1 **Impact of rare and common genetic variants on diabetes**

2 **diagnosis by hemoglobin A1c in multi-ancestry cohorts:**

3 **The Trans-Omics for Precision Medicine Program**

4

5 **Authors**

6 Chloé Sarnowski,[1,35,37,]* Aaron Leong,[2,3,4,35,38,]** Laura M Raffield,[5] Peitao Wu,[1] Paul S de

7 Vries,[6] Daniel DiCorpo,[1] Xiuqing Guo,[7] Huichun Xu,[8] Yongmei Liu,[9] Xiuwen Zheng,[10] Yao

8 Hu,[11] Jennifer A Brody,[12] Mark O Goodarzi,[13] Bertha A Hidalgo,[14] Heather M Highland,[15]

9 Deepti Jain,[10] Ching-Ti Liu,[1] Rakhi P Naik,[16] James A Perry,[17] Bianca C Porneala,[2] Elizabeth

10 Selvin,[18] Jennifer Wessel,[19,20] Bruce M Psaty,[12,21,22] Joanne E Curran,[23] Juan M Peralta,[23] John

11 Blangero,[23] Charles Kooperberg,[11] Rasika Mathias,[18,24] Andrew D Johnson,[25,26] Alexander P

12 Reiner,[11,27] Braxton D Mitchell,[8,28] L Adrienne Cupples,[1,25] Ramachandran S Vasan,[25,29,30]

13 Adolfo Correa,[31,32] Alanna C Morrison,[6] Eric Boerwinkle,[6,33] Jerome I Rotter,[7] Stephen S Rich,[34]

14 Alisa K Manning,[2,3,4] Josée Dupuis,[1,25,36] James B Meigs,[2,3,4,36] on behalf of the Trans-Omics for

15 Precision Medicine (TOPMed) Diabetes and TOPMed Hematology and Hemostasis working

16 groups and the NHLBI TOPMed Consortium.

17

18 **Affiliations**

19 [1] Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118,

20 USA

21 [2] Division of General Internal Medicine, Massachusetts General Hospital, Boston 02114, MA

22 USA

23 [3] Department of Medicine, Harvard Medical School, Boston, MA 02115, USA

24   [4] Programs in Metabolism and Medical & Population Genetics, Broad Institute of MIT and

25   Harvard, Cambridge, MA 02142, USA

26   [5] Department of Genetics, University of North Carolina, Chapel Hill, NC 27514, USA

27   [6] Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental

28   Sciences, School of Public Health, University of Texas Health Science Center at Houston,

29   Houston, TX 77030, USA

30   [7] Institute for Translational Genomics and Population Sciences, LABioMed and Department of

31   Pediatrics at Harbor-UCLA Medical Center, Torrance, CA 90502, USA

32   [8] Division of Endocrinology, Diabetes and Nutrition, Department of Medicine, University of

33   Maryland School of Medicine, Baltimore, MD 21201, USA

34   [9] Department of Epidemiology & Prevention, Wake Forest School of Medicine, Winston-Salem,

35   NC 27101, USA

36   [10] Department of Biostatistics, University of Washington, Seattle, WA 98195, USA

37   [11] Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA

38   98108, USA

39   [12] Cardiovascular Health Research Unit, Department of Medicine, University of Washington,

40   Seattle, WA 98195, USA

41   [13] Division of Endocrinology, Diabetes, and Metabolism, Department of Medicine, Cedars-Sinai

42   Medical Center, Los Angeles, CA 90048, USA

43   [14] University of Alabama at Birmingham, Department of Epidemiology, Birmingham, AL 35294,

44   USA

45   [15] Department of Epidemiology, UNC Gillings School of Global Public Health, University of

46   North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

47   [16] Division of Hematology, Department of Medicine, Johns Hopkins University, Baltimore, MD

48   21205, USA

49   [17] University of Maryland School of Medicine, Baltimore, MD 21205, USA

50   [18] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore,

51   MD 21205, USA

52   [19] Department of Epidemiology, Indiana University Fairbanks School of Public Health,

53   Indianapolis, IN 46202, USA

54   [20] Department of Medicine and Diabetes Translational Research Center, Indiana University

55   School of Medicine, Indianapolis, IN 46202, USA.

56   [21] Kaiser Permanente Washington Health Research Institute, Seattle, WA 98101, USA.

57   [22] Departments of Epidemiology and Health Services, University of Washington, Seattle, WA

58   98195, USA.

59   [23] Department of Human Genetics and South Texas Diabetes and Obesity Institute, University of

60   Texas Rio Grande Valley School of Medicine, Brownsville, TX 78520, USA

61   [24] GeneSTAR Research Program, Department of Medicine, Johns Hopkins University,

62   Baltimore, MD 21205, USA

63   [25] National Heart Lung and Blood institute and Boston University's Framingham Heart Study,

64   Framingham MA 01702, USA

65   [26] Population Sciences Branch, National Heart, Lung, and Blood Institute, National Institutes of

66   Health, Bethesda, MD 20814, USA

67   [27] Department of Epidemiology, University of Washington, Seattle, WA 98195, USA

68   [28] Geriatrics Research and Education Clinical Center, Baltimore Veterans Administration

69   Medical Center, Baltimore, MD 21201, USA

70   [29] Section of Preventive Medicine and Epidemiology, Evans Department of Medicine, Boston

71   University School of Medicine, Boston, MA 02118, USA

72   [30] Whitaker Cardiovascular Institute and Cardiology Section, Evans Department of Medicine,

73   Boston University School of Medicine, Boston, MA 02118, USA

74   [31] Departments of Medicine, Pediatrics and Population Health Science, University of Mississippi

75   Medical Center, Jackson, MS 39216, USA

76   [32] The Jackson Heart Study, Jackson, MS 39213, USA

77   [33] Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA

78   [34] Center for Public Health Genomics, University of Virginia, Charlottesville, VA 22908, USA

79

80   **Joint Authorship**

81   [35] These authors contributed equally to this work

82   [36] These authors contributed equally to this work

83

84   **Present Address**

85   [37] Department of Biostatistics, Boston University School of Public Health, 801 Massachusetts

86   Avenue, MA 02118, Boston, USA

87   [38] Division of General Internal Medicine, Massachusetts General Hospital, 100 Cambridge St

88   16th Floor, MA 02114, Boston, USA

89

90   **Correspondence**

91   *chloesar@bu.edu

92   **ASLEONG@partners.org

Abstract

Hemoglobin A1c (HbA1c) is widely used to diagnose diabetes and assess glycemic control in patients with diabetes. However, nonglycemic determinants, including genetic variation, may influence how accurately HbA1c reflects underlying glycemia. Analyzing the NHLBI Trans-Omics for Precision Medicine (TOPMed) sequence data in 10,338 individuals from five studies and four ancestries (6,158 Europeans, 3,123 African-Americans, 650 Hispanics and 407 East Asians), we confirmed five regions associated with HbA1c (*GCK* in Europeans and African-Americans, *HK1* in Europeans and Hispanics, *FN3K/FN3KRP* in Europeans and *G6PD* in African-Americans and Hispanics) and discovered a new African-ancestry specific low-frequency variant (rs1039215 in *HBG2/HBE1*, minor allele frequency (MAF)=0.03). The most associated *G6PD* variant (p.Val98Met, rs1050828-T, MAF=12% in African-Americans, MAF=2% in Hispanics) lowered HbA1c (-0.88% in hemizygous males, -0.34% in heterozygous females) and explained 23% of HbA1c variance in African-Americans and 4% in Hispanics. Additionally, we identified a rare distinct *G6PD* coding variant (rs76723693 - p.Leu353Pro, MAF=0.5%; -0.98% in hemizygous males, -0.46% in heterozygous females) and detected significant association with HbA1c when aggregating rare missense variants in *G6PD*. We observed similar magnitude and direction of effects for rs1039215 (*HBG2*) and rs76723693 (*G6PD*) in the two largest TOPMed African-American cohorts and replicated the rs76723693 association in the UK Biobank African-ancestry participants. These variants in *G6PD* and *HBG2* were monomorphic in the European and Asian samples. African or Hispanic ancestry individuals carrying *G6PD* variants may be underdiagnosed for diabetes when screened with HbA1c. Thus, assessment of these variants should be considered for incorporation into precision medicine approaches for diabetes diagnosis.

Introduction

Hemoglobin A1c (HbA1c) is a convenient indirect measure of long-term exposure to blood glucose concentrations. HbA1c estimates the proportion of glycated hemoglobin in the blood, an irreversible chemical modification of the hemoglobin molecule by blood glucose.[1] As HbA1c reflects average ambient glycemia over the previous 2-3 months, the life of an erythrocyte, it is a commonly used as a test to diagnose diabetes (MIM: 125853 and 222100) and estimate glycemic control in patients with diabetes. However, non-glycemic variation in HbA1c due to differences in erythrocyte turnover can influence how accurately HbA1c reflects underlying glycemia.[2]

We previously conducted a trans-ethnic genome-wide association study (GWAS) meta-analysis of HbA1c in 159,940 individuals from four ancestries (European, African, East Asian, and South Asian). We identified 60 common (minor allele frequency, MAF, greater than 5%) genetic variants associated with HbA1c, of which 19 were classified as 'glycemic' and 22 as 'erythrocytic' based on the probable biological mechanism through which they appeared to influence HbA1c levels.[3] Genetic variants affecting HbA1c via erythrocyte biological pathways may lead to diagnostic misclassification of ambient glycemia and thus diabetes status. HbA1c GWAS have so far focused on genetic variants imputed to HapMap,[4] therefore genetic discovery efforts have been focused on common variants. Low-frequency ($0.5\% < MAF < 5\%$) and rare ($MAF < 0.5\%$) genetic variants and their associated impact on the diagnostic accuracy of HbA1c have not been systematically examined but are suspected to occur.[5, 6] Further, previous GWAS have shown that the combined effect of HbA1c-related common variants causes differences in HbA1c that were three times greater in individuals of African ancestry compared with those of European ancestry [0.8 (%-units) vs. 0.25 (%-units)].[3] This relatively large difference in HbA1c

140   was mainly driven by a single African-ancestry specific missense variant (p.Val98Met,

141   rs1050828-T) in the Glucose-6-phosphatase Dehydrogenase gene (*G6PD* [MIM:305900])*,* which

142   causes G6PD deficiency (MIM: 300908).[7, 8] As the genetic architecture of HbA1c appears to

143   differ by ancestry, as does type 2 diabetes risk, it is imperative to understand the genetic basis of

144   HbA1c in different ancestral groups to ensure that large-effect ancestry-specific variants, similar

145   to the *G6PD* variant, are uncovered.[9]

146   Here, we sought to confirm common and identify low frequency and rare genetic variants

147   associated with HbA1c through association analyses in diabetes-free individuals from four

148   ancestries using whole genome sequencing (WGS) data from the NIH National Heart, Lung, and

149   Blood Institute (NHLBI) Trans-Omics for Precision Medicine (TOPMed) program. We

150   hypothesized that variants in the low frequency spectrum with relatively large effects will be

151   detected even with modest sample sizes. To uncover additional distinct erythrocytic variants

152   associated with HbA1c, we performed fine-mapping using sequential conditional association

153   analyses of erythrocytic loci that reached genome-wide significance in this study as well as gene-

154   based tests.

155

Methods

Populations and participants

158   We included in our analyses 10,338 TOPMed participants without diabetes from five

159   cohorts: the Old Order Amish study (N=151), the Atherosclerosis Risk in Communities Study

160   (ARIC, N=2,415), the Framingham Heart Study (FHS, N=2,236), the Jackson Heart Study (JHS,

161   N=2,356) and the Multi-Ethnic Study of Atherosclerosis (MESA, N=3,180) representing four

162   ancestry groups: Europeans (EA, N=6,158), African-Americans (AA, N=3,123),

163 Hispanics/Latinos (HA, N=650) and East Asians (AS, N=407) (**Supplementary Table 1**).

164 Descriptions of each cohort are available in the **Supplementary Text**. Diabetes was defined as

165 fasting glucose (FG) $\geq$ 7 mmol/L after $\geq$ 8 hours, HbA1c $\geq$ 6.5%-units, 2-hour glucose by an oral

166 glucose tolerance test $\geq$ 11.1 mmol/L, non-fasting glucose $\geq$ 11.1 mmol/L, physician diagnosed

167 diabetes, self-reported diabetes, or use of an antidiabetic medication. Measures of FG and 2-hour

168 glucose made in whole blood were corrected to plasma levels using the correction factor of 1.13.

169 Individual studies applied further sample exclusions where applicable, including pregnancy and

170 type 1 diabetes status.

171 **Measurement of HbA1c and erythrocytic traits**

172 The National Glycohemoglobin Standardization Program (NGSP) certified assays[10] used

173 to measure HbA1c in each cohort are indicated in **Supplementary Table 2**. HbA1c was

174 expressed in NGSP %-units. Measurements for red blood cell (RBC) count ($\times$1012/L),

175 hemoglobin (HB; g/dL), hematocrit (HCT; %), mean corpuscular volume (MCV; fL), mean

176 corpuscular hemoglobin (MCH; pg), mean corpuscular hemoglobin concentration (MCHC;

177 g/dL), and red blood cell distribution width (RDW; %) were obtained from complete blood count

178 panels performed using standard assays.

179 **Whole genome sequencing**

180 The NHLBI TOPMed program provided WGS, performed at an average depth of 38$\times$ by

181 several sequencing centers (New York Genome Center; Broad Institute of MIT and Harvard;

182 University of Washington Northwest Genomics Center; Illumina Genomic Services; Macrogen

183 Corp.; and Baylor Human Genome Sequencing Center), using DNA from blood. Details

184 regarding the laboratory methods, data processing and quality control are described on the

185 TOPMed website (see URL in the web resources section) and in documents included in each

8

186 TOPMed accession released on the database of Genotypes and Phenotypes (dbGaP). Processing

187 of whole genome sequences was harmonized across genomic centers using a standard pipeline.[11]

188   We used TOPMed 'freeze 5b' that comprised 54,508 samples. Variant discovery and

189 genotype calling were performed jointly across TOPMed parent studies for all samples using the

190 GotCloud pipeline. A support vector machine quality filter was trained using known variants

191 (positive training set) and Mendelian-inconsistent variants (negative training set). The TOPMed

192 data coordinating center performed additional quality control checks for sample identity issues

193 including pedigree errors, sex discrepancies, and genotyping concordance. After site level

194 filtering, TOPMed freeze 5b consisted of ~438 million single nucleotide variants (SNVs) and

195 ~33 million short insertion-deletion variants. Read mapping was done using the 1000 Genomes

196 Project reference sequence versions for human genome build GRCh38.

197 The study was approved by the appropriate institutional review boards (IRB) and informed

198 consent was obtained from all participants.

199 **Statistical analyses**

200   In each ancestry, we performed pooled WGS association analysis of HbA1c using

201 Genesis on the Analysis Commons (see URL in the web resources section).[12] We used linear

202 mixed effect models to test the association of HbA1c with the SNVs individually while adjusting

203 for sex, age at HbA1c measurement and study, allowing for heterogeneous variance across study

204 groups. We accounted for relatedness using an empirical kinship matrix (Genetic Relationship

205 Matrix, GRM). We excluded from our single-SNV analyses variants with a minor allele count

206 (MAC) less than 20 across the combined samples. We used a significance threshold of $P < 2 \times 10^{-8}$

207 to report an association as genome-wide significant for common, low frequency and rare

208 genetic variants, which was slightly more stringent than the widely adopted P-value threshold of

209 5×10^{-8} in GWAS, based on estimations for genome-wide significance for WGS studies in

210 UK10K.[13, 14] For the X chromosome, genotypes were coded as 0 and 2 for males and 0, 1 and 2

211 for females, and sex-stratified analyses (analyses conducted separately in males and females)

212 were also performed. We also performed additional analyses in AA, adjusted on the sickle cell

213 trait (SCT) variant rs334, as well as haplotype analyses with this variant, using the R package

214 haplo.stats.

215 Because loci that influence HbA1c through erythrocytic mechanisms are expected to

216 cause nonglycemic variation in HbA1c, we sought to classify HbA1c-associated loci as

217 'glycemic' or 'erythrocytic' using mediation analyses on FG, HB, MCV, MCH or MCHC. If the

218 HbA1c-variant association effect size changed by more than 25% upon adding FG to the

219 regression model, the variant was classified as 'glycemic'. If a variant was not 'glycemic', and if

220 the effect size changed by more than 25% upon adding HB, MCV, MCH, or MCHC to the

221 regression model, the variant was classified as 'erythrocytic'. If association effect sizes were

222 unchanged by mediation adjustment, then the variant was considered 'unclassified'.

223 If HbA1c-associated variants remained unclassified by the mediation analyses, we used

224 association analysis results with FG to classify them as 'glycemic' ($P < 0.005$), and association

225 analysis results with erythrocytic traits (HB, MCV, MCH, MCHC, RBC, HCT or RDW) to

226 classify them as 'erythrocytic' ($P < 0.005$). Association results for FG were obtained from a

227 pooled association analysis performed using a linear mixed effect model in Genesis, adjusted on

228 age, age squared, sex, BMI and self-report ancestry in TOPMed freeze 5b (N=26,883).

229 Association results for erythrocytic traits were obtained from several sources. We first used

230 single variant association analyses performed using a score test in Genesis and adjusted for age,

231 sex, and study, with a GRM, while allowing for heterogeneous variance across study groups, in

10

232   TOPMed freeze 5b (N=25,080). The number of individuals included in each analysis, per study,

233   is available in **Supplementary Table 3**. Analyses were performed at the University of

234   Washington using the TOPMed pipeline. We then used association results from published

235   GWAS in European (UK Biobank and INTERVAL studies; N=173,480),[15] Hispanic (Hispanic

236   Community Health Study and Study of Latinos (HCHS/SOL); N=12,502),[16] and African-

237   American participants (Continental Origins and Genetic Epidemiology [COGENT] Network;

238   N~16,500,[17] and the Candidate gene Association Resource (CARe) Project N=7,112),[18] and

239   exome genotyping in 130,273 multi-ethnic individuals.[19]

240   For loci classified as 'erythrocytic', we performed sequential conditional analyses to

241   determine the number of distinct signals in each region. The regions were defined based on

242   linkage disequilibrium (LD) plots. We used a Bonferroni correction for the number of SNVs in

243   the region with MAC $\geq$ 20 to define a signal as distinct. We also performed gene-based and

244   burden tests in Genesis using different selection of rare genetic variants (MAF $\leq$ 1%) based on

245   functional annotations (missense, high confidence loss of function or synonymous variants).

246   Results from each analysis were combined across ancestries or across sex by meta-

247   analysis using METAL.[20] The heterogeneity test between males and females was calculated

248   using the following formula: $\frac{(\beta males - \beta females)^2}{SE males^2 + SE females^2 - 2 \times r \times SE males \times SE females}$. This assessed the

249   difference in effect sizes between males (βmales) and females (βfemales) while accounting for

250   correlation (r) between male and female statistics due to relatedness and was calculated outside

251   of METAL. LD calculations in the TOPMed data and regional plots were done, within ancestry

252   group, using the Omics Analysis, Search and Information System (OASIS, see URL in the web

253   resources section). Functional annotations were performed using the WGS Annotator

254   (WGSA).[21]

255   We used the UK Biobank (UKBB, (see URL in the web resources section), a prospective cohort

256   study with deep genetic and phenotypic data collected on approximately 500,000 individuals

257   from across the United Kingdom, aged between 40 and 69 at recruitment, as an independent

258   sample for external replication of our findings. The centralized analysis of the genetic data,

259   including genotype quality, population structure and relatedness of the genetic data, and

260   efficient phasing and genotype imputation has been described extensively elsewhere.[22] Two

261   similar arrays were used for genotyping (Applied Biosystems UK Biobank Lung Exome Variant

262   Evaluation and UK Biobank Axiom Arrays) and pre-phasing was performed using markers

263   present on both arrays. Phasing on the autosomes was carried out using SHAPEIT3 and 1000

264   Genomes phase 3 panel to help with the phasing of non-European ancestry samples. Imputations

265   were carried out using the IMPUTE4 program with the Haplotype Reference Consortium (HRC)

266   reference panel or with a merged UK10K and 1000 Genomes phase 3 reference panel. For

267   chromosome X, haplotype estimation and genotype imputation were carried out separately on the

268   pseudo-autosomal and non-pseudo autosomal regions.

269   We identified UKBB African-ancestry participants using the following six self-reported

270   ethnicities: "Caribbean", "African", "Black or Black British", "Any other Black background",

271   "White and Black African" and "White and Black Caribbean". We excluded participants with

272   diabetes defined by the use of antidiabetic medication, self-reported physician diagnosis, FG ≥ 7

273   mmol/L or non-fasting glucose ≥ 11.1 mmol/L.

274   We generated principal components using smartPCA (see URL in the web resources

275   section)[23, 24] based on 72,300 common genetic variants in low LD selected with PLINK.[25] For

276   each PC, ethnic outliers lying more than 6SD away from the mean were excluded. The same

277   genetic variants were used to calculate an empirical kinship matrix using EPACTS (see URL

278   for EPACTS documentation in the web resources section). We used a linear mixed effect

279   model in R to evaluate the association of each variant with HbA1c, adjusting for age, sex, 10 PCs

280   and using the empirical kinship matrix.

281

Results

283   **HbA1c-associated regions using WGS**

284   We included in our analyses 10,338 TOPMed participants representing four ancestry

285   groups: Europeans (EA, N=6,158), African-Americans (AA, N=3,123), Hispanics/Latinos (HA,

286   N=650) and East Asians (AS, N=407; **Supplementary Table 1**). TOPMed studies were

287   composed of middle to older-aged participants of EA, AA, HA or AS ancestry with comparable

288   mean HbA1c, mean fasting glucose (FG) and mean hemoglobin (HB, **Supplementary Table 4**).

289   A total of 13,079,661 variants (EA), 21,443,543 variants (AA), 9,567,498 variants (HA), and

290   6,567,324 variants (AS) passed filters and were included in the analyses. QQ-plots and

291   Manhattan plots of WGS associations with HbA1c from ancestry-specific analyses and the meta-

292   analysis are provided in **Figure 1 and Supplementary Figures 1-2**. Using a significance

293   threshold of $P < 2 \times 10^{-8}$ to report an association as genome-wide significant, we detected five

294   regions associated with HbA1c, including one novel locus (low-frequency AA-specific variant,

295   rs1039215 in *HBG2* [MIM:142250] */HBE1* [MIM:142100], MAF=0.03), and a sixth locus

296   reaching suggestive evidence (rare AA-specific variant, rs551601853, near *XPNPEP1*

297   [MIM:602443], MAF=0.003, $P < 5 \times 10^{-7}$). Regional plots for the *HBG2/HBE1* and the

298   *XPNPEP1* loci are provided in **Supplementary Figures 3 and 4**. The four other single

299   nucleotide variants (SNVs) were located in regions previously identified as associated with

300   HbA1c in trans-ethnic meta-analyses[3]: rs2971670 in *GCK* on chromosome 7 ($P_{meta}=1.7 \times 10^{-9}$,

301 mainly associated in EA and AA, r^2=0.59, D'=1 with rs4607517, the index SNV in published

302 GWAS), rs17476364 in *HK1* [MIM:142600] on chromosome 10 (P_{meta}=3.1×10^{-21}, mainly

303 associated in EA and HA, r^2=0.10, D'=1 with rs10823343, r^2=0.18, D'=0.50 with rs4745982,

304 index SNVs in published GWAS), rs113373052 in *FN3K* [MIM:608425] */FN3KRP*

305 [MIM:611683] on chromosome 17 (P_{meta}=4.5×10^{-10}, associated in EA, r^2=0.92, D'=0.99, with

306 rs1046896, index SNV in published GWAS) and rs1050828 in *G6PD* on chromosome X

307 (P_{meta}=5.1×10^{-210}, associated in AA and HA, monomorphic in the other ancestries)[3]. The top

308 SNV in each region detected at the genome-wide threshold ($P < 2 \times 10^{-8}$) is indicated in **Table 1**

309 and at the sub-genome-wide threshold ($P < 5 \times 10^{-7}$) in **Supplementary Table 5**.

310

311 **Classification of HbA1c-associated loci by their biological pathways**

312 Because loci that influence HbA1c through erythrocytic pathways are expected to cause

313 nonglycemic variation in HbA1c, we sought to classify HbA1c-associated loci as 'glycemic' or

314 'erythrocytic' using mediation analyses and association analyses with glycemic and erythrocytic

315 traits. Mediation analyses and look-up in WGS analysis of FG in TOPMed classified the *GCK*

316 [MIM:138079] variants as glycemic (**Supplementary Tables 6, 7 & 8**). Association analyses

317 with erythrocytic traits in published GWAS showed that rs17476364-C allele, in the *HK1* gene,

318 was positively associated with HB, mean corpuscular volume (MCV), mean corpuscular

319 hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), red blood cell count

320 (RBC), hematocrit (HCT) and red blood cell distribution width (RDW) in EA[15] and that

321 rs1050828-T allele, in the *G6PD* gene, was positively associated with MCH in AA and MCV in

322 AA and HA and negatively associated with HCT and HB in AA, and RBC and RDW in AA and

323 HA[16-18] (**Supplementary Tables 6, 8 and 9**). Results of the mediation and association analyses

14

324   are available in **Supplementary Tables 6, 8 and 9** for the genome-wide variants and

325   **Supplementary Tables 7 and 10** for the sub-genome variants. Among the 26 variants meeting

326   the sub-genome-wide threshold, four were common and 22 were low-frequency or rare. Two of

327   the common HbA1c-associated variants were reported previously:[3] the glycemic variant at

328   *G6PC2* [MIM:608058] and the erythrocytic variant at *TMPRSS6* [MIM:609862]. Among the

329   genome-wide significant loci, *HK1* (EA) and *G6PD* (AA and HA) were classified as

330   'erythrocytic', as was the low frequency variant rs1039215 in *HGB2/HBE1*, negatively

331   associated with HB, HCT, MCV and MCH. The significant *FN3K/FN3KRP* and *XPNPEP1*

332   variants were unclassified.

333

334   **Characterization of distinct signals at erythrocytic HbA1c-associated loci**

335   For loci classified as 'erythrocytic' (*HK1*, *HBG2/HBE1* and *G6PD*), we performed

336   sequential conditional analyses to detect distinct signals in addition to the top SNV. The regions

337   were defined as +/- 60kb region around *HK1*, +/- 250kb region around *HBG2/HBE1* and +/-

338   500kb region around *G6PD*, based on Linkage Disequilibrium (LD) plots (**Supplementary**

339   **Figures 3, 5 and 6**). We did not detect secondary associations in the *HK1* and *HBG2/HBE1*

340   regions at a threshold of 4.3×10^{-5} and 8.5×10^{-6} respectively. Interestingly, in the *G6PD* region,

341   the top SNV (rs1050828, p.Val98Met, T-allele frequency 12% in AA, 2% in HA, 0% in EA and

342   AS) was associated with lower HbA1c in both AA and HA. This SNV accounted for 23% of

343   HbA1c variance in AA and 4% in HA. The LD in the *G6PD* region is complex, with a strong

344   haplotype effect in AA and HA (**Figure 2**). By performing conditional analyses on the top SNV

345   (rs1050828) we were able to detect an additional rare signal (G-allele frequency 0.5%) in AA

346   (rs76723693, p.Leu353Pro, $B_{cond}=-0.50$, $P_{cond}=2.8 \times 10^{-15}$) that was distinct from rs1050828

15

347 (r2=0.0006, D'=1, **Figure 3**). The threshold to detect an additional signal was fixed to 1.7×10^{-5}

348 in AA and 4.5×10^{-5} in HA. The three other SNVs located in the same region that were sub-

349 genome-wide significant in the main analysis (rs143745197, rs184539426 and rs189305788)

350 were not significant after adjusting for both rs1050828 and rs76723693. We detected significant

351 or suggestive associations using gene-based ($P=9.7 \times 10^{-11}$) and burden tests ($P=4.7 \times 10^{-5}$) when

352 aggregating 15 missense rare variants with MAF $\leq$ 1% in *G6PD* (**Supplementary Table 11**). As

353 association from SKAT was more significant than the association of rs76723693 from single-

354 SNV association analysis, it suggested that several rare missense variants in *G6PD* were

355 associated with HbA1c. We performed additional association and conditional analyses in *G6PD*

356 for missense variants with 10 < minor allele count < 20. In addition to rs76723693, one rare

357 missense variant (rs5030872, MAF=0.002) was suggestively associated with HbA1c (B=-0.64,

358 $P=3.4 \times 10^{-6}$) and became significantly associated with HbA1c when adjusting for both rs1050828

359 and rs76723693 ($B_{cond}=-0.71$, $P_{cond}=1.3 \times 10^{-9}$).

360   The *HBG2/HBE1* new variant rs1039215 ($P=2.0 \times 10^{-9}$) is located at 327 kb from the SCT

361 variant rs334 ($P=2.9 \times 10^{-8}$). A recent paper reported rs1039215 to be in LD ($r^2>0.2$) with the

362 SCT variant (rs334, $r^2=0.24$, D'=-0.83) and correlated with *HBG2* gene expression in whole

363 blood.[26] To determine if rs1039215 was distinct from rs334, we calculated LD and performed

364 conditional analyses. In TOPMed, rs1039215 was in modest LD with rs334 ($r^2=0.30$, D'=0.88,

365 **Supplementary Figure 3**) and conditioning on rs334 attenuated its association (decrease in p-

366 value from 2.0×10^{-9} to 10^{-3}, **Supplementary Table 12**), suggesting that the association of

367 rs1039215 with HbA1c may be explained by the SCT rs334 variant. Nevertheless, the haplotype

368 rs334-A / rs1039215-G was more significantly associated with HbA1c than the haplotypes

369 rs334-A / rs1039215-A and rs334-T / rs1039215-G (**Supplementary Table 13**). Thus, the

370 association of the new variant rs1039215 in *HBG2/HBE1* could be partially distinct from the

371 known association at the SCT variant. In our AA sample, 12 individuals carried two copies of

372 rs1050828-T allele and at least one copy of rs1039215-G allele and had a mean HbA1c of 4.60

373 (0.41) *versus* 5.65 (0.39) for the 2,372 individuals that carried no risk allele at the two variants

374 (**Supplementary Table 14**).

375

376 **HbA1c-lowering effects of *G6PD* variants in AA and HA**

377 In sex-stratified analyses in AA, the *G6PD* rs76723693 variant was associated with a

378 decrease in HbA1c of 0.98%-units (95% CI 0.51–1.44) per allele in hemizygous men and 0.46%-

379 units (95% CI 0.26–0.66) in heterozygous women; whereas, the *G6PD* rs1050828 variant was

380 associated with a decrease in HbA1c of 0.88%-units (95% CI 0.81–0.95) per allele in

381 hemizygous men and 0.34%-units (95% CI 0.30–0.39) in heterozygous women (**Table 2 and**

382 **Figure 4**). We detected heterogeneity in effect sizes between males and females for rs1050828

383 (P_{het}=0.008) but not for rs76723693 (P_{het}=0.19). In HA, the *G6PD* rs1050828 variant was

384 associated with a decrease in HbA1c of 0.84%-units (95% CI 0.47–1.22) per allele in

385 hemizygous men and 0.25%-units (95% CI 0.06–0.45) in heterozygous women (**Table 3 and**

386 **Figure 4**). Upon adjusting on the sickle cell trait (SCT) variant, rs334, the association of these

387 two SNVs did not attenuate (**Supplementary Table 12**).

388

389 **Population reclassification of diabetes diagnosis by *G6PD* variants**

390 To assess the potential public health impact of *G6PD* variants in AA, we designed a

391 hypothetical scenario, using National Health and Nutrition Examination Survey (NHANES, see

392 URL in the web resources section) 2015-2016, to estimate the number of African Americans

393   and Hispanics whose diagnosis of diabetes would have been missed if screened with a single

394   HbA1c measurement using the 6.5% diagnostic threshold and if the effects of the two *G6PD*

395   variants rs1050828 and rs76723693 were not accounted for. We restricted the NHANES analytic

396   sample to adults aged ≥ 18 years who self-identified as Non-Hispanic Black or Mexican

397   American/Other Hispanic respectively. In non-Hispanic Blacks (N=1,227), an estimated 2.32%

398   with HbA1c < 6.5%-units may be considered to have diabetes upon accounting for the effect size

399   and observed allele frequency of the common *G6PD* variant, rs1050828. An additional 0.13%

400   with HbA1c < 6.5%-units may be considered to have diabetes upon accounting for the effect of

401   the rare *G6PD* variant, rs76723693. In Mexican American/Other Hispanic (N=1,768), an

402   additional 0.26% with HbA1c < 6.5%-units may be considered to have diabetes due to the effect

403   of rs1050828. According to the 2016 United States Census Bureau, approximately 30.14 and

404   39.33 million adults identified themselves as African American and Hispanic adults respectively,

405   suggesting that 740,000 African American adults, of which 40,000 would be attributed to the

406   rare variant, rs76723693, and 100,000 Hispanic adults with diabetes would remain undiagnosed

407   when screened by a single HbA1c measurement if this genetic information was not taken into

408   account (**Supplementary Table 15**).

409

410   **Replication of results in the UKBB**

411   We selected a total of 5,964 African-ancestry non-diabetic UKBB participants for

412   replication. In this sample, 2,461 participants were males, mean age (SD) was 51.1 (7.7), and

413   mean HbA1c (SD) was 5.50 (0.44). The four variants of interest (rs334 (SCT), rs1039215

414   (HBG2), rs76723693 and rs1050828 (G6PD)) had good quality of imputation (info score > 0.50,

415   **Supplementary Table 16**). We observed significant and consistent associations of the two

18

416   *G6PD* variants with HbA1c (rs1050828-T, B=-0.36, P<2.2x10^{-16} and rs76723693-G, B=-0.21,

417   P=0.007, **Supplementary Table 16**). The association of the rare *G6PD* rs76723693-G variant

418   was even more significant when both *G6PD* variants were in the model (B=-0.27, P=6.7×10^{-5}).

419   We did not detect significant associations of *HBG2* and *SCT* variants with HbA1c in the UKBB.

420

Discussion

422   Through deep large-scale WGS association analysis in 10,338 individuals from four

423   different ancestries and fine-mapping of HbA1c-associated loci, we identified common, low

424   frequency and rare genetic variants that influence nonglycemic variation in HbA1c. We

425   confirmed four known regions associated with HbA1c (the glycemic *GCK*, erythrocytic *HK1* and

426   *G6PD*, and unclassified *FN3K/FN3KRP* regions) and discovered two new AA-specific low

427   frequency or rare erythrocytic variants in *G6PD* (rs76723693) and *HBG2/HBE1* (rs1039215).

428   The magnitude and direction of effect of the association of rs76723693 and rs1039215 with

429   HbA1c was similar in the two largest TOPMed AA cohorts (JHS and MESA, **Supplementary**

430   **Table 17**). The association of rs76723693 with HbA1c was replicated in UKBB African-

431   ancestry participants. We detected significant association using gene-based test when

432   aggregating rare missense variants in *G6PD*, indicating that others rare missense variants in

433   *G6PD*, in addition to rs76723693, are associated with HbA1c. We showed that the association of

434   the *HBG2/HBE1* variant with HbA1c was partially distinct from the known SCT rs334 variant in

435   *HBB* [MIM:141900] suggesting that genetic variation at hemoglobin genes other than *HBB* may

436   also influence HbA1c. Individuals carrying both the *G6PD* variant (rs1050828) and the

437   *HBG2/HBE1* variant (rs1039215) had even lower HbA1c than those carrying only one of the two

438   variants. Hemizygous males and homozygous females for the *G6PD* rs1050828-T allele who

439   carried one or more copies of the *HBG2* rs1039215-G allele had a mean HbA1c that was 1.05%-

440   unit lower than those carrying none of these alleles.

441   The novel HbA1c variant in AA, rs1039215 on chromosome 11, lies in an intron of the

442   hemoglobin subunit gamma 2 gene *HBG2* and the hemoglobin subunit epsilon 1 gene *HBE1*.

443   *HBG2,* in addition to *HBG1,* encodes the gamma chain of hemoglobin, which combines with 2

444   alpha chains to form fetal hemoglobin. Fetal hemoglobin is known to interfere with the

445   measurement of HbA1c by some laboratory assays,[27-29] and so persistence of fetal hemoglobin

446   may be a mechanism through which rs1039215 influences HbA1c measurements. While a small

447   degree of analytic interference with SCT has been reported for the Tosoh 2.2 and G7 assays used

448   by JHS, MESA, and ARIC to measure HbA1c (see NGSP website URL in the web resources

449   section),[30, 31] no interference has been reported for the Bio-Rad Variant II Turbo assay used by

450   UKB[32] (**Supplementary Table 2**). Thus, assay interference may explain the lack of association

451   of rs334 with HbA1c in UKBB. Alternatively, as our haplotype analysis indicated that the

452   haplotype containing both rs334-A in *HBB* and rs1039215-G in *HBG2* was likely the main driver

453   of this signal in the region, the true causal variant could be another variant within this haplotype

454   that partially explains the reported association of rs334 with HbA1c.

455   Both variants identified in *G6PD* in AA are missense and pathogenic variants for G6PD

456   deficiency (ClinVar), an X-linked genetic disorder characterized by a defect glucose-6-phosphate

457   dehydrogenase enzymatic action, causing erythrocytes to break down prematurely. Thus, the

458   mechanism through which these *G6PD* variants lower HbA1c is most likely through shortening

459   of erythrocytic lifespan (**Supplementary Table 18**). Assuming random X-inactivation, we

460   would expect the difference between hemizygous males and no rs1050828-T variant to be two

461   times larger than the difference between heterozygous females and no rs1050828-T variant;

462 however, in this study, we observed a slightly lower than expected effect in females. We posit

463 that this heterogeneity of genetic effects by sex may be due to X-linked dosage compensation

464 and not actual sex differences.[33] While these *G6PD* variants lower HbA1c in AA, similar to

465 sickle cell trait,[5] we note that the lowering effect of these genetic variants on HbA1c do not

466 explain the higher mean HbA1c in AA compared to European American with similar glycemia

467 that has been reported in observational studies.[34, 35]

468 We found that rs1050828, the common *G6PD* variant in AA, was also associated with

469 HbA1c in HA (sample composed of 51% Mexican, 14% Dominican, 14% Puerto Rican, 11%

470 South American, and 10% others) suggesting that carriers of this variant are likely to be

471 underdiagnosed if a single HbA1c measurement is used to screen two of the largest racial/ethnic

472 minorities in the United States for diabetes. If *G6PD* variants were ignored, we estimate that

473 approximately 840,000 African American and Hispanic adults with diabetes would remain

474 undiagnosed. As Hispanic populations are ancestrally diverse including African Ancestry (~7%

475 of HA in our study have > 50% AA ancestry and ~25% have > 15% AA ancestry), *G6PD*

476 variants were expected to occur at some lower frequency in HA compared to AA. Given that

477 there is no universal screening for G6PD deficiency in North America, asymptomatic carriers are

478 unlikely to know their G6PD status.[7] While misclassification of diabetes status at the population

479 level was largely driven by the common G6PD variant, rs10105828, and not the rare rs76723693

480 variant (0.5% of the AA population), in this era of precision medicine there is a growing need for

481 clinical practice to account for the impact of rare genetic variants with clinically meaningful

482 effects on routinely performed diagnostic tests even if the proportion of carriers in the population

483 is small. We also predict that other *G6PD* deficiency variants, particularly common in African,

484 Mediterranean and Asian populations,[36] will impact HbA1c's utility as a diagnostic measure.

21

485   Until genetic information is universally accessible to adjust diagnostic thresholds of imperfect

486   biomarkers, our findings support current clinical practice guidelines by the American Diabetes

487   Association which recommends the use of both FG and HbA1c in combination for diabetes

488   diagnosis, especially given that the diagnostic level of HbA1c is conservative relative to FG.[37]

489   Strengths of this investigation include the analysis of extremely high-quality deep

490   sequence data from a large, multi-ancestry sample including five studies and four ancestries to

491   investigate the association of genetic variation across the full allelic spectrum with HbA1c, a

492   diagnostic test that is central to the care of patients with diabetes. As previous GWAS on HbA1c

493   had included a smaller sample of Hispanic ancestry, this study represents one of the first large-

494   scale genetic discovery efforts for HbA1c that include multiple ethnic minorities in the United

495   States. Our study demonstrates that both common and rare genetic variation differ across

496   ancestries and informs the unique genetic architecture of HbA1c. We acknowledge limitations.

497   Our hypothetical scenario using NHANES data to assess the population impact of *G6PD*

498   genetics on diabetes screening does not take into account the full complement of clinical and

499   genetic information contributing to nonglycemic variation in HbA1c, some of which may raise,

500   and not lower, HbA1c reducing the risk of under-detection.[38] For instance, structural variants

501   (e.g., copy number variation) were not included in our analyses. Nevertheless, our findings

502   highlight the potential for underdiagnosis in carriers of these variants, which disproportionately

503   affect certain ethnic minorities, if HbA1c is used as the only diagnostic test for diabetes. Our

504   sample sizes, and thus our power, are limited compared to GWAS, particularly for Hispanic and

505   Asian populations. For instance, we did not detect an association signal for rs76723693 in HA

506   because only one individual in our HA population carried a copy of the rare variant. Including

507   larger samples of individuals of non-European ancestries will improve the characterization of

22

508   rare variants with clinically meaningful effects on HbA1c. We acknowledge that the genome-

509   wide threshold used in this paper may not represent the true number of independent genetic

510   variants (SNVs as well as insertion-deletion variants) tested using our sequence data.

511   In this study we identified common and rare genetic variants that cause nonglycemic

512   differences in HbA1c with important clinical and public health implications. Because HbA1c is

513   commonly used in diabetes screening and management worldwide, disregarding the effects of

514   these rare and common genetic variants that tend to occur in high diabetes-risk minority groups,

515   could result in delayed diagnosis or under-detection of diabetes in carriers of these variants. To

516   avoid such disparities in care, an assessment of these variants should be considered for

517   incorporation into precision medicine approaches for diabetes diagnosis.

518

Supplementary Data

520   Supplemental data contains cohort descriptions, 18 tables and 6 figures.

521

Acknowledgements

26

608   **The Multi-Ethnic Study of Atherosclerosis**

27

28

Web Resources

Analysis Commons: http://analysiscommons.com

EPACTS: https://genome.sph.umich.edu/wiki/EPACTS

HbA1c Assay Interferences information on the National Glycohemoglobin Standardization

Program (NGSP) website: http://www.ngsp.org/interf.asp

National Health and Nutrition Examination Survey (NHANES):

https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?BeginYear=2015

OASIS: https://omicsoasis.github.io

SmartPCA: http://www.hsph.harvard.edu/alkes-price/software

TOPMed Pipeline: https://github.com/UW-GAC/analysis_pipeline

TOPMed website: www.nhlbiwgs.org

Details regarding TOPMed laboratory methods, data processing and quality control:

https://www.nhlbiwgs.org/methods

UK Biobank: https://www.ukbiobank.ac.uk/

Funding

668   HMH was supported by NHLBI training grants T32 HL007055 and T32 HL129982, and

669   American Diabetes Association Grant #1-19-PDF-045.

670   **Declarations of Interest**

671   The authors have no conflict of interest to declare.

672   The views expressed in this manuscript are those of the authors and do not necessarily represent

673   the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or

674   the U.S. Department of Health and Human Services.

675

References

677   1. Mortensen, H.B., Christophersen, C. (1983). Glucosylation of human haemoglobin a in red

678   blood cells studied in vitro. Kinetics of the formation and dissociation of haemoglobin A1c. Clin.

679   Chim. Acta *134*, 317-326.

680   2. Leong, A., Meigs, J.B. (2015). Type 2 Diabetes Prevention: Implications of Hemoglobin A1c

681   Genetics. Rev. Diabet. Stud. *12*, 351-362.

682   3. Wheeler, E., Leong, A., Liu, C.T., Hivert, M.F., Strawbridge, R.J., Podmore, C., Li, M., Yao,

683   J., Sim, X., Hong, J. et al. (2017). Impact of common genetic determinants of Hemoglobin A1c

684   on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-

685   wide meta-analysis. PLoS Med. *14*, e1002383.

686   4. International HapMap Consortium, Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A.,

687   Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P. et al. (2007). A second

688   generation human haplotype map of over 3.1 million SNPs. Nature *449*, 851-861.

689   5. Lacy, M.E., Wellenius, G.A., Sumner, A.E., Correa, A., Carnethon, M.R., Liem, R.I., Wilson,

690   J.G., Sacks, D.B., Jacobs, D.R.,Jr, Carson, A.P. et al. (2017). Association of Sickle Cell Trait

691   With Hemoglobin A1c in African Americans. JAMA *317*, 507-515.

692   6. Strickland, S.W., Campbell, S.T., Little, R.R., Bruns, D.E., Bazydlo, L.A.L. (2018).

693   Recognition of rare hemoglobin variants by hemoglobin A1c measurement procedures. Clin.

694   Chim. Acta *476*, 67-74.

695   7. Leong, A. (2007). Is there a need for neonatal screening of glucose-6-phosphate

696   dehydrogenase deficiency in Canada? Mcgill J. Med. *10*, 31-34.

697   8. Motulsky, A.G., Stamatoyannopoulos, G. (1966). Clinical implications of glucose-6-

698   phosphate dehydrogenase deficiency. Ann. Intern. Med. *65*, 1329-1334.

699   9. Paterson, A.D. (2017). HbA1c for type 2 diabetes diagnosis in Africans and African

700   Americans: Personalized medicine NOW! PLoS Med. *14*, e1002384.

701   10. Little, R.R., Rohlfing, C.L., Sacks, D.B., National Glycohemoglobin Standardization

702   Program (NGSP) Steering Committee. (2011). Status of hemoglobin A1c measurement and goals

703   for improvement: from chaos to order for improving diabetes care. Clin. Chem. *57*, 205-214.

704   11. Regier, A.A., Farjoun, Y., Larson, D.E., Krasheninina, O., Kang, H.M., Howrigan, D.P.,

705   Chen, B.J., Kher, M., Banks, E., Ames, D.C. et al. (2018). Functional equivalence of genome

706   sequencing analysis pipelines enables harmonized variant calling across human genetics projects.

707   Nat. Commun. *9*, 4038-018-06159-4.

708   12. Brody, J.A., Morrison, A.C., Bis, J.C., O'Connell, J.R., Brown, M.R., Huffman, J.E., Ames,

709   D.C., Carroll, A., Conomos, M.P., Gabriel, S. et al. (2017). Analysis commons, a team approach

710   to discovery in a big-data environment for genetic epidemiology. Nat. Genet. *49*, 1560-1563.

711   13. Kanai, M., Tanaka, T., Okada, Y. (2016). Empirical estimation of genome-wide significance

712   thresholds based on the 1000 Genomes Project data set. J. Hum. Genet. *61*, 861-866.

713   14. Xu, C., Tachmazidou, I., Walter, K., Ciampi, A., Zeggini, E., Greenwood, C.M., UK10K

714   Consortium. (2014). Estimating genome-wide significance for whole-genome sequencing

715   studies. Genet. Epidemiol. *38*, 281-290.

716   15. Astle, W.J., Elding, H., Jiang, T., Allen, D., Ruklisa, D., Mann, A.L., Mead, D., Bouman, H.,

717   Riveros-Mckay, F., Kostadima, M.A. et al. (2016). The Allelic Landscape of Human Blood Cell

718   Trait Variation and Links to Common Complex Disease. Cell *167*, 1415-1429.e19.

719   16. Hodonsky, C.J., Jain, D., Schick, U.M., Morrison, J.V., Brown, L., McHugh, C.P.,

720   Schurmann, C., Chen, D.D., Liu, Y.M., Auer, P.L. et al. (2017). Genome-wide association study

721   of red blood cell traits in Hispanics/Latinos: The Hispanic Community Health Study/Study of

722   Latinos. PLoS Genet. *13*, e1006760.

723   17. Chen, Z., Tang, H., Qayyum, R., Schick, U.M., Nalls, M.A., Handsaker, R., Li, J., Lu, Y.,

724   Yanek, L.R., Keating, B. et al. (2013). Genome-wide association analysis of red blood cell traits

725   in African Americans: the COGENT Network. Hum. Mol. Genet. *22*, 2529-2538.

726   18. Lo, K.S., Wilson, J.G., Lange, L.A., Folsom, A.R., Galarneau, G., Ganesh, S.K., Grant, S.F.,

727   Keating, B.J., McCarroll, S.A., Mohler, E.R.,3rd et al. (2011). Genetic association analysis

728   highlights new loci that modulate hematological trait variation in Caucasians and African

729   Americans. Hum. Genet. *129*, 307-317.

730   19. Chami, N., Chen, M.H., Slater, A.J., Eicher, J.D., Evangelou, E., Tajuddin, S.M., Love-

731   Gregory, L., Kacprowski, T., Schick, U.M., Nomura, A. et al. (2016). Exome Genotyping

732   Identifies Pleiotropic Variants Associated with Red Blood Cell Traits. Am. J. Hum. Genet. *99*, 8-

733   21.

734   20. Willer, C.J., Li, Y., Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of

735   genomewide association scans. Bioinformatics *26*, 2190-2191.

736   21. Liu, X., White, S., Peng, B., Johnson, A.D., Brody, J.A., Li, A.H., Huang, Z., Carroll, A.,

737   Wei, P., Gibbs, R. et al. (2016). WGSA: an annotation pipeline for human genome sequencing

738   studies. J. Med. Genet. *53*, 111-112.

739   22. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A.,

740   Vukcevic, D., Delaneau, O., O'Connell, J. et al. (2018). The UK Biobank resource with deep

741   phenotyping and genomic data. Nature *562*, 203-209.

742   23. Patterson, N., Price, A.L., Reich, D. (2006). Population structure and eigenanalysis. PLoS

743   Genet. *2*, e190.

744   24. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., Reich, D. (2006).

745   Principal components analysis corrects for stratification in genome-wide association studies. Nat.

746   Genet. *38*, 904-909.

747   25. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J.,

748   Sklar, P., de Bakker, P.I., Daly, M.J. et al. (2007). PLINK: a tool set for whole-genome

749   association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559-575.

750   26. Shriner, D., Rotimi, C.N. (2018). Whole-Genome-Sequence-Based Haplotypes Reveal Single

751   Origin of the Sickle Allele during the Holocene Wet Phase. Am. J. Hum. Genet. *102*, 547-556.

752   27. Little, R.R., Roberts, W.L. (2009). A review of variant hemoglobins interfering with

753   hemoglobin A1c measurement. J. Diabetes Sci. Technol. *3*, 446-451.

754   28. Rohlfing, C.L., Connolly, S.M., England, J.D., Hanson, S.E., Moellering, C.M., Bachelder,

755   J.R., Little, R.R. (2008). The effect of elevated fetal hemoglobin on hemoglobin A1c results: five

756   common hemoglobin A1c methods compared with the IFCC reference method. Am. J. Clin.

757   Pathol. *129*, 811-814.

758   29. Sarnowski, C., Hivert, M.F. (2018). Impact of Genetic Determinants of HbA1c on Type 2

759   Diabetes Risk and Diagnosis. Curr. Diab Rep. *18*, 52-018-1022-4.

760   30. Roberts, W.L., Safar-Pour, S., De, B.K., Rohlfing, C.L., Weykamp, C.W., Little, R.R.

761   (2005). Effects of hemoglobin C and S traits on glycohemoglobin measurements by eleven

762   methods. Clin. Chem. *51*, 776-778.

763   31. Rohlfing, C., Hanson, S., Little, R.R. (2017). Measurement of Hemoglobin A1c in Patients

764   With Sickle Cell Trait. JAMA *317*, 2237.

765   32. Mongia, S.K., Little, R.R., Rohlfing, C.L., Hanson, S., Roberts, R.F., Owen, W.E., D'Costa,

766   M.A., Reyes, C.A., Luzzi, V.I., Roberts, W.L. (2008). Effects of hemoglobin C and S traits on

767   the results of 14 commercial glycated hemoglobin assays. Am. J. Clin. Pathol. *130*, 136-140.

768   33. Sidorenko, J., Kassam, I., Kemper, K., Zeng, J., Lloyd-Jones, L., Montgomery, G.W.,

769   Gibson, G., Metspalu, A., Esko, T., Yang, J. et al. (2018). The effect of X-linked dosage

770   compensation on complex trait variation. bioRxiv.

771   34. Bergenstal, R.M., Gal, R.L., Connor, C.G., Gubitosi-Klug, R., Kruger, D., Olson, B.A.,

772   Willi, S.M., Aleppo, G., Weinstock, R.S., Wood, J. et al. (2017). Racial Differences in the

773   Relationship of Glucose Concentrations and Hemoglobin A1c Levels. Ann. Intern. Med. *167*,

774   95-102.

775   35. Ziemer, D.C., Kolm, P., Weintraub, W.S., Vaccarino, V., Rhee, M.K., Twombly, J.G.,

776   Narayan, K.M., Koch, D.D., Phillips, L.S. (2010). Glucose-independent, black-white differences

777   in hemoglobin A1c levels: a cross-sectional analysis of 2 studies. Ann. Intern. Med. *152*, 770-

778   777.

779   36. Gomez-Manzo, S., Marcial-Quino, J., Vanoye-Carlo, A., Serrano-Posada, H., Ortega-

780   Cuellar, D., Gonzalez-Valdez, A., Castillo-Rodriguez, R.A., Hernandez-Ochoa, B., Sierra-

781   Palacios, E., Rodriguez-Bustamante, E. et al. (2016). Glucose-6-Phosphate Dehydrogenase:

782   Update and Analysis of New Mutations around the World. Int. J. Mol. Sci. *17*,

783   10.3390/ijms17122069.

784   37. American Diabetes Association. (2019). 2. Classification and Diagnosis of Diabetes:

785   Standards of Medical Care in Diabetes-2019. Diabetes Care *42*, S13-S28.

786   38. Jun, G., Sedlazeck, F.J., Chen, H., Yu, B., Qi, Q., Krasheninina, O., Carroll, A., Liu, X.,

787   Mansfield, A., Zarate, S. et al. (2018). **PgmNr 3186/W: Identification of novel structural**

788   **variations affecting common and complex disease risks with >16,000 whole genome**

789   **sequences from ARIC and HCHS/SOL**.

790

791

Figure Legends:

792

793 **Figure 1:** Manhattan-plots of the meta-analysis of HbA1c in non-diabetic individuals in

794 TOPMed cohorts. The –log10(P-value) for each single nucleotide variant on the y-axis is plotted

795 against the build 38 genomic position on the x-axis (chromosomal coordinate). The dashed

796 horizontal line indicates the genome-wide significance threshold of $P = 2\times10^{-8}$. The y-axis was
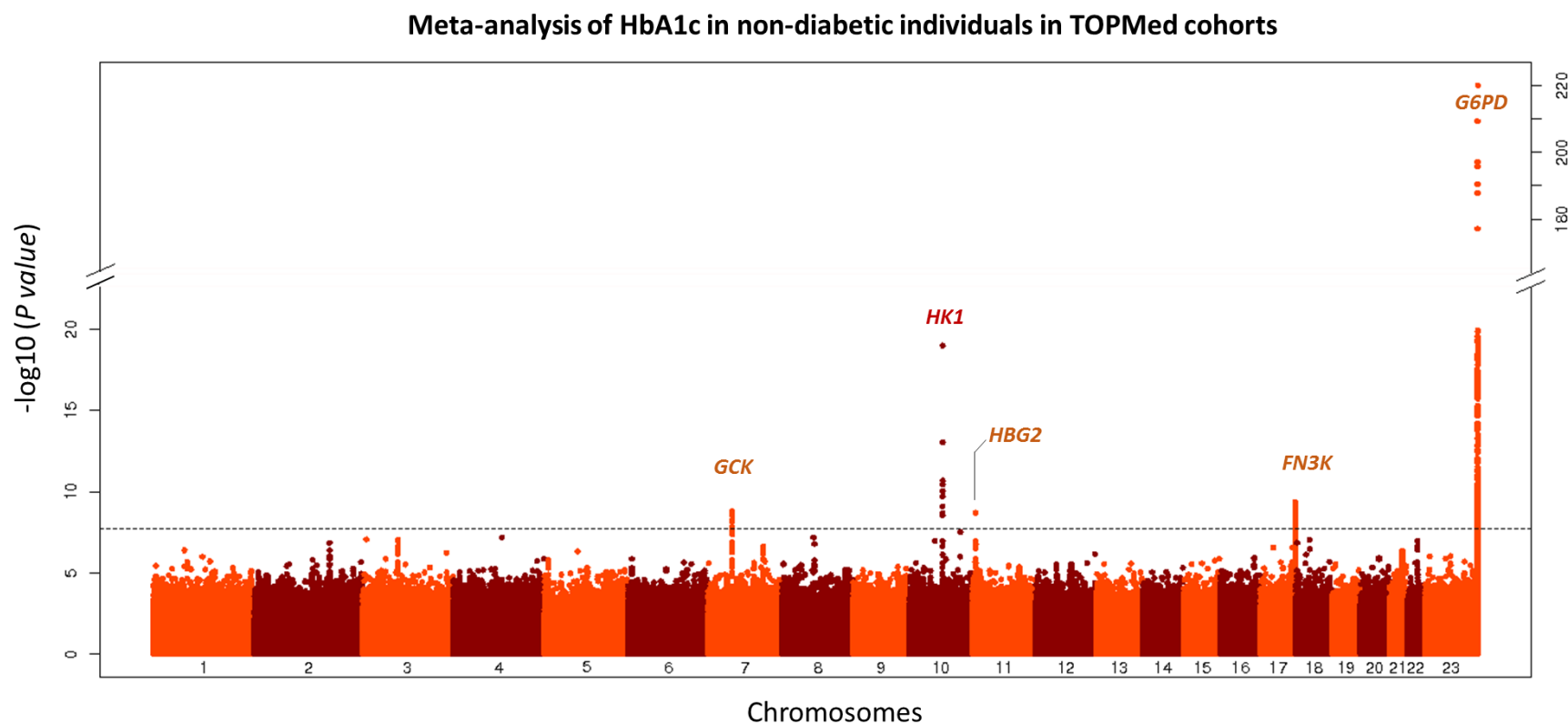
797 truncated for ease of interpretation.

798 **Figure 2:** Regional HbA1c association plots in the *G6PD* region in non-diabetic African-

799 Americans and Hispanics in TOPMed cohorts. Association plots are displayed separately in

800 African-Americans (left-side) and in Hispanics (right-side). Single nucleotide variants are plotted

801 with their P-values (-log10 values, y-axis) as a function of build 38 genomic position on

802 chromosome X (x-axis). The local linkage disequilibrium (LD) structure (r2 - top or D' - bottom)

803 between the top associated single nucleotide variant rs1050828 (red triangle) and correlated

804 proxies is indicated in color according to a blue to red scale from r2/D'=0 to 1 and was

805 calculated separately in non-diabetic African-Americans and Hispanics TOPMed cohorts.

806 **Figure 3:** Sequential conditional analyses in the *G6PD* region in non-diabetic African-

807 Americans and Hispanics in TOPMed cohorts. Association plots are displayed separately in

808 African-Americans (left-side) and in Hispanics (right-side). Single nucleotide variants are plotted

809 with their P-values (-log10 values, y-axis) as a function of build 38 genomic position on

810 chromosome X (x-axis). The local linkage disequilibrium (LD) structure (r2 - top or D' - bottom)

811 between the top associated single nucleotide variant (red triangle) and correlated proxies is

812 indicated in color according to a blue to red scale from r2/D'=0 to 1 and was calculated

813 separately in non-diabetic African-Americans and Hispanics TOPMed cohorts. Significance

814 thresholds to declare an association signal as distinct are indicated by a red line ($P=1.7\times10^{-5}$ in

815   African-Americans and P=4.5×10^{-5} in Hispanics, based on the number of single nucleotide

816   variants in the region with a minor allele count greater than 20).

817   **Figure 4:** Mean HbA1c levels according to genotypes at both *G6PD* variants (rs76723693 and

818   rs1050828) and stratified by sex. Plots are displayed separately in African-Americans (left and

819   center) and in Hispanics (right). Due to the limited sample size of African-Americans women

820   carrying two copies of the rare G allele at rs76723693, they are not represented on this plot. In

821   AA, the *G6PD* rs76723693 variant was associated with a decrease in HbA1c of 0.98%-units

822   (95% Confidence Interval (CI) 0.51–1.44) per allele in hemizygous men and 0.46%-units (95%

823   CI 0.26–0.66) in heterozygous women; whereas, the *G6PD* rs1050828 variant was associated

824   with a decrease in HbA1c of 0.88%-units (95% CI 0.81–0.95) per allele in hemizygous men and

825   0.34%-units (95% CI 0.30–0.39) in heterozygous women In HA, the *G6PD* rs1050828 variant

826   was associated with a decrease in HbA1c of 0.84%-units (95% CI 0.47–1.22) per allele in

827   hemizygous men and 0.25%-units (95% CI 0.06–0.45) in heterozygous women.

828 **Figure 1:** Manhattan-plots of the meta-analysis of HbA1c in non-diabetic individuals in TOPMed cohorts.


Meta-analysis of HbA1c in non-diabetic individuals in TOPMed cohorts

829

830 The –log10(P-value) for each single nucleotide variant on the y-axis is plotted against the build 38 genomic position on the x-axis

831 (chromosomal coordinate). The dashed horizontal line indicates the genome-wide significance threshold of $P = 2{\times}10^{-8}$. The y-axis

832 was truncated for ease of interpretation.

833

834 **Figure 2:** Regional HbA1c association plots in the *G6PD* region in non-diabetic African-Americans and Hispanics in TOPMed
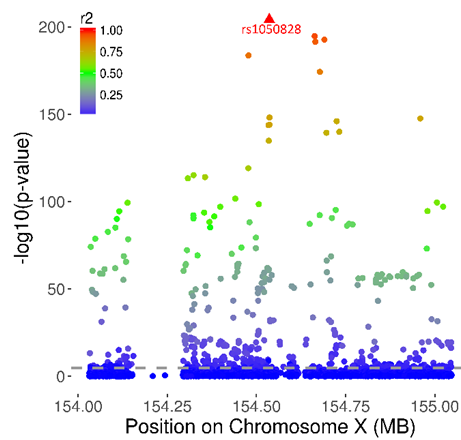
835 cohorts
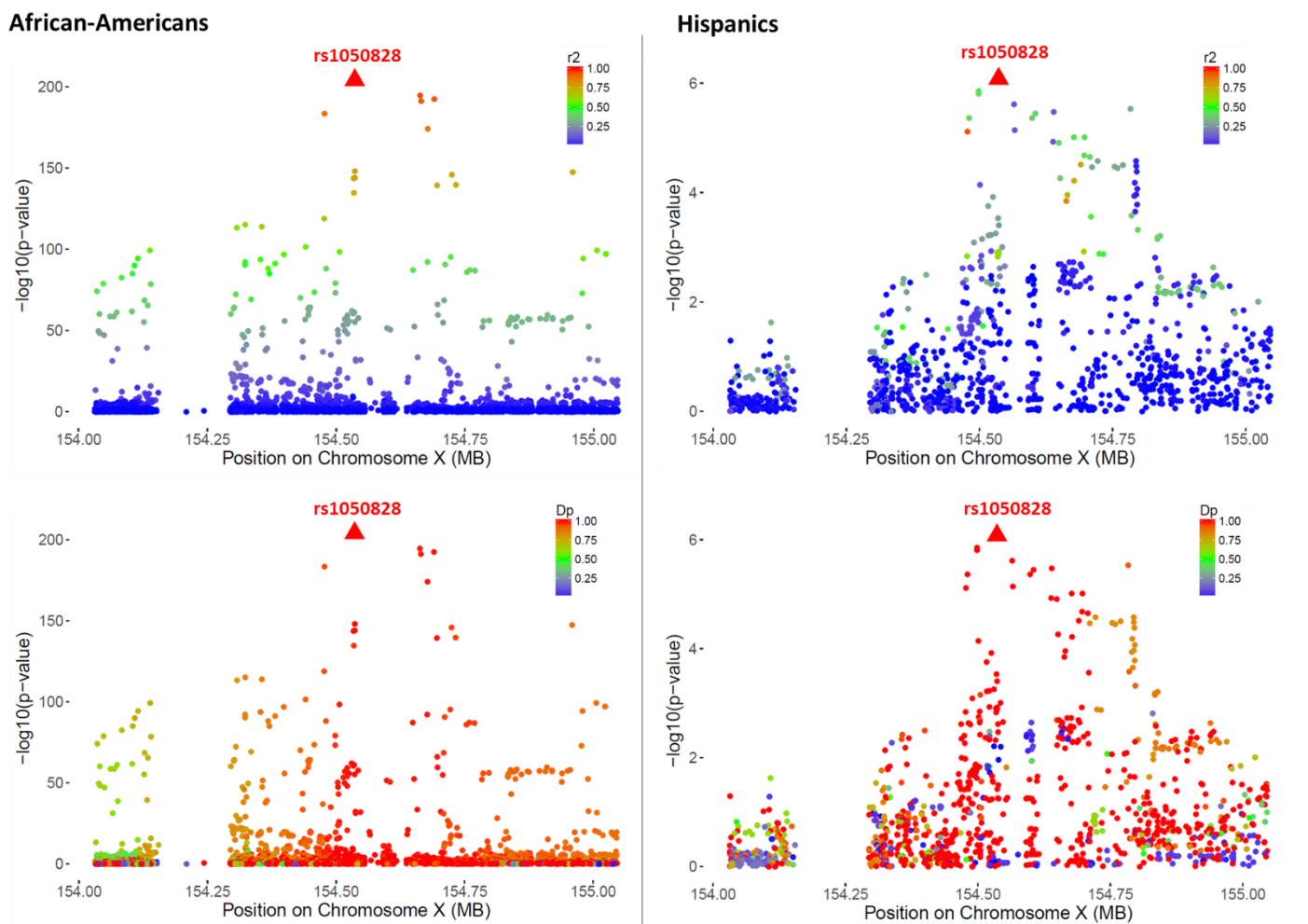
837   Association plots are displayed separately in African-Americans (left-side) and in Hispanics (right-side). Single nucleotide variants are

838   plotted with their P-values (-log10 values, y-axis) as a function of build 38 genomic position on chromosome X (x-axis). The local

839   linkage disequilibrium (LD) structure (r2 - top or D' - bottom) between the top associated single nucleotide variant rs1050828 (red

840   triangle) and correlated proxies is indicated in color according to a blue to red scale from r2/D'=0 to 1 and was calculated separately in

841   non-diabetic African-Americans and Hispanics TOPMed cohorts.

842 **Figure 3:** Sequential conditional analyses in the *G6PD* region in non-diabetic African-

843 Americans and Hispanics in TOPMed cohorts

844
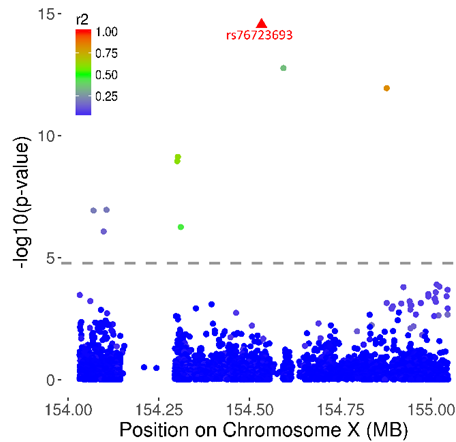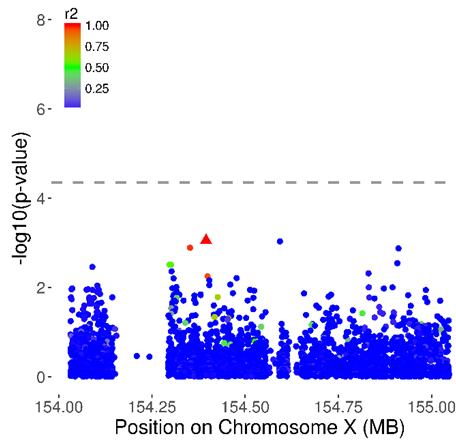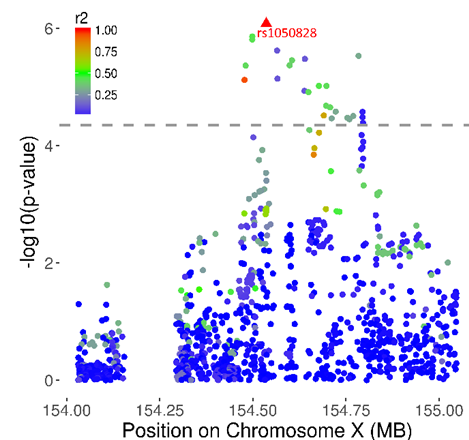
845   Association plots are displayed separately in African-Americans (left-side) and in Hispanics

846   (right-side). Single nucleotide variants are plotted with their P-values (-log10 values, y-axis) as a

847   function of build 38 genomic position on chromosome X (x-axis). The local linkage

848   disequilibrium (LD) structure (r2 - top or D' - bottom) between the top associated single

849   nucleotide variant (red triangle) and correlated proxies is indicated in color according to a blue to

850   red scale from r2/D'=0 to 1 and was calculated separately in non-diabetic African-Americans
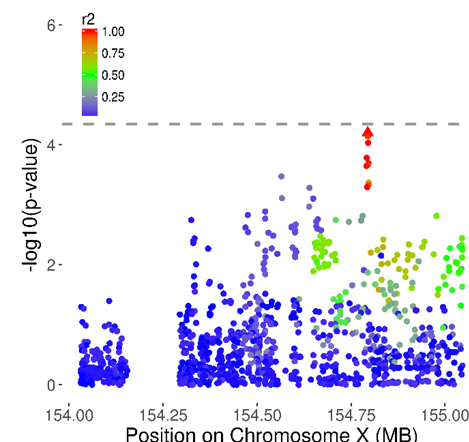
851   and Hispanics TOPMed cohorts. Significance thresholds to declare an association signal as

852   distinct are indicated by a red line (P=1.7$\times$10^{-5} in African-Americans and P=4.5$\times$10^{-5} in

853   Hispanics, based on the number of single nucleotide variants in the region with a minor allele

854   count greater than 20).

Figure 4: Mean HbA1c levels according to genotypes at both G6PD variants (rs76723693 and rs1050828) and stratified by sex
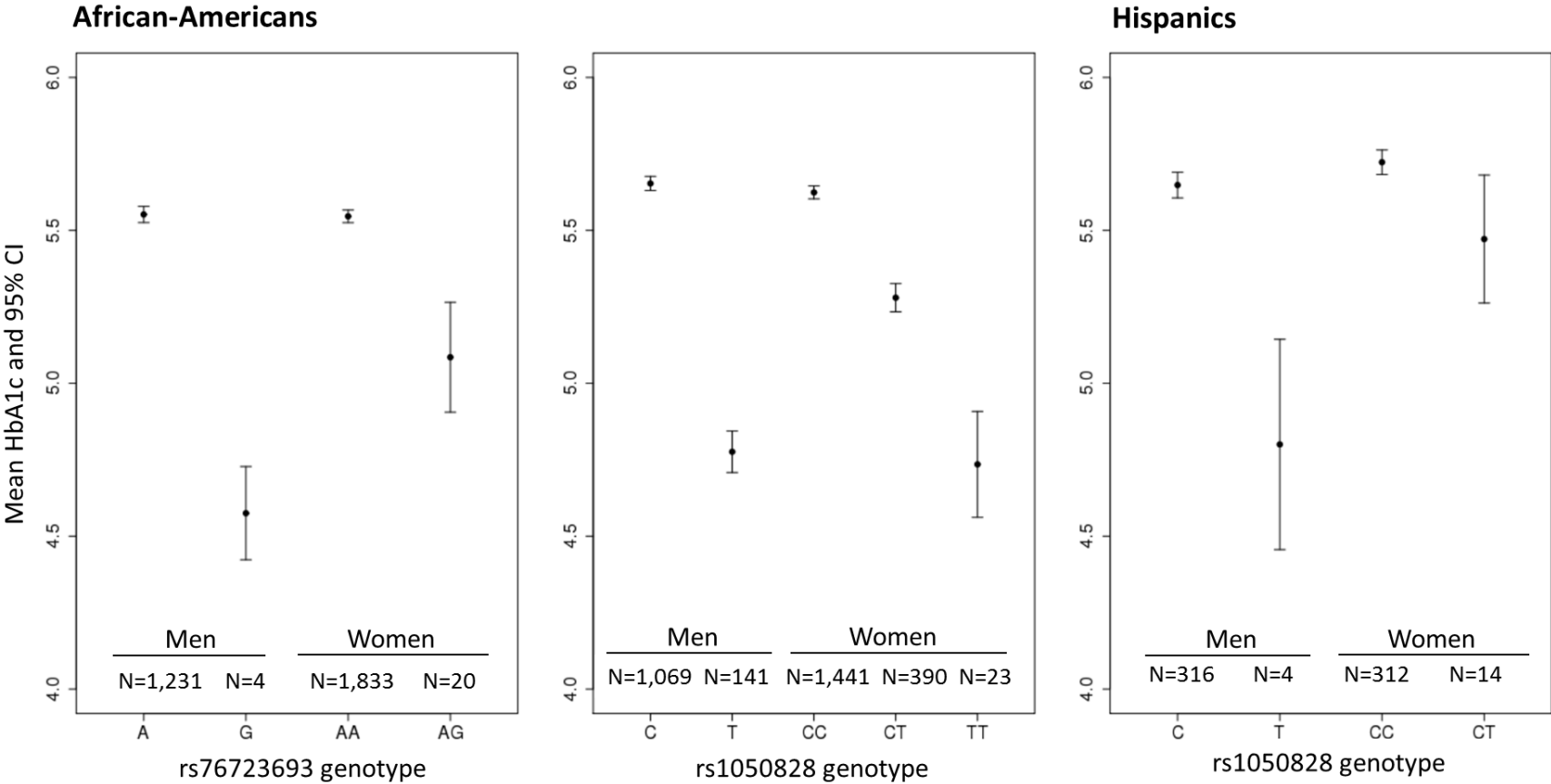
857   Plots are displayed separately in African-Americans (left and center) and in Hispanics (right). Due to the limited sample size of

858   African-Americans women carrying two copies of the rare G allele at rs76723693, they are not represented on this plot.

859   In AA, the *G6PD* rs76723693 variant was associated with a decrease in HbA1c of 0.98%-units (95% Confidence Interval (CI) 0.51–

860   1.44) per allele in hemizygous men and 0.46%-units (95% CI 0.26–0.66) in heterozygous women; whereas, the *G6PD* rs1050828

861   variant was associated with a decrease in HbA1c of 0.88%-units (95% CI 0.81–0.95) per allele in hemizygous men and 0.34%-units

862   (95% CI 0.30–0.39) in heterozygous women In HA, the *G6PD* rs1050828 variant was associated with a decrease in HbA1c of 0.84%-

863   units (95% CI 0.47–1.22) per allele in hemizygous men and 0.25%-units (95% CI 0.06–0.45) in heterozygous women.

864 **Table 1:** Most associated single nucleotide variants in the regions detected at the genome-wide level ($P<2\times10^{-8}$) by ancestry and in the

865 meta-analysis of HbA1c in non-diabetic individuals in TOPMed cohorts

MarkerID[b]	rsID	Closest Gene	Ref/Alt Allele	Meta-analysis			Europeans (N=6,158)				African-Americans (N=3,123)				Hispanics (N=650)				Asians (N=407)			
				Beta	SE	*P value*	MAF[a]	Beta	SE	*P value*	MAF[a]	Beta	SE	*P value*	MAF[a]	Beta	SE	*P value*	MAF[a]	Beta	SE	*P value*
7:44186502	rs2971670	*GCK*	C/T	0.04	0.01	1.7E-09	0.18	0.03	0.01	1.6E-05	0.18	0.06	0.01	1.6E-04	0.21	0.05	0.03	0.04	0.19	0.04	0.03	0.20
10:69334748	rs17476364	*HK1*	T/C	-0.09	0.01	3.1E-21	0.10	-0.08	0.01	3.2E-18	0.02	-0.09	0.04	0.04	0.06	-0.15	0.04	5.4E-04				
11:5464062	rs1039215	*HBG2/ HBE1*	A/G	-0.21	0.04	2.0E-09					0.03	-0.21	0.04	2.0E-09								
17:82739582	rs113373052	*FN3K*	C/T	0.03	0.01	4.5E-10	0.32	0.03	0.01	4.0E-08	0.29	0.02	0.01	0.06	0.39	0.03	0.02	0.15	0.49	0.05	0.02	0.04
X:154536002	rs1050828	*G6PD*	C/T	-0.41	0.01	5.1E-210					0.12	-0.41	0.01	8.4E-205	0.02	-0.37	0.07	8.3E-07				

866

867 [a]MAF: minor allele frequency

868 [b]MarkerID is defined as Chromosome:Position with positions on Build 38.

869

870 **Table 2:** Combined and sex-stratified conditional association analyses of HbA1c for the two distinct single nucleotide variants in the

871 G6PD region in non-diabetic African-Americans in TOPMed cohorts

MarkerID[c]	rsID	Ref/Alt Allele	MAF[a]	Pooled analysis (N=3,123)			Conditional analysis on rs1050828			Males (N=1,269)			Females (N=1,854)			Heterogeneity P value[b]
				Beta	SE	*P value*	Beta	SE	*P value*	Beta	SE	*P value*	Beta	SE	*P value*	
X:154533025	rs76723693	A/G	0.005	-0.44	0.07	2.2×10^{-9}	-0.50	0.06	2.8×10^{-15}	-0.52	0.12	6.8×10^{-6}	-0.33	0.10	6.2×10^{-4}	0.19
X:154536002	rs1050828	C/T	0.12	-0.41	0.01	8.4×10^{-205}	-	-	-	-0.44	0.02	2.3×10^{-148}	-0.37	0.02	2.7×10^{-69}	0.008

872 [a]MAF: Minor allele frequency

873 [b]Heterogeneity P value: test of difference in effect sizes between males and females, accounting for the correlation between males and

874 females statistics

875 [c]MarkerID is defined as Chromosome:Position. Positions are indicated on Build 38.

876

Table 3: Analysis of HbA1c associations with rs1050828, X:154536002, alternate allele T, in non-diabetic African-Americans and Hispanics in TOPMed cohorts

	African-Americans						Hispanics					
	N	MAF[a]	Beta	SE	*P value*	% of variance explained	N	MAF[a]	Beta	SE	*P value*	% of variance explained
All	3,123	0.12	-0.41	0.01	8.4E-205	0.23	650	0.02	-0.37	0.07	8.3E-07	0.04
Males	1,269	0.12	-0.44	0.02	2.3E-148	0.35	323	0.01	-0.45	0.11	3.5E-05	0.05
Females	1,854	0.12	-0.37	0.02	2.7E-69	0.14	327	0.02	-0.28	0.1	4.5E-03	0.02

[a]MAF: Minor allele frequency