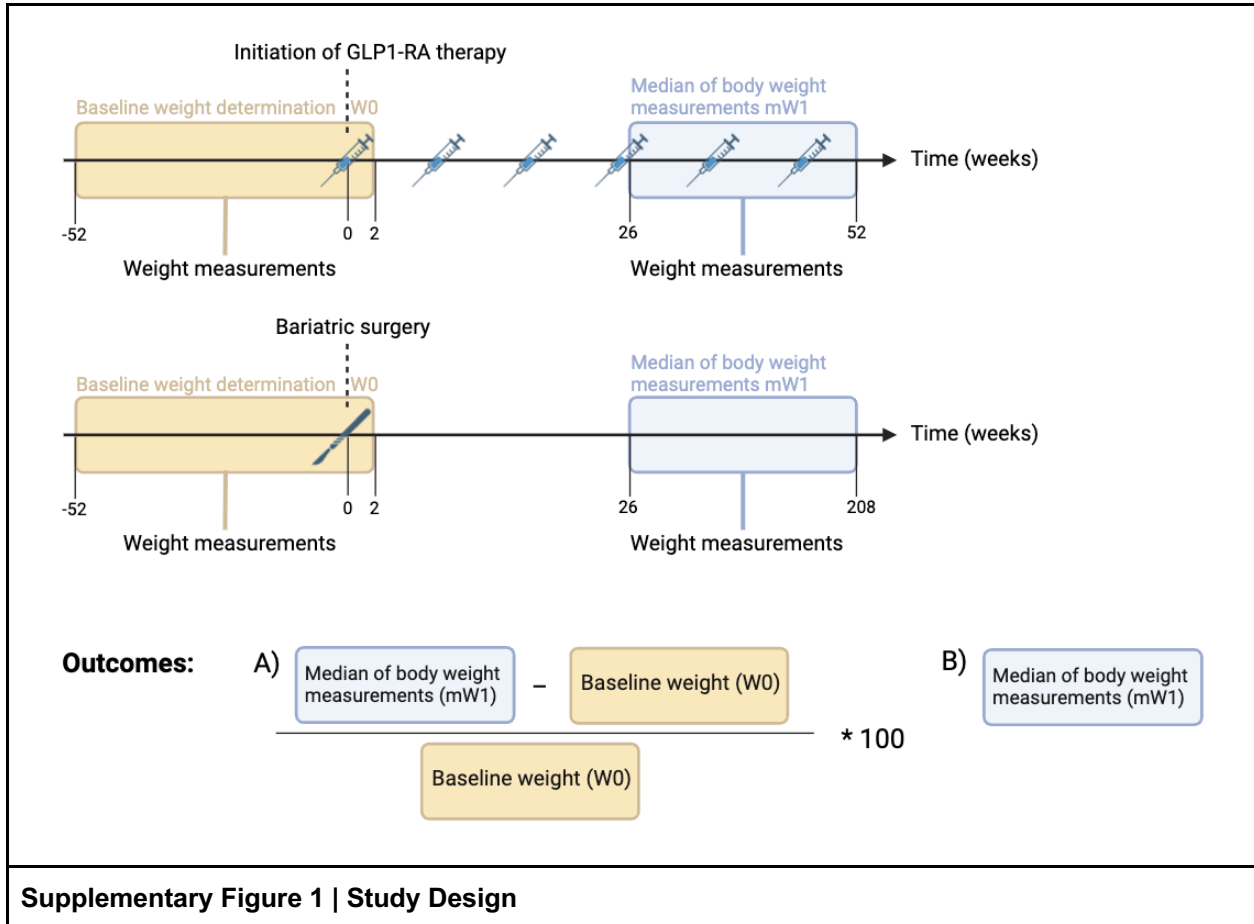


1 Supplementary Figure 1 – Study Design

2

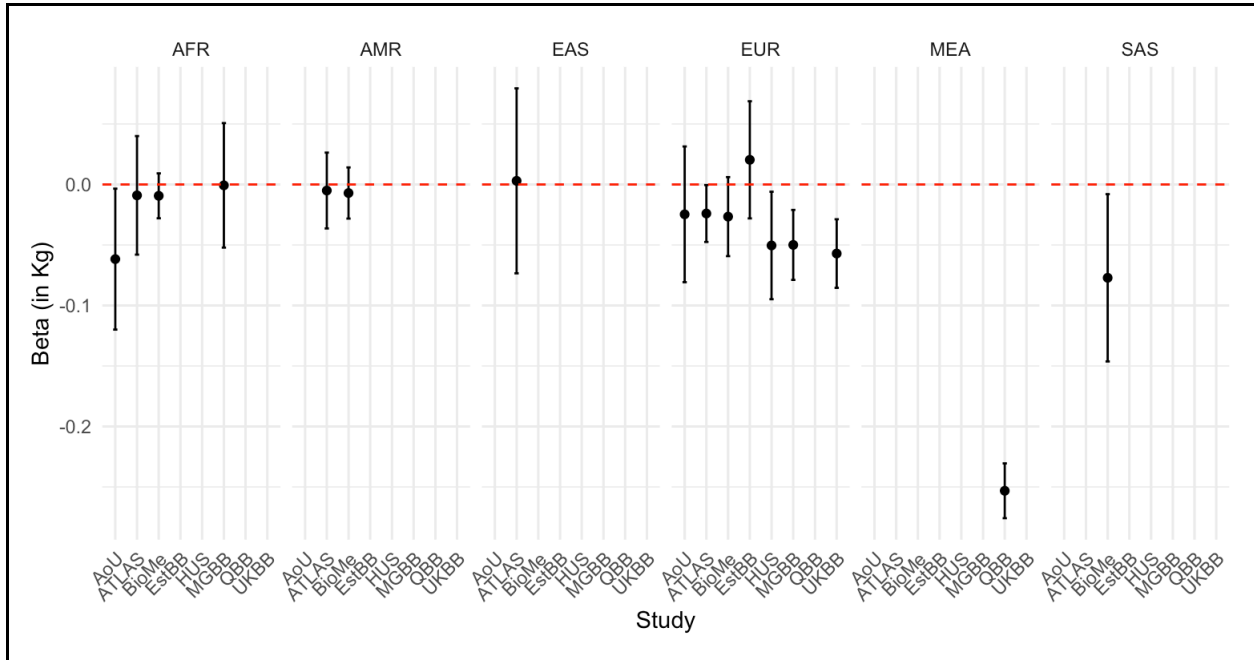


3

4

Supplementary Figure 1 | Study Design

5 Supplementary Figure 2 – Effect sizes of baseline
 6 weight on weight loss amongst GLP1-RA users
 7 across studies by ancestries
 8

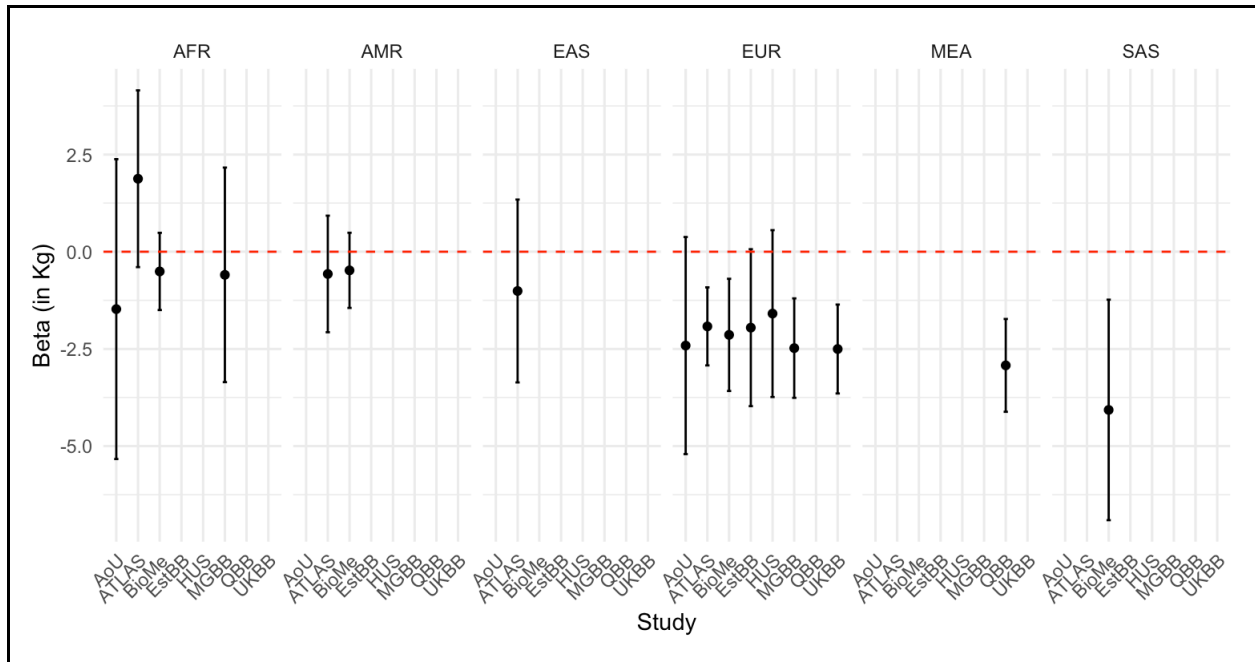


Supplementary Figure 2 | Effect sizes of baseline weight on weight loss amongst GLP1-RA users across studies by ancestries. Points represent beta coefficients in Kg and error bars represent the 95% confidence intervals.
 AoU: All of Us, ATLAS: UCLA-Atlas Biobank, BioMe: BioMe Biobank, ESTBB: Estonian Biobank, HUS: Helsinki University Hospital, MGGB: Mass General Brigham Biobank, QBB: Qatar Biobank, UKBB: UK Biobank.

9
 10
 11

12 Supplementary Figure 3 – Effect sizes of sex on
 13 weight loss amongst GLP1-RA users across studies
 14 by ancestries

15



Supplementary Figure 3 | Effect sizes of sex on weight loss amongst GLP1-RA users across studies by ancestries. Points represent beta coefficients in Kg and error bars represent the 95% confidence intervals.

AoU: All of Us, ATLAS: UCLA-Atlas Biobank, BioMe: BioMe Biobank, ESTBB: Estonian Biobank, HUS: Helsinki University Hospital, MGGB: Mass General Brigham Biobank, QBB: Qatar Biobank, UKBB: UK Biobank.

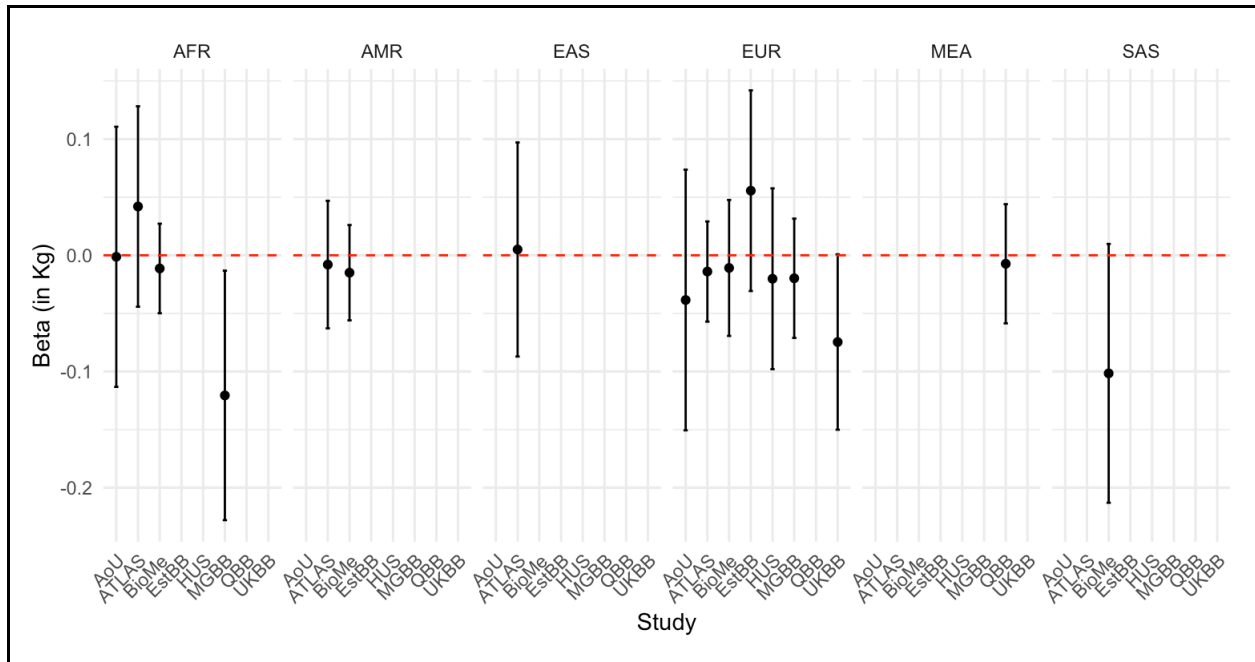
16

17

18

19 Supplementary Figure 4 – Effect sizes of age at
 20 treatment initiation on weight loss amongst GLP1-
 21 RA users across studies by ancestries

22



Supplementary Figure 4 | Effect sizes of age at treatment initiation on weight loss amongst GLP1-RA users across studies by ancestries. Points represent beta coefficients in Kg and error bars represent the 95% confidence intervals.

AoU: All of Us, ATLAS: UCLA-Atlas Biobank, BioMe: BioMe Biobank, ESTBB: Estonian Biobank, HUS: Helsinki University Hospital, MGBB: Mass General Brigham Biobank, QBB: Qatar Biobank, UKBB: UK Biobank.

23

24 Supplementary Material – Cohort descriptions and 25 ancestry definitions

26 All of Us (AoU)

27 The All of Us Research Program is a longitudinal cohort study aiming to enroll at least one
28 million individuals across a diverse population in the USA. Over 245,000 individuals have
29 genotype data. The sensitivity for single nucleotide variants was > 98.7% and the precision was
30 > 99.9%.

31 Ancestry categories were based on gnomAD and inferred using PCA data. The patterns of
32 ancestry and admixture were compared to self-identified race and ethnicity, and continuous
33 ancestry inference using genome-wide genotypes resulted in concordant estimates. We
34 restricted participants to an unrelated subset. The participants were further filtered based on the
35 criteria described in the analysis plan (e.g., per ancestry minimum allele count/sample size after
36 filtering on availability of weight measurements and prescription and/or bariatric surgery).

37 Bialystok Bariatric Surgery Study (BBSS)

38 The Bialystok Bariatric Surgery Study (BBSS) is a prospective cohort study of patients
39 undergoing bariatric surgery at the First Clinical Department of General and Endocrine Surgery
40 at the Medical University of Bialystok. This is the primary receiving center for patients referred
41 for bariatric surgery in the province of Podlaskie Voivodeship and the largest center by number
42 of bariatric surgeries performed in northeastern Poland. This center specializes in several
43 bariatric surgical techniques including Roux-en-Y gastric bypass, gastric banding, and sleeve
44 gastrectomy. For this study, we selected only patients who underwent sleeve gastrectomy since
45 it represents the vast majority (>90%) of all interventions performed at the center and in order to
46 eliminate confounding variation in surgical technique. The BBSS began in 2015 and consisted
47 of a battery of baseline tests established one month prior to the intervention and repeated at
48 one, three, six, and twelve-month follow-up clinical visits. Subsequently, patients are examined
49 every year after the first year. At each visit, all subjects underwent physical examination, body
50 composition analysis, and blood testing, as well as completed diet and physical activity
51 questionnaires. All subjects give their informed consent for inclusion before participating in the
52 study. The study is conducted in accordance with the Declaration of Helsinki, and the protocol

53 was approved by the Ethics Committee of the Medical University of Bialystok (Project
54 identification code: R-I-002/546/2015).

55 BioMe

56 The BioMe Biobank, founded in 2007, is an ongoing electronic health record (EHR)-linked
57 biorepository that enrolls participants non-selectively from across the Mountain Sinai Health
58 System (MSHS). Over 60,000 participants have been recruited from >26 outpatient sites
59 located in Manhattan and Queens, and recruitment is ongoing. We restricted participants to an
60 unrelated subset (based on 2nd degree or greater KING-derived kinship coefficients) who had
61 been genotyped with either the Global Screening Array (GSA) or Global Diversity Array (GDA).
62 We then filtered this population to those adhering to the criteria described in the analysis plan
63 (e.g., per-ancestry minimum allele count/sample size after filtering on availability of weight
64 measurements and prescription and/or bariatric surgery).

65 Genetically determined ancestry group based on ADMIXTURE with $K=10$, with grouping based
66 on which 1000 Genomes phase 3 reference samples (with superpopulation labels as reference)
67 showed the highest proportion for a given class. The choice for $K=10$ was based on a
68 combination of minimizing cross-validation errors for admixture analysis, as well as minimizing
69 the Frobenius distance of each ADMIXTURE matrix K with the matrix outputted by Genetic
70 Relationship and Fingerprinting (GrafPop).

71

72 Estonian Biobank (ESTBB)

73 ESTBB is a population-based biobank that involves over 212,000 adult participants (about 20%
74 of the Estonian adult population). The Estonian Biobank Project was initiated in 1999; data
75 collection began in 2002. Initially, participants were recruited by general practitioners across
76 Estonia from among individuals visiting general practitioners' offices and hospitals. Consenting
77 adults donated blood samples for DNA extraction, underwent anthropometric, blood pressure,
78 and resting heart rate measurement, and provided extensive information on demographic
79 characteristics, genealogy, education, occupation, and lifestyle, as well as health status and
80 medical history (Leitsalu et al., 2014, 2015). From 2017 onward, the process of joining ESTBB
81 was simplified: new joiners were required to sign an online consent form, visit a healthcare
82 provider or pharmacy to provide a blood sample, and fill in an online questionnaire. Additional

83 data are collected through linkage with various national databases and registries and through
84 subsequent studies involving different subsets of participants.

85 Participants were genotyped using the Illumina Global Screening Array (GSA) v1.0, v2.0, and
86 v2.0_EST. Samples were genotyped and PLINK format files created using Illumina
87 GenomeStudio v2.0.4. Individuals with call rates below 95% or whose sex in phenotype data did
88 not match X chromosome heterozygosity were excluded. Before imputation, variants were
89 filtered by call rate <95%, Hardy–Weinberg equilibrium p -value < $1e^{-4}$ (autosomal variants only),
90 and minor allele frequency <1%. A population-specific imputation reference panel of 2,297 WGS
91 samples was used (Mitt et al., 2017).

92 We aligned participants to 21 major ancestry groups similarly to Privé (2021). We removed
93 participants with non-European assigned group ancestry, keeping Europeans, Finns and
94 Italians. <https://doi.org/10.1093/bioinformatics/btac348>.

95

96 Helsinki University Hospital / Helsinki Biobank (HUS)

97 The Helsingin Biopankki (Helsinki Biobank) was founded in 2015 by HUS Helsinki University
98 Hospital, University of Helsinki, the Joint Municipality for Health and Social Services in the
99 Kymenlaakso Valley (Kymsote) and the South Karelia Social and Health Care District (Exote),
100 and is the largest hospital biobank in Finland. Approximately 80,000 patients have given a blood
101 sample based on biobank consent and used for research purposes. Diagnostic formalin-fixed
102 paraffin-embedded tissue samples (FFPE) from approximately one million patients have been
103 transferred to the biobank's sample collection. The biobank's samples were combined with
104 clinical patient data from the hospital's database (HUS Datapool) in accordance with the
105 description in the research plan recommended by the ethics committee. Data that are run
106 together can usually include the sampling date, the donor's gender, histological and cytological
107 examinations, diagnoses, care measures and laboratory examination results.

108

109 Mass General Brigham Biobank (MGBB)

110 Mass General Brigham is an integrated healthcare system located in Boston, Massachusetts for
111 more than 1.5 million individuals per year. The Mass General Brigham Biobank (MGBB) has
112 recruited over 140,000 participants since 2008. Over 65,000 participants have genotype data
113 sequenced using the Illumina Multi-Ethnic Genotyping MEG) assay or the Illumina Global

114 Screening Array (GSA). All genotyped biospecimens had at least 99% call rate per array and
115 were confirmed for sex concordance using sex in EHR and sex computed from array data.
116 Ancestry was defined using high-dimensional principal components, specifically top 30 genetic
117 PCs and a network-based clustering approach. We restricted participants to an unrelated
118 subset. The participants were further filtered based on the criteria described in the analysis plan
119 (e.g., per ancestry minimum allele count/sample size after filtering on availability of weight
120 measurements and prescription and/or bariatric surgery).

121

122 UCLA-ATLAS (Atlas Biobank)

123 ATLAS is an EHR-linked biobank across multiple institutions within the UCLA Health System.
124 After participants signed informed consent for participation to the UCLA ATLAS Community
125 Health Initiative, biological samples in the form of de-identified blood samples were collected
126 and subsequently processed for DNA extraction and genotyping. The data analyzed in this
127 manuscript draws from a “frozen snapshot” of ATLAS data encompassing all available samples
128 up to February 15, 2024, totaling N=53829. The corresponding EHR data for ATLAS
129 participants was sourced from the UCLA Data Discovery Repository (DDR). This repository
130 operates under the auspices of the UCLA Health Office of Health Informatics Analytics and the
131 UCLA Institute of Precision Health. Patient Recruitment and Sample Collection for Precision
132 Health Activities at UCLA is an approved study by the UCLA Institutional Review Board.
133 Extensive details on ATLAS genotyping and quality control were previously reported in [PMID:
134 36777178]. In short, participants in the ATLAS initiative were genotyped using a custom
135 genotyping array constructed from the Global Screening Array with the Infinium Global
136 Screening Array-24 Kit (Illumina, San Diego, California), a multi-disease drop-in panel aligned
137 with the GRCh38 assembly. SNPs were subjected to removal if they were unmapped, strand-
138 ambiguous, duplicates, or exhibited >5% missingness. Samples with missingness exceeding
139 >5% were excluded, as were duplicates (including identical twins or triplets) with preference
140 given to individuals with the lowest missing rate in each pair. Imputation to the
141 TOPMedFreeze5 was conducted using the Michigan Imputation Server [PMID: 27571263]. After
142 imputation, SNPs with a quality score (R2) below 0.90 were excluded, for polygenic score
143 calculation SNPs with a MAF below 0.1% were filtered out.
144 The genetic ancestry of ATLAS individuals was estimated by assessing their proximity to 1000
145 Genome super populations on the principal component (PC) space. We computed the top 20
146 PCs using the bigsnpr R software with default parameters. Subsequently, leveraging the super

147 population label and PCs of the 1000 Genome individuals, we train a K-nearest neighbors
148 model to assign genetic ancestry labels to each ATLAS individual. To optimize model
149 performance, we employed 10-fold cross-validation to select the optimal hyper-parameter (k)
150 from a range of values (k = 5, 10, 15, 20). In instances where an individual was assigned to
151 multiple ancestries with probability larger than 0.5 or was not assigned to any cluster, their
152 ancestry was labelled as unknown. We clustered 94% of the ATLAS participants into one of the
153 five 1000 Genome super population (AFR, AMR, EAS, EUR, SAS) while the remaining
154 individuals' ancestry could not be ascertained.

155 UK Biobank (UKBB)

156 The UK Biobank (UKBB) is a large-scale biomedical database and research resource containing
157 in-depth genetic and health information from half a million UK participants. Upon signing
158 informed consent, participants provided biological samples, including de-identified blood
159 samples, which were processed for DNA extraction and genotyping. The dataset analyzed in
160 this manuscript is derived from a “frozen snapshot” of the UKBB data, encompassing all
161 available samples up to January 31, 2024, totaling N=502,536. Corresponding electronic health
162 record (EHR) data for UKBB participants were sourced from the UK Biobank Resource, which
163 operates under the auspices of the UK Biobank Steering Committee and has received ethical
164 approval from the North West Multi-centre Research Ethics Committee. Detailed information on
165 UKBB genotyping and quality control has been previously reported in
166 https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/genotyping_qc.pdf. Samples were genotyped
167 at the Affymetrix Research Services Laboratory in Santa Clara, California, USA. Upon receipt of
168 a 96-well plate containing 94 UK Biobank samples, Affymetrix added two control individuals
169 (from 1000 Genomes) to the same well positions on each plate. Genotypes were then called
170 from the resulting intensities in batches of ~4,700 samples (~4,800 including the controls) using
171 the Affymetrix Power Tools software and the Affymetrix Best Practices Workflow. After
172 genotype calling, Affymetrix performed quality control in each batch separately, to exclude
173 SNPs with poor cluster properties. If a SNP did not meet the Affymetrix prescribed QC
174 thresholds in a given batch, it was set to missing in all individuals from that batch. Affymetrix
175 also checked sample quality (such as DNA concentration) and genotype calls were provided
176 only for samples with sufficient DNA metrics.

177 The ancestry composition of the UK Biobank participants was defined based on self-reported
178 ethnicities and genetic principal component analysis (PCA). Participants self-reported their
179 ethnic backgrounds into categories such as White, Mixed, Asian or Asian British, Black or Black

180 British, and Other ethnic groups, with the majority (~94%) identifying as White British. Genetic
181 PCA was used to capture the underlying population structure and define genetic ancestry more
182 precisely, categorizing participants into groups such as European, South Asian, African, East
183 Asian, and other ancestries. This approach ensured a nuanced understanding of the genetic
184 and demographic diversity within the cohort, enabling robust and reliable biomedical research.

185 To capture population structure specific to the UK Biobank cohort, we performed principal
186 component analysis of ~150,000 UK Biobank samples using ~100,000 SNPs. These PCs can
187 be used to identify samples with similar ancestry or to control for population structure in
188 association studies.

189

190 Qatar Biobank (QBB)

191 This study was based on the second release of 14,060 Qatari participants (all from the same
192 ancestry, Middle-Eastern - MEA) from the population based Qatar Genome Program (QGP).
193 Ethical approval was provided by the institutional review board of the Qatar Biobank (QBB)
194 (Protocol no. QF-QBB-RES-ACC-00205). A detailed description of the recruitment process and
195 collection of phenotypic data by the QGP and the QBB was previously reported at
196 <https://doi.org/10.1093/aje/kwz084>. A signed consent was obtained from all included
197 participants by the QBB, and WGS data used in the current study was obtained from the QGP
198 as previously described (<https://doi.org/10.1002/humu.24336>). All samples were sequenced on
199 Illumina HiSeq X instruments (Illumina, San Diego, CA, USA) to obtain WGS with a target
200 average depth of coverage of 30x. The read mapping, variant calling and joint variant calling
201 were performed using Sentieon's DNaseq pipeline v201808.03 (Sentieon, San Jose, CA, USA),
202 following the BWA-GATK Best Practice Workflow and using the GRCh38/hg38 reference
203 genome. The produced gVCF were jointly called to produce one multisample VCF file (msVCF)
204 for the whole cohort. Clinical data are extracted from the EMR of the Hamad Medical
205 Corporation (HMC).