A, Weng L-C, Choi SH, Akkad A-D, Batra P, Tucker
   NR, Hall AW, Roselli C, Benjamin EJ, Vellarikkal SK, Gupta RM, Stegmann CM,
   Juric D, Stone JR, Vasan RS, Ho JE, Hoffmann U, Lubitz SA, Philippakis AA,
   Lindsay ME, Ellinor PT. Deep learning enables genetic analysis of the human
   thoracic aorta. *Nat Genet.* 2021;
- 546 11. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a
  547 multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD):
  548 The TRIPOD Statement. *Annals of Internal Medicine*. 2015;162:55.
- 549 12. Collins R. UK Biobank Protocol [Internet]. 2007;Available from:
   550 https://www.ukbiobank.ac.uk/media/gnkeyh2q/study-rationale.pdf
- 13. Denny JC, Rutter J, Goldstein DB, Philippakis A, Smoller J, Jenkins G, Dishman E.
  The "All of Us" Research Program. *New England Journal of Medicine*.
  2019;381:668–676.
- 14. Rogers IS, Massaro JM, Truong QA, Mahabadi AA, Kriegel MF, Fox CS,
  Thanassoulis G, Isselbacher EM, Hoffmann U, O'Donnell CJ. Distribution,
  determinants, and normal reference values of thoracic and abdominal aortic
  diameters by computed tomography (from the Framingham Heart Study). *Am J Cardiol.* 2013;111:1510–1516.
- 15. Pirruccello JP, Rämö JT, Choi SH, Chaffin MD, Kany S, Nekoui M, Chou EL,
  Jurgens SJ, Friedman SF, Juric D, Stone JR, Batra P, Ng K, Philippakis AA,
  Lindsay ME, Ellinor PT. The Genetic Determinants of Aortic Distention. *Journal of the American College of Cardiology*. 2023;81:1320–1335.
- 563 16. Khurshid S, Reeder C, Harrington LX, Singh P, Sarma G, Friedman SF, Di Achille
  564 P, Diamant N, Cunningham JW, Turner AC, Lau ES, Haimovich JS, Al-Alusi MA,
  565 Wang X, Klarqvist MDR, Ashburner JM, Diedrich C, Ghadessi M, Mielke J, Eilken
  566 HM, McElhinney A, Derix A, Atlas SJ, Ellinor PT, Philippakis AA, Anderson CD, Ho
  567 JE, Batra P, Lubitz SA. Cohort design and natural language processing to reduce
  568 bias in electronic health records research. *NPJ Digit Med*. 2022;5:47.
- Mbatchou J, Barnard L, Backman J, Marcketta A, Kosmicki JA, Ziyatdinov A,
  Benner C, O'Dushlaine C, Barber M, Boutkov B, Habegger L, Ferreira M, Baras A,
  Reid J, Abecasis G, Maxwell E, Marchini J. Computationally efficient wholegenome regression for quantitative and binary traits. *Nat Genet*. 2021;53:1097–
  1103.
- Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P,
  Green J, Landray M, Liu B, Matthews P, Ong G, Pell J, Silman A, Young A,
  Sprosen T, Peakman T, Collins R. UK biobank: an open access resource for
  identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine*. 2015;12:e1001779.
- Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic
  D, Delaneau O, O'Connell J, Cortes A, Welsh S, Young A, Effingham M, McVean
  G, Leslie S, Allen N, Donnelly P, Marchini J. The UK Biobank resource with deep
  phenotyping and genomic data. *Nature*. 2018;562:203.
- 583 20. Ge T, Chen C-Y, Ni Y, Feng Y-CA, Smoller JW. Polygenic prediction via Bayesian 584 regression and continuous shrinkage priors. *Nature Communications*.

585 2019;10:1776.

- 586 21. Karlson EW, Boutin NT, Hoffnagle AG, Allen NL. Building the Partners HealthCare
  587 Biobank at Partners Personalized Medicine: Informed Consent, Return of Research
  588 Results, Recruitment Lessons and Operational Considerations. *J Pers Med*589 [Internet]. 2016 [cited 2020 Jul 14];6. Available from:
- 590 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4810381/
- 591 22. Kowalski MH, Qian H, Hou Z, Rosen JD, Tapia AL, Shan Y, Jain D, Argos M,
- 592 Arnett DK, Avery C, Barnes KC, Becker LC, Bien SA, Bis JC, Blangero J,
- 593 Boerwinkle E, Bowden DW, Buyske S, Cai J, Cho MH, Choi SH, Choquet H, 594 Cupples LA, Cushman M, Daya M, Vries PS de, Ellinor PT, Faraday N, Fornage M,
- 595 Gabriel S, Ganesh S, Graff M, Gupta N, He J, Heckbert SR, Hidalgo B, Hodonsky
- 596 C, Irvin MR, Johnson AD, Jorgenson E, Kaplan R, Kardia SL, Kelly TN,
- 597 Kooperberg C, Lasky-Su JA, Loos RJF, Lubitz SA, Mathias RA, McHugh CP,
- 598 Montgomery C, Moon J-Y, Morrison AC, Palmer ND, Pankratz N, Papanicolaou
- 599 GJ, Peralta JM, Peyser PA, Rich SS, Rotter JI, Silverman EK, Smith JA, Smith NL,
- Taylor KD, Thornton TA, Tiwari HK, Tracy RP, Wang T, Weiss ST, Weng LC,
  Wiggins KL, Wilson JG, Yanek LR, Zöllner S, North KN, Auer PL, Consortium NTO for PM (TOPMed), Hematology & amp Topm, Group HW, Raffield LM, Reiner AP,
  Li Y. Use of >100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed)
  Consortium whole genome sequences improves imputation quality and detection of
- 605 rare variant associations in admixed African and Hispanic/Latino populations.
   606 *bioRxiv*. 2019;683201.
- 607 23. Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, Peltonen L, 608 Dermitzakis E, Schaffner SF, Yu F, Peltonen L, Dermitzakis E, Bonnen PE, 609 Altshuler DM, Gibbs RA, de Bakker PIW, Deloukas P, Gabriel SB, Gwilliam R, Hunt 610 S, Inouye M, Jia X, Palotie A, Parkin M, Whittaker P, Yu F, Chang K, Hawes A, 611 Lewis LR, Ren Y, Wheeler D, Gibbs RA, Marie Muzny D, Barnes C, Darvishi K, 612 Hurles M, Korn JM, Kristiansson K, Lee C, McCarroll SA, Nemesh J, Dermitzakis 613 E, Keinan A, Montgomery SB, Pollack S, Price AL, Soranzo N, Bonnen PE, Gibbs 614 RA, Gonzaga-Jauregui C, Keinan A, Price AL, Yu F, Anttila V, Brodeur W, Daly MJ, 615 Leslie S, McVean G, Moutsianas L, Nguyen H, Schaffner SF, Zhang Q, Ghori MJR, 616 McGinnis R, McLaren W, Pollack S, Price AL, Schaffner SF, Takeuchi F, 617 Grossman SR, Shlyakhter I, Hostetter EB, Sabeti PC, Adebamowo CA, Foster 618 MW, Gordon DR, Licinio J, Cristina Manca M, Marshall PA, Matsuda I, Ngare D, 619 Ota Wang V, Reddy D, Rotimi CN, Royal CD, Sharp RR, Zeng C, Brooks LD, 620 McEwen JE. The International HapMap 3 Consortium, Principal investigators, 621 Project coordination leaders, Manuscript writing group, Genotyping and QC, 622 ENCODE 3 sequencing and SNP discovery, Copy number variation typing and 623 analysis, Population analysis, Low frequency variation analysis, Linkage 624 disequilibrium and haplotype sharing analysis, et al. Integrating common and rare
- 625 genetic variation in diverse human populations. *Nature*. 2010;467:52–58.
- Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: Using Sequence and
  Genotype Data to Estimate Haplotypes and Unobserved Genotypes. *Genet Epidemiol.* 2010;34:816–834.
- 629 25. Venner E, Muzny D, Smith JD, Walker K, Neben CL, Lockwood CM, Empey PE,
   630 Metcalf GA, Kachulis C, All of Us Research Program Regulatory Working Group,

- Mian S, Musick A, Rehm HL, Harrison S, Gabriel S, Gibbs RA, Nickerson D, Zhou
  AY, Doheny K, Ozenberger B, Topper SE, Lennon NJ. Whole-genome sequencing
  as an investigational device for return of hereditary disease risk and
  pharmacogenomic results as part of the All of Us Research Program. *Genome*
- 635 *Med.* 2022;14:34.
- 636 26. Lim M, Hastie T. Learning Interactions via Hierarchical Group-Lasso
- Regularization. Journal of Computational and Graphical Statistics. 2015;24:627–654.
- 639 27. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or
  640 more correlated receiver operating characteristic curves: a nonparametric
  641 approach. *Biometrics*. 1988;44:837–845.
- 642 28. Pencina MJ, Steyerberg EW, D'Agostino RB. Extensions of net reclassification
  643 improvement calculations to measure usefulness of new biomarkers. *Stat Med.*644 2011;30:11–21.
- 29. Yengo L, Vedantam S, Marouli E, Sidorenko J, Bartell E, Sakaue S, Graff M,
  29. Yengo L, Vedantam S, Marouli E, Sidorenko J, Bartell E, Sakaue S, Graff M,
  Eliasen AU, Jiang Y, Raghavan S, Miao J, Arias JD, Mukamel RE, Spracklen CN,
  Yin X, Chen S-H, Ferreira T, Ji Y, Karedera T, Lull K, Lin K, Malden DE, MedinaGomez C, Machado M, Moore A, Rueger S, Group GC-HW, Team 23andMe
  Research, Program VMV, Initiative) D (DiscovEHR and MCH, Network) eMERGE
- 650 (Electronic MR and G, Study LC, Center RG, Consortium TP, Group USS,
- 651 Chasman DI, Cho YS, Heid IM, McCarthy MI, Ng MCY, O'Donnell CJ, Rivadeneira
  652 F, Thorsteinsdottir U, Sun YV, Thai ES, Boehnke M, Deloukas P, Justice AE,
  653 Lindgren CM, Loos RJF, Mohlke KL, North KE, Stefansson K, Walters RG, Winkler
  654 TM, Yaung KL, Job RD, Yang J, Falsa T, Aasimaa TL, Auton A, Abasasia CD
- TW, Young KL, Loh P-R, Yang J, Esko T, Assimes TL, Auton A, Abecasis GR,
  Willer CJ, Locke AE, Berndt SI, Lettre G, Frayling TM, Okada Y, Wood AR,
  Visscher PM, Hirschhorn JN. A Saturated Map of Common Genetic Variants
  Associated with Human Height from 5.4 Million Individuals of Diverse Ancestries
  [Internet]. 2022 [cited 2022 Jan 11]. Available from:
- 659 https://www.biorxiv.org/content/10.1101/2022.01.07.475305v1
- 660 30. Gurdasani D, Carstensen T, Fatumo S, Chen G, Franklin CS, Prado-Martinez J, 661 Bouman H, Abascal F, Haber M, Tachmazidou I, Mathieson I, Ekoru K, DeGorter 662 MK, Nsubuga RN, Finan C, Wheeler E, Chen L, Cooper DN, Schiffels S, Chen Y, 663 Ritchie GRS, Pollard MO, Fortune MD, Mentzer AJ, Garrison E, Bergström A, 664 Hatzikotoulas K, Adeyemo A, Doumatey A, Elding H, Wain LV, Ehret G, Auer PL, 665 Kooperberg CL, Reiner AP, Franceschini N, Maher D, Montgomery SB, Kadie C, 666 Widmer C, Xue Y, Seeley J, Asiki G, Kamali A, Young EH, Pomilla C, Soranzo N, 667 Zeggini E, Pirie F, Morris AP, Heckerman D, Tyler-Smith C, Motala AA, Rotimi C, 668 Kaleebu P, Barroso I, Sandhu MS. Uganda Genome Resource Enables Insights 669 into Population History and Genomic Discovery in Africa. Cell. 2019;179:984-670 1002.e36.

672

671

## 673 Figure Legends

### 674 Figure 1: Study cohorts

675 The sample diagram depicts participant subsets. UK Biobank participants with MRI data

- 676 were split by random ID into a GWAS group (blue, N=39,524) and a residual group.
- 677 Individuals in the residual group related within 3 degrees of kinship were removed
- 678 (N=557), and the residual group was then split into a training group (purple, N=4,896)
- and an internal validation group (yellow, N=4,962). As depicted in the legend, the
- 680 GWAS and PRS development was conducted in the GWAS group (blue, N=39,524).

681 Clinical model development was conducted in the GWAS and training groups (blue and

682 purple, N=44,420). The PRS was incorporated into the clinical score using a linear

683 model to produce the AORTA Gene model in the training group (purple, N=4,896). All

684 models were validated in the internal validation set (yellow, N=4,962) and the external

validation sets (orange): MGB, FHS, and *All of Us*. MGB: Mass General Brigham

686 Biobank. FHS: Framingham Heart Study. MRI: magnetic resonance imaging. GWAS:

687 genome-wide association study.

#### 688 Figure 2: Score calibration curves in the UK Biobank validation set

689 Counterclockwise from top left: AORTA Gene, followed by the clinical AORTA Score,

690 the age and sex model, and the age/sex/genetics model. The x axis represents the

- 691 predicted ascending aortic diameter; the y axis represents the measured ascending
- aortic diameter; both are truncated at the same points so that the x and y axes have the
- 693 same span. Each point represents one of the 4,962 UK Biobank internal validation set
- 694 participants; orange points represent women while navy points represent men. The blue

695 line shows the smoothed average value, and only extends along the x axis to the limits 696 of the observed data. The black line shows the line of ideal calibration, where a 1 cm 697 greater score would be met by a 1 cm greater aortic diameter. For all plotted scores, the 698 slopes were statistically indistinguishable from one and the intercepts were statistically 699 indistinguishable from zero as detailed in the Results.

### **Figure 3: Variance in a ortic diameter explained by the models**

For each cohort, the variance in ascending aortic diameter explained by the model

derived from age and sex (navy blue), the clinical model (AORTA Score, slate gray),

and the clinical model incorporating the polygenic score (AORTA Gene, orange) is

depicted for the internal validation set and the three external validation sets. MGB: Mass

705 General Brigham Biobank. FHS: Framingham Heart Study.

### 706 Figure 4: Receiver operator characteristic curves

Receiver operator characteristic (ROC) curves for the AORTA Gene model (red) and the clinical
AORTA Score model (blue) in the validation cohorts. The dashed diagonal line represents the
no-information baseline. Counterclockwise from top left: UK Biobank internal validation, MGB,
FHS, and *All of Us*. The AUROC for detecting aortic diameter ≥4cm for the AORTA Gene model
in these cohorts was, respectively, 0.834, 0.808, 0.856, and 0.827. The respective AUROC for
the clinical AORTA Score was 0.765, 0.767, 0.818, and 0.791. MGB: Mass General Brigham
Biobank. FHS: Framingham Heart Study.

# 714 Tables

## 715 Table 1: Participant characteristics

	UK Biobank MRI Training	UK Biobank MRI Internal Validation	MGB TTE External Validation	FHS CT External Validation	<i>All of Us</i> TTE External Validation	<i>All of Us</i> Aneurysm External Validation
N	44,420	4,962	5,469	1,298	610	164,789
Men	21,053 (47.4%)	2,344 (47.2%)	2,580 (47.2%)	617 (47.5%)	246 (40.3%)	70,861 (43.0%)
Women	23,367 (52.6%)	2,618 (52.8%)	2,889 (52.8%)	681 (52.5%)	364 (59.7%)	93,928 (57.0%)
Age (years)	65.2 (7.7)	65.4 (7.7)	61.1 (14.4)	59.8 (9.0)	66.1 (11.0)	60.3 (10.9)
Medical history						
Hypertension	13,299 (29.9%)	1,501 (30.2%)	4,535 (82.9%)	532 (41.0%)	426 (69.8%)	57,845 (35.1%)
Hyperlipidemia	9,644 (21.7%)	1,103 (22.2%)	942 (17.2%)	257 (19.8%)	155 (25.4%)	6,379 (3.9%)
Diabetes	1,277 (2.9%)	160 (3.2%)	1,750 (32.0%)	115 (8.9%)	210 (34.4%)	25,568 (15.5%)
Heart rate (bpm)	68 (11)	68 (11)	76 (15)	65 (11)	69 (11)	72 (12)
Systolic blood pressure (mmHg)	140 (18)	141 (18)	128 (17)	126 (18)	129 (18)	129 (18)
Diastolic blood pressure (mmHg)	79 (10)	79 (10)	75 (10)	75 (9)	77 (11)	78 (11)
Height (cm)	169 (9)	169 (9)	169 (10)	168 (9)	166 (10)	168 (10)
Weight (kg)	75.2 (14.7)	75.4 (14.9)	84.4 (21.2)	80 (17.3)	88.9 (24.9)	84.5 (21.8)
BMI (kg/m2)	26.3 (4.2)	26.3 (4.3)	29.5 (6.6)	28.3 (5.2)	32.0 (8.4)	30.0 (7.3)
Ascending aortic diameter (cm)	3.18 (0.35)	3.19 (0.35)	3.20 (0.48)	3.45 (0.36)	3.34 (0.45)	NA
Aortic diameter ≥4cm	1,040 (2.3%)	113 (2.3%)	335 (6.1%)	116 (8.9%)	65 (10.7%)	NA
Prevalent aneurysm	NA	NA	NA	NA	NA	1,904 (1.2%)
Incident aneurysm	NA	NA	NA	NA	NA	1,632 (1.0%)

716 All values are mean (standard deviation) or count (percent of total). BPM: beats per

717 minute. MGB: Mass General Brigham. BMI: body mass index. MRI: magnetic resonance

718 imaging. CT: computed tomography. TTE: transthoracic echocardiography.

Figures

# Figure 1: Study cohorts





# Figure 2: Score calibration curves in the UK Biobank validation set



# Figure 3: Variance in aortic diameter explained by the models



# Figure 4: Receiver operator characteristic curves