The American Journal of Human Genetics, Volume *105*

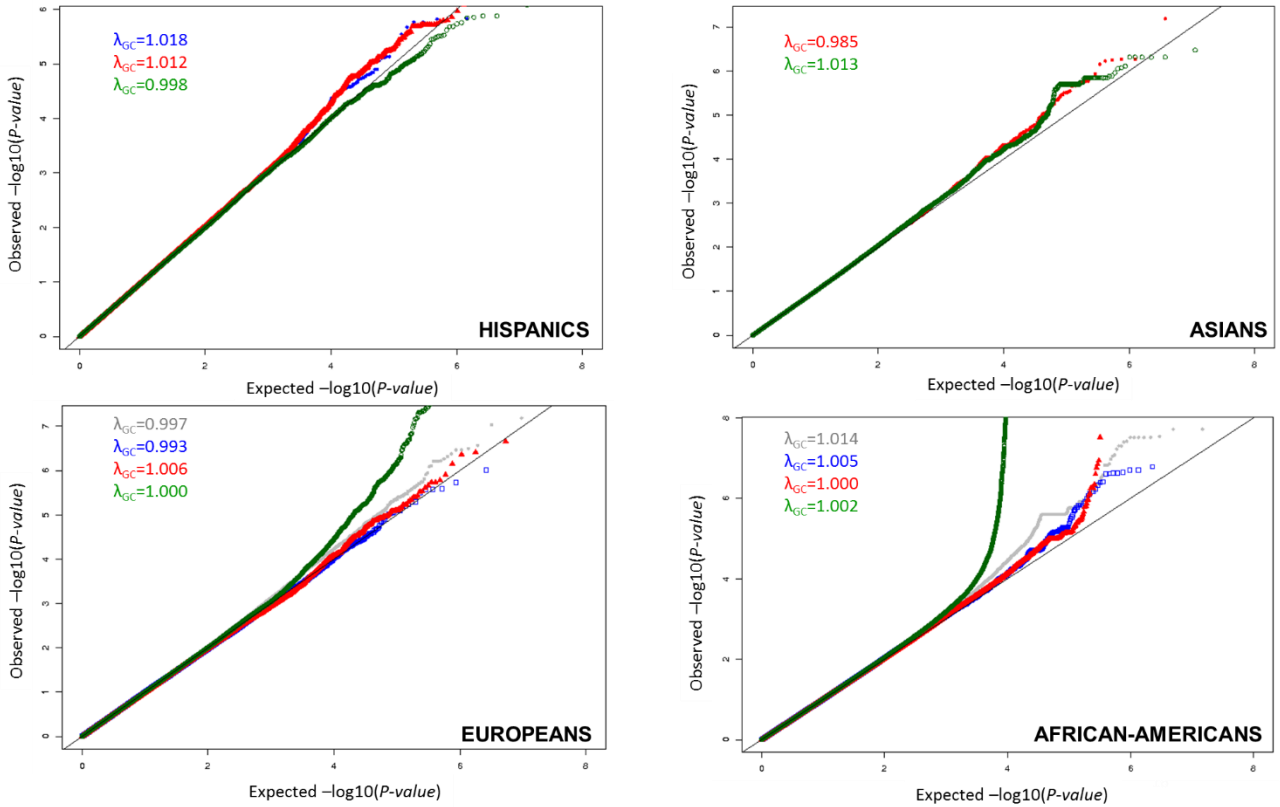
Supplemental Data

Impact of Rare and Common Genetic Variants on Diabetes Diagnosis by Hemoglobin A1c in Multi-Ancestry Cohorts: The Trans-Omics for Precision Medicine Program

Chloé Sarnowski, Aaron Leong, Laura M. Raffield, Peitao Wu, Paul S. de Vries, Daniel DiCorpo, Xiuqing Guo, Huichun Xu, Yongmei Liu, Xiuwen Zheng, Yao Hu, Jennifer A. Brody, Mark O. Goodarzi, Bertha A. Hidalgo, Heather M. Highland, Deepti Jain, Ching-Ti Liu, Rakhi P. Naik, Jeffrey R. O'Connell, James A. Perry, Bianca C. Porneala, Elizabeth Selvin, Jennifer Wessel, Bruce M. Psaty, Joanne E. Curran, Juan M. Peralta, John Blangero, Charles Kooperberg, Rasika Mathias, Andrew D. Johnson, Alexander P. Reiner, Braxton D. Mitchell, L. Adrienne Cupples, Ramachandran S. Vasan, Adolfo Correa, Alanna C. Morrison, Eric Boerwinkle, Jerome I. Rotter, Stephen S. Rich, Alisa K. Manning, Josée Dupuis, James B. Meigs, TOPMed Diabetes Working Group, TOPMed Hematology Working Group, TOPMed Hemostasis Working Group, and National Heart, Lung, and Blood Institute TOPMed Consortium

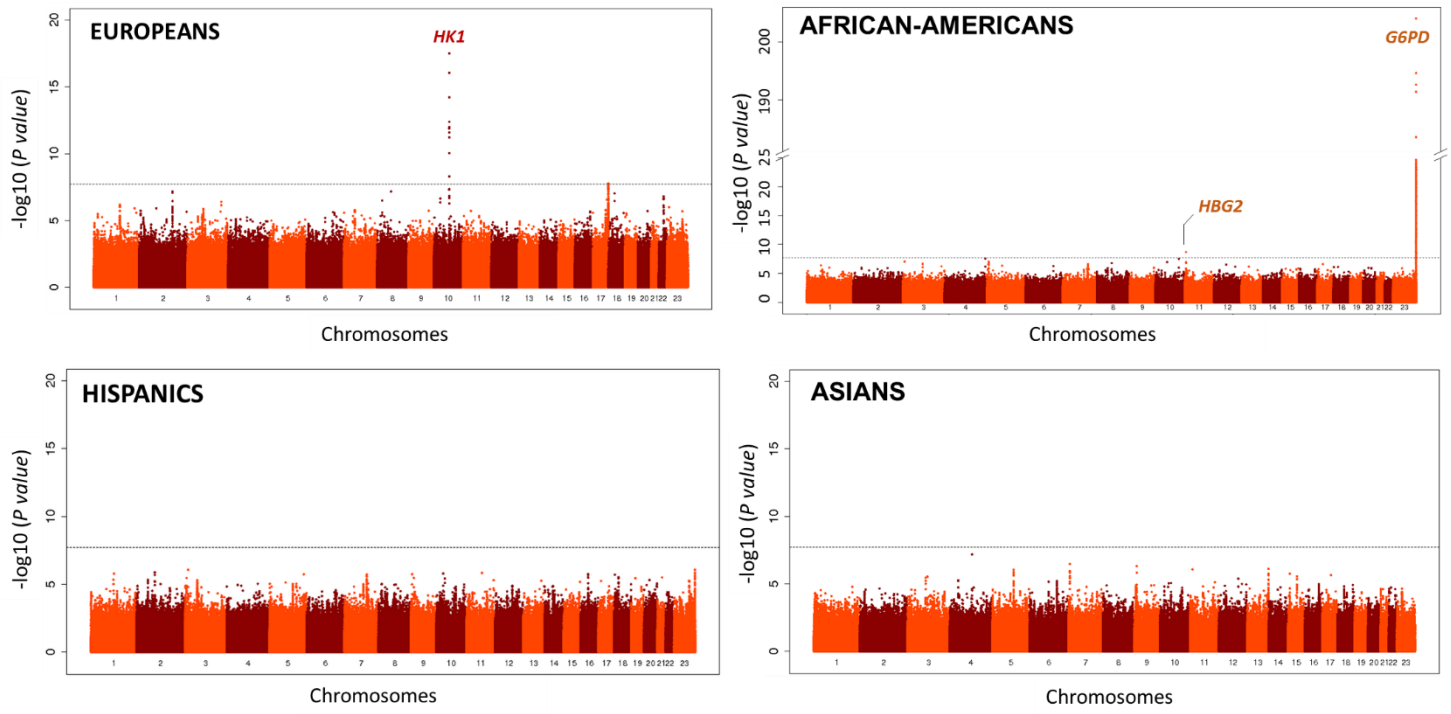
Supplementary Figures

Figure S1: Quantile-Quantile (QQ)-plots by ancestry of the association analysis of HbA1c in non-diabetic individuals in TOPMed cohorts stratified on minor allele frequency (MAF)



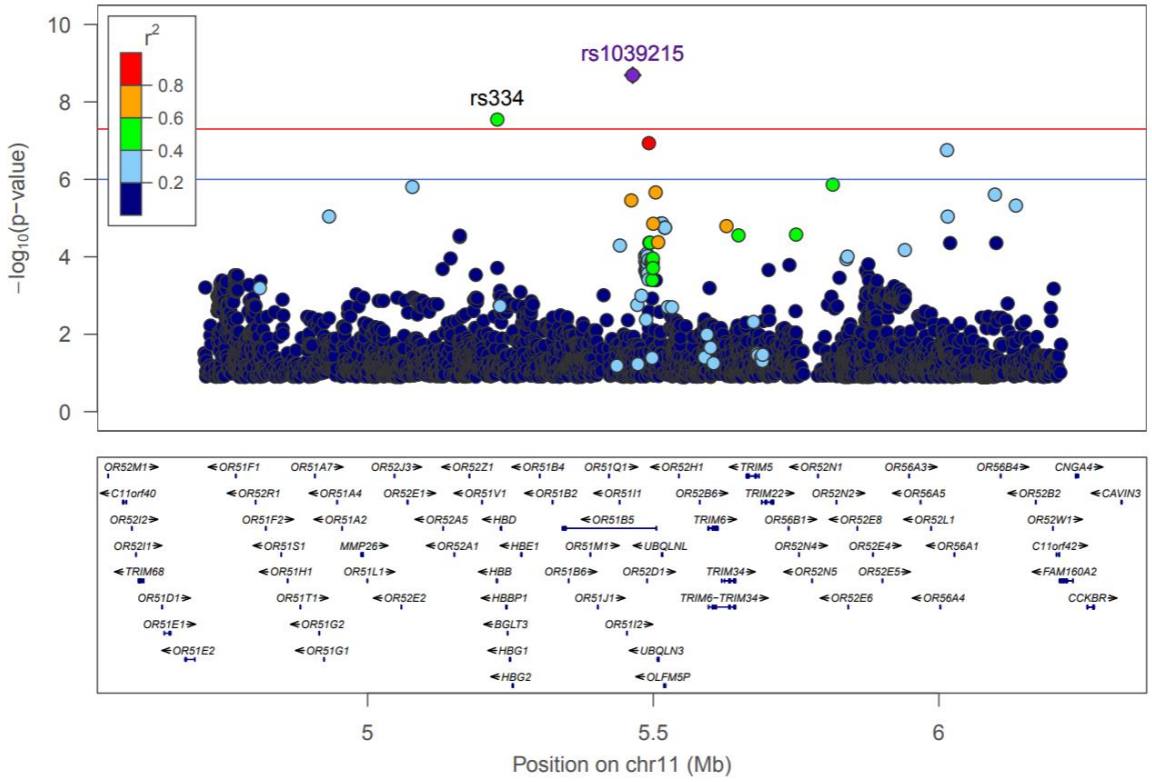
The dots represent the distribution of observed ordered –log10(*P-values*) against the theoretical model distribution of expected ordered –log10(*P-values*). The solid black line represents the theoretical model distribution of expected –log10(*P-values*) under the null distribution. The genomic inflation factor (λ_{GC}) is defined as the ratio of the median of the empirically observed distribution of the test statistic to the expected median, thus quantifying the extent of the inflation and the excess false positive rate. Association results were stratified by minor allele frequency (MAF) of single nucleotide variants: MAF ≥ 5% are indicated in green; 5% > MAF ≥ 1% are indicated in red; 1% > MAF ≥ 0.5% are indicated in blue; 0.5% > MAF ≥ 0.1% are indicated in grey.

Figure S2: Manhattan-plots by ancestry of the association analysis of HbA1c in non-diabetic individuals in TOPMed cohorts.



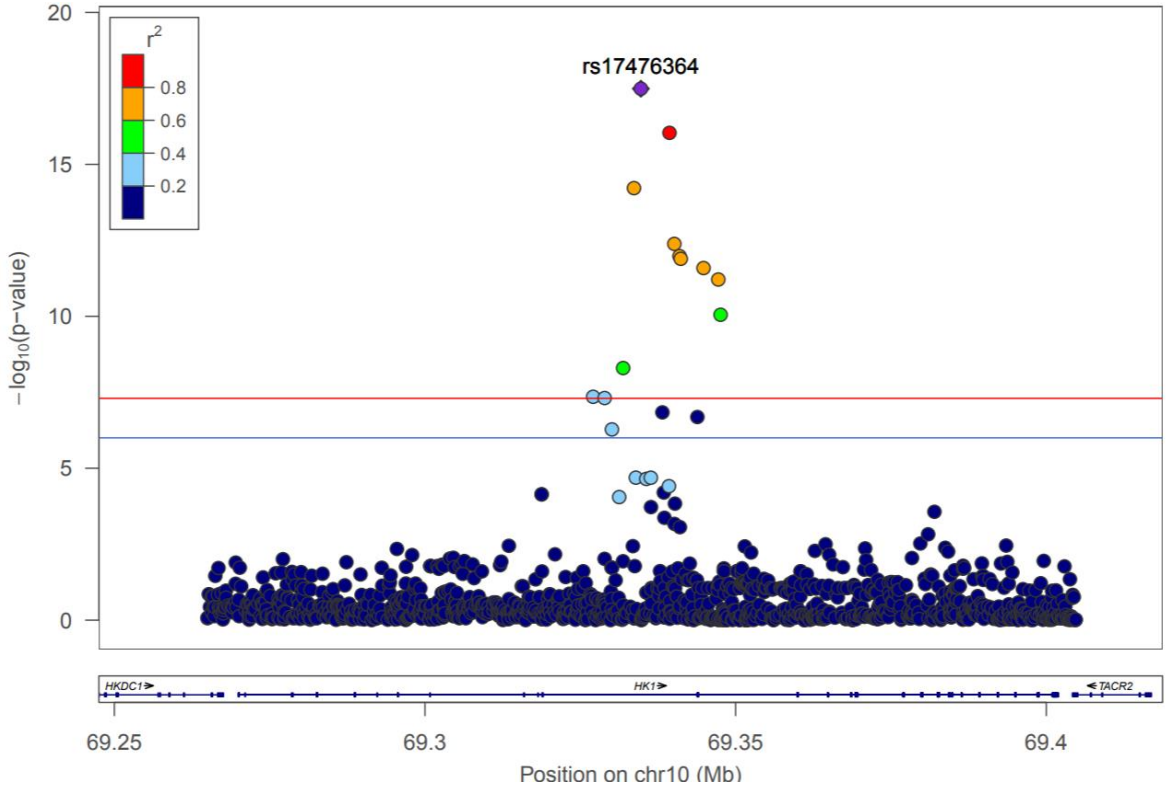
The –log10(*P-value*) for each single nucleotide variant on the y-axis is plotted against the build 38 genomic position on the x-axis (chromosomal coordinate). The dashed horizontal line indicates the genome-wide significance threshold of $P = 2\times10^{-8}$. The y-axis was truncated for ease of interpretation.

Figure S3: Regional HbA1c association plot in the *HBG2/HBE1* region (+/- 750kb around *HBG2*).



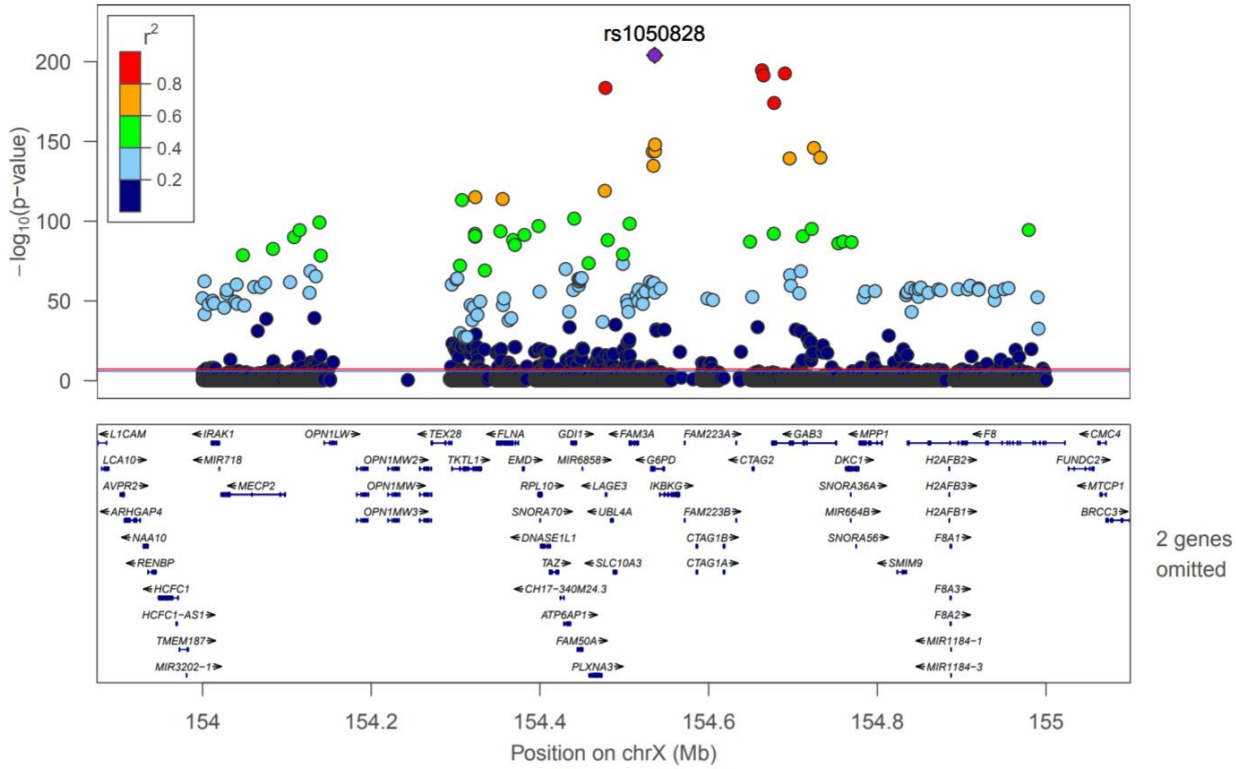
Single nucleotide variants are plotted with their P-values (-log10 values, left y-axis) as a function of build 38 genomic position on chromosome 11 (x-axis). Estimated recombination rates (right y-axis) are plotted to reflect the local linkage disequilibrium (LD) structure around the top associated single nucleotide variant rs1039215 (purple diamond) and correlated proxies (according to a blue to red scale from r2=0 to 1). LD was calculated in non-diabetic African ancestry TOPMed cohorts.

Figure S4: Regional HbA1c association plot in the *HK1* region (+/- 70kb around *HK1*).



Single nucleotide variants are plotted with their P-values (-log10 values, left y-axis) as a function of build 38 genomic position on chromosome 10 (x-axis). Estimated recombination rates (right y-axis) are plotted to reflect the local linkage disequilibrium (LD) structure around the top associated single nucleotide variant rs17476364 (purple diamond) and correlated proxies (according to a blue to red scale from r2=0 to 1). LD was calculated in non-diabetic European ancestry TOPMed cohorts.

Figure S5: Regional HbA1c association plot in the *G6PD* region (+/- 500kb around *G6PD*).



Single nucleotide variants are plotted with their P-values (-log10 values, left y-axis) as a function of build 38 genomic position on chromosome X (x-axis). Estimated recombination rates (right y-axis) are plotted to reflect the local linkage disequilibrium (LD) structure around the top associated single nucleotide variant rs1050828 (purple diamond) and correlated proxies (according to a blue to red scale from r2=0 to 1). LD was calculated in non-diabetic African ancestry TOPMed cohorts.

Table S1: Counts of non-diabetic individuals by TOPMed cohorts and ancestry groups included in HbA1c association analyses

	Ancestry				
Cohort	European	African-American	Hispanic	Asian	**Totals**
Amish	151	--	--	--	**151**
Atherosclerosis Risk in Communities Study (ARIC)	2,315	100	--	--	**2,415**
Framingham Heart Study (FHS)	2,236	--	--	--	**2,236**
Jackson Heart Study (JHS)	--	2,356	--	--	**2,356**
Multi-Ethnic Study of Atherosclerosis (MESA)	1,456	667	650	407	**3,180**
Totals	**6,158**	**3,123**	**650**	**407**	**10,338**

Table S2: Assays used to measure HbA1c by TOPMed cohorts included in HbA1c association analyses and in UK Biobank

Cohort	Assay	Interference[a]	
		HbS (SCT)	**Elevated HbF**
Amish	Quest Diagnostics using Immunoturbidimetry (Roche Cobas Intregra)	Higher	?
Atherosclerosis Risk in Communities Study (ARIC)	2003-04: Tosoh 2.2 Plus HPLC instrument	Slightly lower*	?
	2007-08: Tosoh G7 instrument	Slightly lower*	Yes if >30%
Framingham Heart Study (FHS)	Roche Hitachi 911, Roche Diagnostics, HbA1c - turbidimetric immunoassay	No**	?
Jackson Heart Study (JHS)	Tosoh 2.2 HPLC system (Tosoh Corporation, Tokyo, Japan) at University of Minnesota	Slightly lower*	?
Multi-Ethnic Study of Atherosclerosis (MESA)	Exam 2: Tosoh 2.2 Plus HPLC instrument	Slightly lower*	?
	Exam 5: Tosoh G7 instrument	Slightly lower*	Yes if >30%
UK Biobank	Bio-Rad Variant II Turbo	No**	Yes if >5%

[a] Slightly lower*: Some assay inference has been reported with HbS but not >±7% at 6 or 9% HbA1c; No**: Assay inference has not been reported to be >±7% at 6 or 9% HbA1c. ?: Not reported

Table S3: Number of TOPMed individuals, per study, included in the TOPMed red blood cell traits association analyses

Cohort	HB[a]	HCT[b]	MCH[c]	MCHC[d]	MCV[e]	RBC[f]	RDW[g]
Amish	1,021	1,021	1,021	1,021	1,021	1,021	0
Atherosclerosis Risk in Communities Study (ARIC)	3,475	3,475	1,778	1,779	3,315	1,778	1,073
Cardiovascular Health Study (CHS)	69	69	0	0	0	0	0
Framingham Heart Study (FHS)	2,826	2,826	2,826	2,826	2,826	2,826	2,080
Genetic Study of Atherosclerosis Risk (GeneSTAR)	1,585	1,585	1,506	1,506	1,506	1,506	0
Jackson Heart Study (JHS)	2,981	2,981	2,675	2,675	2,674	2,675	2,674
Multi-Ethnic Study of Atherosclerosis (MESA)	2,240	2,240	2,240	2,240	2,240	2,240	0
San Antonio Family Study (SAFS)	971	971	968	968	965	970	0
Women's Health Initiative (WHI)	9,912	9,909	1,286	1,286	1,286	1,286	1,285
Total	25,080	25,077	14,300	14,301	15,833	14,302	7,112

[a]HB: Hemoglobin
[b]MCV: Mean Corpuscular Volume
[c]MCH: Mean Corpuscular Hemoglobin
[d]MCHC: Mean Corpuscular Hemoglobin Concentration
[e]HCT: Hematocrit
[f]RBC: Red Blood Cell count
[g]RDW: Red cell Distribution Width

Table S4: Characteristics of non-diabetic individuals by TOPMed cohorts and ancestry groups included in HbA1c association analyses

Cohort	N	Men, N (%)	Age mean years (SD)	HbA1c mean % (SD)	Fasting Glucose (FG) mean mmol/L (SD)	Hemoglobin (Hb) mean mg/L (SD)
Amish European ancestry	151	69 (45.7)	45.1 (15.8)	4.9 (0.4)	5.0 (0.5)	--
ARIC European ancestry	2,315	1,117 (48.3)	57.5 (5.6)	5.4 (0.4)	6.4 (0.6)	14.0 (1.3)
ARIC African ancestry	100	41 (41.0)	56.8 (5.4)	5.6 (0.4)	6.3 (0.5)	13.1 (1.4)
FHS European ancestry Offspring Cohort	1,211	522 (43.1)	71.7 (8.0)	5.6 (0.3)	5.5 (0.5)	13.7 (1.2)
FHS European ancestry Third Generation	1,025	473 (46.2)	46.1 (9.0)	5.4 (0.3)	5.2 (0.5)	14.2 (1.2)
JHS African ancestry	2,356	915 (38.8)	54.0 (13.0)	5.5 (0.5)	5.0 (0.5)	13.2 (1.5)
MESA European ancestry	1,456	703 (48.3)	69.9 (9.5)	5.5 (0.3)	5.1 (0.5)	--
MESA African ancestry	667	313 (46.9)	68.6 (9.6)	5.7 (0.4)	5.1 (0.5)	--
MESA Hispanic ancestry	650	323 (49.7)	68.2 (9.7)	5.7 (0.4)	5.3 (0.5)	--
MESA Asian ancestry	407	206 (50.6)	68.4 (9.3)	5.7 (0.4)	5.2 (0.5)	--

Table S5: Most associated single nucleotide variants in the regions detected at the sub-genome-wide level ($P<5\times10^{-7}$) by ancestry and in the meta-analysis of HbA1c in non-diabetic individuals in TOPMed cohorts

				Meta-analysis			Europeans (N=6,158)				African-Americans (N=3,123)			
MarkerID[a]	**rsID**	**Gene**	**Ref/Alt Allele**	**Beta**	**SE**	***P value***	**AAF[b]**	**Beta**	**SE**	***P value***	**AAF[b]**	**Beta**	**SE**	***P value***
1:69561493	rs546782425	*LRRC7*	G/A	-0.37	0.07	4.10E-07	--	--	--	--	0.01	-0.37	0.07	4.10E-07
2:168900844	rs1402837	*G6PC2*	C/T	0.03	0.01	1.40E-07	0.23	0.04	0.01	6.80E-08	0.29	0.01	0.01	0.29
3:5940489	rs541283713	*EDEM1*	G/T	-0.48	0.09	8.90E-08	--	--	--	--	0.004	-0.48	0.09	8.90E-08
3:73834413	rs73103979	*PDZRN3*	C/T	-0.03	0.01	8.90E-08	0.26	-0.03	0.01	1.40E-06	0.06	-0.03	0.02	0.18
4:100766435	rs184930483	*EMCN*	T/C	-0.33	0.06	6.50E-08	--	--	--	--	--	--	--	--
5:73310829	rs148873180	*FOXD1*	G/A	-0.32	0.06	4.70E-07	--	--	--	--	0.01	-0.32	0.06	4.70E-07
7:117730753	rs77974976	*CTTNBP2*	G/T	-0.35	0.07	2.40E-07	--	--	--	--	0.01	-0.35	0.07	2.40E-07
7:117730990		*CTTNBP2*	TG/T	-0.35	0.07	2.40E-07	--	--	--	--	0.01	-0.35	0.07	2.40E-07
7:117734487	rs115069847	*CTTNBP2*	A/G	-0.34	0.07	3.90E-07	--	--	--	--	0.01	-0.34	0.07	3.90E-07
8:57797437	rs182571425	*FAM110B*	T/G	-0.35	0.06	6.60E-08	0.002	-0.35	0.06	6.60E-08	--	--	--	--
8:59954108	rs148165085	*CA8*	C/G	-0.40	0.08	1.60E-07	--	--	--	--	0.01	-0.40	0.08	1.60E-07
10:53414779	rs185736000	*PCDH15*	G/A	-0.48	0.09	1.10E-07	--	--	--	--	0.004	-0.48	0.09	1.10E-07
10:108997093	rs551601853	*XPNPEP1*	A/G	-0.55	0.10	3.10E-08	--	--	--	--	0.003	-0.55	0.10	3.1E-08
11:5078417	rs77362408	*OR52E2*	C/G	-0.15	0.03	3.10E-07	--	--	--	--	0.04	-0.14	0.03	1.60E-06
11:5492424	rs73392143	*HBG2/HBE1*	G/T	-0.20	0.04	1.20E-07	--	--	--	--	0.02	-0.20	0.04	1.20E-07
11:6014322	rs76128436	*OR56A4*	C/T	-0.19	0.04	1.80E-07	--	--	--	--	0.03	-0.19	0.04	1.80E-07
17:32533283	rs140882821	*MYO1D*	G/A	-0.19	0.04	2.80E-07	--	--	--	--	0.02	-0.19	0.04	2.50E-07
17:78138386	rs144120533	*TMC8*	C/T	-0.26	0.05	2.70E-07	0.003	-0.26	0.05	2.70E-07	--	--	--	--
18:2770634		*SMCHD1*	G/GT	0.17	0.03	1.40E-07	0.001	0.16	0.11	0.14	0.03	0.17	0.03	8.40E-07
18:34585879	rs187431535	*DTNA*	T/A	0.25	0.05	9.20E-08	0.004	0.25	0.05	9.20E-08	--	--	--	--
18:34712254	rs183229631	*DTNA*	G/A	0.25	0.05	3.40E-07	0.003	0.25	0.05	3.40E-07	--	--	--	--
21:37519947	rs11701836	*DYRK1A*	A/G	-0.03	0.01	4.40E-07	0.26	-0.02	0.01	1.70E-03	0.32	-0.05	0.01	1.10E-04
22:37066896	rs855791	*TMPRSS6*	G/A	0.03	0.01	1.10E-07	0.44	0.03	0.01	1.60E-07	0.16	0.01	0.02	0.53
X:153542357	rs143745197	*ATP2B3*	G/A	-0.31	0.06	2.30E-07	--	--	--	--	0.01	-0.31	0.06	2.30E-07
X:154593327	rs184539426	*CTAG1A*	G/T	-0.23	0.04	1.40E-07	--	--	--	--	0.01	-0.23	0.04	1.40E-07
X:154877653	rs189305788	*F8*	C/A	-0.40	0.08	2.40E-07	--	--	--	--	0.01	-0.40	0.08	2.40E-07

Table S5 (continued):

				Meta-analysis			Hispanics (N=650)				Asians (N=407)			
MarkerID[a]	**rsID**	**Gene**	**Ref/Alt Allele**	**Beta**	**SE**	***P value***	**AAF[b]**	**Beta**	**SE**	***P value***	**AAF[b]**	**Beta**	**SE**	***P value***
1:69561493	rs546782425	*LRRC7*	G/A	-0.37	0.07	4.10E-07	--	--	--	--	--	--	--	--
2:168900844	rs1402837	*G6PC2*	C/T	0.03	0.01	1.40E-07	0.22	0.02	0.03	0.45	0.38	0.01	0.03	0.63
3:5940489	rs541283713	*EDEM1*	G/T	-0.48	0.09	8.90E-08	--	--	--	--	--	--	--	--
3:73834413	rs73103979	*PDZRN3*	C/T	-0.03	0.01	8.90E-08	0.13	-0.06	0.03	0.08	0.01	-0.24	0.12	0.04
4:100766435	rs184930483	*EMCN*	T/C	-0.33	0.06	6.50E-08	--	--	--	--	0.04	-0.33	0.06	6.50E-08
5:73310829	rs148873180	*FOXD1*	G/A	-0.32	0.06	4.70E-07	--	--	--	--	--	--	--	--
7:117730753	rs77974976	*CTTNBP2*	G/T	-0.35	0.07	2.40E-07	--	--	--	--	--	--	--	--
7:117730990	--	*CTTNBP2*	TG/T	-0.35	0.07	2.40E-07	--	--	--	--	--	--	--	--
7:117734487	rs115069847	*CTTNBP2*	A/G	-0.34	0.07	3.90E-07	--	--	--	--	--	--	--	--
8:57797437	rs182571425	*FAM110B*	T/G	-0.35	0.06	6.60E-08	--	--	--	--	--	--	--	--
8:59954108	rs148165085	*CA8*	C/G	-0.40	0.08	1.60E-07	--	--	--	--	--	--	--	--
10:53414779	rs185736000	*PCDH15*	G/A	-0.48	0.09	1.10E-07	--	--	--	--	--	--	--	--
10:108997093	rs551601853	*XPNPEP1*	A/G	-0.55	0.10	3.10E-08	--	--	--	--	--	--	--	--
11:5078417	rs77362408	*OR52E2*	C/G	-0.15	0.03	3.10E-07	0.005	-0.39	0.17	0.02	--	--	--	--
11:5492424	rs73392143	*HBG2/HBE1*	G/T	-0.20	0.04	1.20E-07	--	--	--	--	--	--	--	--
11:6014322	rs76128436	*OR56A4*	C/T	-0.19	0.04	1.80E-07	--	--	--	--	--	--	--	--
17:32533283	rs140882821	*MYO1D*	G/A	-0.19	0.04	2.80E-07	0.005	-0.08	0.17	0.61	--	--	--	--
17:78138386	rs144120533	*TMC8*	C/T	-0.26	0.05	2.70E-07	--	--	--	--	--	--	--	--
18:2770634	--	*SMCHD1*	G/GT	0.17	0.03	1.40E-07	0.01	0.15	0.13	0.26	--	--	--	--
18:34585879	rs187431535	*DTNA*	T/A	0.25	0.05	9.20E-08	--	--	--	--	--	--	--	--
18:34712254	rs183229631	*DTNA*	G/A	0.25	0.05	3.40E-07	--	--	--	--	--	--	--	--
21:37519947	rs11701836	*DYRK1A*	A/G	-0.03	0.01	4.40E-07	0.28	-0.02	0.02	0.44	0.37	-0.06	0.03	0.02
22:37066896	rs855791	*TMPRSS6*	G/A	0.03	0.01	1.10E-07	0.45	0.05	0.02	0.02	0.44	-0.01	0.02	0.57
X:153542357	rs143745197	*ATP2B3*	G/A	-0.31	0.06	2.30E-07	--	--	--	--	--	--	--	--
X:154593327	rs184539426	*CTAG1A*	G/T	-0.23	0.04	1.40E-07	--	--	--	--	--	--	--	--
X:154877653	rs189305788	*F8*	C/A	-0.40	0.08	2.40E-07	--	--	--	--	--	--	--	--

[a]MarkerID defined as Chromosome:Position. Positions are indicated on Build 38
[b]AAF: Alternate Allele Frequency

Table S6: Results of the mediation analyses for the most associated single nucleotide variants in the regions detected at the genome-wide level ($P<2\times10^{-8}$) in the meta-analysis of HbA1c in non-diabetic individuals in TOPMed cohorts

MarkerID[a]	rsID	Closest Gene	Ref/Alt Allele	HbA1c (N=10,338)			FG[b] (N=10,185)			HB[c] (N=6,411)			Class
				Beta	SE	*P value*	Beta	SE	*P value*	Beta	SE	*P value*	
7:44186502	rs2971670	*GCK*	C/T	0.04	0.01	1.70E-09	0.03	0.01	2.70E-05	0.04	0.01	4.60E-08	Glycemic
10:69334748	rs17476364	*HK1*	T/C	-0.09	0.01	3.10E-21	-0.09	0.01	1.60E-22	-0.09	0.01	3.40E-17	Unclassified
11:5464062	rs1039215	*HBG2/HBE1*	A/G	-0.21	0.04	2.00E-09	-0.23	0.13	2.60E-11	-0.21	0.04	7.50E-08	Unclassified
17:82739582	rs113373052	*FN3K*	C/T	0.03	0.01	4.50E-10	0.03	0.01	1.30E-11	0.03	0.01	2.20E-07	Unclassified
X:154533025	rs76723693	*G6PD*	A/G	-0.44	0.07	2.16E-09	-0.43	0.07	7.13E-10	-0.40	0.09	7.64E-06	Unclassified
X:154536002	rs1050828	*G6PD*	C/T	-0.41	0.01	5.10E-210	-0.40	0.01	7.00E-229	-0.42	0.02	2.40E-156	Unclassified

Table S6 (continued):

MarkerID[a]	rsID	Closest Gene	Ref/Alt Allele	MCHC[d] (N=3,779)			MCH[e] (N=3,779)			MCV[f] (N=6,178)			Class
				Beta	SE	*P value*	Beta	SE	*P value*	Beta	SE	*P value*	
7:44186502	rs2971670	*GCK*	C/T	0.05	0.01	2.20E-06	0.05	0.01	2.60E-06	0.04	0.01	3.90E-08	Glycemic
10:69334748	rs17476364	*HK1*	T/C	-0.08	0.02	1.30E-07	-0.08	0.02	3.20E-07	-0.09	0.01	2.00E-17	Unclassified
11:5464062	rs1039215	*HBG2/HBE1*	A/G	-0.21	0.04	1.00E-06	-0.24	0.04	1.50E-08	-0.25	0.04	1.30E-09	Unclassified
17:82739582	rs113373052	*FN3K*	C/T	0.03	0.01	1.80E-04	0.03	0.01	1.30E-04	0.03	0.01	1.50E-07	Unclassified
X:154533025	rs76723693	*G6PD*	A/G	-0.44	0.09	6.54E-07	-0.43	0.09	9.32E-07	-0.40	0.09	2.96E-06	Unclassified
X:154536002	rs1050828	*G6PD*	C/T	-0.41	0.02	7.80E-133	-0.39	0.02	2.40E-124	-0.38	0.02	5.20E-126	Unclassified

[a]MarkerID defined as Chromosome:Position. Positions are indicated on Build 38.
[b]FG: fasting glucose
[c]HB: Hemoglobin
[d]MCHC: Mean Corpuscular Hemoglobin Concentration levels
[e]MCH: Mean Corpuscular Hemoglobin
[f]MCV: Mean Corpuscular Volume

Table S7: Results of the mediation analyses for the most associated single nucleotide variants in the regions detected at the sub-genome-wide level ($P<5\times10^{-7}$) in the meta-analysis of HbA1c in non-diabetic individuals in TOPMed cohorts

				HbA1c (N=10,338)			FG[b] (N=10,185)			HB[c] (N=6,411)			
MarkerID[a]	**rsID**	**Closest Gene**	**Ref/Alt**	**Beta**	**SE**	***P value***	**Beta**	**SE**	***P value***	**Beta**	**SE**	***P value***	**Class**
1:69561493	rs546782425	*LRRC7*	G/A	-0.37	0.07	4.10E-07	-0.32	0.07	5.62E-06	-0.39	0.08	1.31E-06	Unclassified
2:168900844	rs1402837	*G6PC2*	C/T	0.03	0.01	1.40E-07	0.02	0.01	3.57E-04	0.03	0.01	3.33E-06	Glycemic
3:5940489	rs541283713	*EDEM1*	G/T	-0.48	0.09	8.90E-08	-0.49	0.09	1.20E-08	-0.52	0.10	1.50E-07	Unclassified
3:73834413	rs73103979	*PDZRN3*	C/T	-0.03	0.01	8.90E-08	-0.03	0.01	2.10E-06	-0.03	0.01	2.38E-05	Unclassified
4:100766435	rs184930483	*EMCN*	C/T	-0.33	0.06	6.50E-08	0.27	0.06	2.33E-06	--	--	--	Unclassified
5:73310829	rs148873180	*FOXD1*	G/A	-0.32	0.06	4.70E-07	-0.29	0.06	3.92E-06	-0.37	0.08	2.24E-06	Unclassified
7:117730753	rs77974976	*CTTNBP2*	G/T	-0.35	0.07	2.40E-07	-0.30	0.07	8.40E-06	-0.41	0.08	2.99E-07	Unclassified
7:117730990	--	*CTTNBP2*	TG/T	-0.35	0.07	2.40E-07	-0.30	0.07	8.40E-06	-0.41	0.08	2.99E-07	Unclassified
7:117734487	rs115069847	*CTTNBP2*	G/A	0.34	0.07	3.90E-07	0.29	0.07	9.37E-06	0.39	0.08	5.89E-07	Unclassified
8:57797437	rs182571425	*FAM110B*	G/T	0.35	0.06	6.60E-08	0.31	0.06	4.98E-07	0.27	0.08	7.24E-04	Unclassified
8:59954108	rs148165085	*CA8*	G/C	0.40	0.08	1.60E-07	0.37	0.07	4.81E-07	0.44	0.09	5.91E-07	Unclassified
10:53414779	rs185736000	*PCDH15*	G/A	-0.48	0.09	1.10E-07	-0.45	0.09	3.87E-07	-0.48	0.11	8.86E-06	Unclassified
10:108997093	rs551601853	*XPNPEP1*	A/G	-0.55	0.10	3.10E-08	-0.53	0.10	3.70E-08	-0.58	0.11	1.16E-07	Unclassified
11:5078417	rs77362408	*OR52E2*	G/C	0.15	0.03	3.10E-07	0.16	0.03	4.11E-09	0.14	0.03	3.29E-05	Unclassified
11:5492424	rs73392143	*HBG2/HBE1*	G/T	-0.20	0.04	1.20E-07	-0.21	0.04	4.01E-09	-0.20	0.04	2.99E-06	Unclassified
11:6014322	rs76128436	*OR56A4*	C/T	-0.19	0.04	1.80E-07	-0.20	0.04	2.04E-08	-0.17	0.04	4.53E-05	Unclassified
17:32533283	rs140882821	*MYO1D*	G/A	-0.19	0.04	2.80E-07	-0.18	0.04	4.48E-07	-0.22	0.04	9.28E-08	Unclassified
17:78138386	rs144120533	*TMC8*	C/T	-0.26	0.05	2.70E-07	-0.23	0.05	3.18E-06	-0.28	0.06	5.44E-07	Unclassified
18:2770634	--	*SMCHD1*	GT/G	-0.17	0.03	1.40E-07	-0.15	0.03	7.18E-07	-0.14	0.04	3.40E-04	Unclassified
18:34585879	rs187431535	*DTNA*	T/A	0.25	0.05	9.20E-08	0.23	0.05	3.03E-07	0.28	0.06	2.06E-06	Unclassified
18:34712254	rs183229631	*DTNA*	G/A	0.25	0.05	3.40E-07	0.24	0.05	3.99E-07	0.27	0.06	7.71E-06	Unclassified
21:37519947	rs11701836	*DYRK1A*	G/A	0.03	0.01	4.40E-07	0.03	0.01	6.34E-09	0.02	0.01	1.11E-03	Unclassified
22:37066896	rs855791	*TMPRSS6*	G/A	0.03	0.01	1.10E-07	0.03	0.00	1.02E-08	0.02	0.01	2.98E-04	Unclassified
X:153542357	rs143745197	*ATP2B3*	G/A	-0.31	0.06	2.30E-07	-0.32	0.06	2.33E-08	-0.33	0.07	2.93E-06	Unclassified
X:154593327	rs184539426	*CTAG1A*	G/T	-0.23	0.04	1.40E-07	-0.24	0.04	2.59E-08	-0.23	0.05	1.24E-05	Unclassified
X:154877653	rs189305788	*F8*	C/A	-0.40	0.08	2.40E-07	-0.40	0.07	9.30E-08	-0.38	0.09	1.36E-05	Unclassified

Table S7 (continued):

				MCHC[d] (N=3,779)			MCH[e] (N=3,779)			MCV[f] (N=6,178)			
MarkerID[a]	**rsID**	**Closest Gene**	**Ref/Alt**	**Beta**	**SE**	***P value***	**Beta**	**SE**	***P value***	**Beta**	**SE**	***P value***	**Class**
1:69561493	rs546782425	*LRRC7*	G/A	-0.40	0.08	1.32E-06	-0.38	0.08	4.55E-06	-0.36	0.08	3.34E-06	Unclassified
2:168900844	rs1402837	*G6PC2*	C/T	0.03	0.01	1.00E-03	0.03	0.01	1.06E-03	0.03	0.01	2.22E-06	Glycemic
3:5940489	rs541283713	*EDEM1*	G/T	-0.46	0.10	1.08E-05	-0.45	0.10	1.21E-05	-0.43	0.10	1.16E-05	Unclassified
3:73834413	rs73103979	*PDZRN3*	C/T	-0.02	0.01	8.33E-02	-0.02	0.01	1.14E-01	-0.03	0.01	7.46E-05	Unclassified
4:100766435	rs184930483	*EMCN*	C/T	--	--	--	--	--	--	--	--		Unclassified
5:73310829	rs148873180	*FOXD1*	G/A	-0.36	0.08	5.33E-06	-0.35	0.08	7.73E-06	-0.33	0.08	1.33E-05	Unclassified
7:117730753	rs77974976	*CTTNBP2*	G/T	-0.35	0.08	4.25E-05	-0.32	0.08	1.01E-04	-0.32	0.08	3.48E-05	Unclassified
7:117730990	--	*CTTNBP2*	TG/T	-0.35	0.08	4.25E-05	-0.32	0.08	1.01E-04	-0.32	0.08	3.48E-05	Unclassified
7:117734487	rs115069847	*CTTNBP2*	G/A	0.33	0.08	7.19E-05	0.31	0.08	1.72E-04	0.31	0.08	6.14E-05	Unclassified
8:57797437	rs182571425	*FAM110B*	G/T	--	--	--	--	--	--	0.27	0.08	5.29E-04	Unclassified
8:59954108	rs148165085	*CA8*	G/C	0.44	0.10	8.96E-06	0.41	0.10	1.91E-05	0.39	0.09	1.74E-05	Unclassified
10:53414779	rs185736000	*PCDH15*	G/A	-0.30	0.13	1.87E-02	-0.28	0.12	2.16E-02	-0.33	0.12	5.39E-03	Unclassified
10:108997093	rs551601853	*XPNPEP1*	A/G	-0.48	0.12	4.50E-05	-0.47	0.12	4.80E-08	-0.46	0.11	6.80E-05	Unclassified
11:5078417	rs77362408	*OR52E2*	G/C	0.13	0.03	1.07E-04	0.16	0.03	1.75E-06	0.17	0.03	5.64E-07	Unclassified
11:5492424	rs73392143	*HBG2/HBE1*	G/T	-0.21	0.04	3.52E-06	-0.24	0.04	9.30E-08	-0.23	0.04	4.11E-08	Unclassified
11:6014322	rs76128436	*OR56A4*	C/T	-0.18	0.04	3.64E-05	-0.20	0.04	5.91E-06	-0.18	0.04	2.18E-05	Unclassified
17:32533283	rs140882821	*MYO1D*	G/A	-0.22	0.04	5.47E-07	-0.21	0.04	6.67E-07	-0.21	0.04	7.39E-07	Unclassified
17:78138386	rs144120533	*TMC8*	C/T	-0.25	0.08	2.51E-03	-0.23	0.08	4.83E-03	-0.27	0.06	1.53E-06	Unclassified
18:2770634	--	*SMCHD1*	GT/G	-0.14	0.04	8.07E-04	-0.15	0.04	1.89E-04	-0.15	0.04	1.66E-04	Unclassified
18:34585879	rs187431535	*DTNA*	T/A	0.33	0.08	1.45E-05	0.33	0.08	1.82E-05	0.29	0.06	1.37E-06	Unclassified
18:34712254	rs183229631	*DTNA*	G/A	0.32	0.08	6.10E-05	0.31	0.08	8.36E-05	0.28	0.06	4.57E-06	Unclassified
21:37519947	rs11701836	*DYRK1A*	G/A	0.03	0.01	1.20E-03	0.03	0.01	1.51E-03	0.02	0.01	2.74E-03	Unclassified
22:37066896	rs855791	*TMPRSS6*	G/A	0.02	0.01	4.02E-03	0.02	0.01	2.96E-02	0.02	0.01	9.61E-03	Unclassified
X:153542357	rs143745197	*ATP2B3*	G/A	-0.33	0.07	2.08E-06	-0.31	0.07	8.07E-06	-0.30	0.07	9.56E-06	Unclassified
X:154593327	rs184539426	*CTAG1A*	G/T	-0.25	0.05	4.68E-06	-0.24	0.05	4.00E-06	-0.23	0.05	7.27E-06	Unclassified
X:154877653	rs189305788	*F8*	C/A	-0.42	0.09	1.49E-06	-0.40	0.09	2.29E-06	-0.37	0.08	6.53E-06	Unclassified

[a]MarkerID defined as Chromosome:Position. Positions are indicated on Build 38.
[b]FG: fasting glucose
[c]HB: Hemoglobin
[d]MCHC: mean corpuscular hemoglobin concentration levels
[e]MCH: mean corpuscular hemoglobin
[f]MCV: mean corpuscular volume

Table S8: Association of the most associated single nucleotide variants in the regions detected at the genome-wide level ($P<2\times10^{-8}$) in the meta-analysis of HbA1c in non-diabetic individuals in TOPMed cohorts with fasting glucose (FG, N=26,883) and erythrocytic traits (HCT, HB, MCV, MCH, MCHC, RBC, RDW, N=25,080)

				HbA1c		FG[c]		HB[d]		HCT[e]		MCV[f]	
MarkerID[a]	**rsID**	**Closest Gene**	**Ref/Alt Allele**	**Beta**	***P value***	**Beta**	***P value***	**Score[b]**	***P value***	**Score[b]**	***P value***	**Score[b]**	***P value***
7:44186502	rs2971670	*GCK*	C/T	0.04	1.70E-09	0.06	7.40E-28	33.37	0.68	7.79	0.78	-2.35	0.87
10:69334748	rs17476364	*HK1*	T/C	-0.09	3.10E-21	-0.01	0.17	420.08	1.60E-12	140.77	2.10E-12	35.6	4.60E-04
11:5464062	rs1039215	*HBG2*	A/G	-0.21	2.00E-09	-0.02	0.41	-49.73	4.90E-04	-20.35	3.60E-05	-5.75	0.02
17:82739582	rs113373052	*FN3K*	C/T	0.03	4.50E-10	0.003	0.5	59.46	0.54	39.33	0.23	-11.41	0.48
X:154533025	rs76723693	*G6PD*	A/G	-0.44	0.07	-0.09	0.10	-12.50	0.05	-2.69	0.22	1.85	0.08
X:154536002	rs1050828	*G6PD*	C/T	-0.41	5.10E-210	-0.01	0.42	-246.54	3.40E-13	-76.5	6.30E-11	66.83	2.90E-31

Table S8 (continued):

				MCH[g]		MCHC[h]		RBC[i]		RDW[j]		
MarkerID*	**rsID**	**Closest Gene**	**Ref/Alt Allele**	**Score[b]**	***P value***	**Score[b]**	***P value***	**Score[b]**	***P value***	**Score[b]**	***P value***	**Class**
7:44186502	rs2971670	*GCK*	C/T	-21.06	0.54	62.55	0.37	91.29	0.57	91.04	0.03	Glycemic
10:69334748	rs17476364	*HK1*	T/C	64.16	0.01	-28.79	0.55	267.44	0.02	35.47	0.23	Erythrocytic
11:5464062	rs1039215	*HBG2*	A/G	-15.07	0.01	7.2	0.65	-11.97	0.71	16.87	0.06	Unclassified
17:82739582	rs113373052	*FN3K*	C/T	-55.38	0.16	-142.06	0.08	331.35	0.08	-54.06	0.28	Unclassified
X:154533025	rs76723693	*G6PD*	A/G	2.30	0.39	-12.20	0.09	-41.38	0.004	-15.84	1.46E-04	Erythrocytic
X:154536002	rs1050828	*G6PD*	C/T	122.46	1.10E-16	-152.61	6.20E-05	-911.53	1.90E-32	-248.94	6.30E-34	Erythrocytic

[a]MarkerID defined as Chromosome:Position. Positions are indicated on Build 38.
[b]from a pooled ancestry score test
[c]FG: Fasting Glucose
[d]HB: Hemoglobin
[e]HCT: Hematocrit,
[f]MCV: Mean Corpuscular Volume,
[g]MCH: Mean Corpuscular Hemoglobin
[h]MCHC: Mean Corpuscular Hemoglobin Concentration
[i]RBC: Red Blood Cell count
[j]RDW: Red cell Distribution Width

Table S9: Association of the most associated single nucleotide variants in the regions detected at the genome-wide level ($P<2\times10^{-8}$) in the meta-analysis of HbA1c in non-diabetic individuals in TOPMed cohorts with erythrocytic traits (HCT, HB, MCV, MCH, RBC, RDW) in published genome-wide association studies composed of people of European, Hispanic and African ancestry

Ancestry	PMID	Trait	Effect/Other Allele	EAF[a]	Beta	SE	*P value*
rs17476364 (10:69,334,748) *HK1*							
European	27863252	HCT[b]	C/T	0.11	0.15	0.006	7.7E-159
European	27863252	HB[c]	C/T	0.11	0.15	0.006	1.2E-156
European	27863252	MCH[d]	C/T	0.11	0.08	0.006	3.7E-44
European	27863252	MCV[e]	C/T	0.11	0.09	0.006	7.3E-55
European	27863252	RBC[f]	C/T	0.11	0.08	0.006	1.8E-48
European	27863252	RDW[g]	C/T	0.11	0.04	0.006	2.0E-11
rs1050828 (X:154,536,002) *G6PD*							
African American	21153663	HCT	T/C	0.11	-0.22	0.03	1.2E-13
African American	21153663	HB	T/C	0.11	-0.24	0.03	1.2E-15
Hispanic	28453575	MCV	T/C	0.02	1.92	0.22	1.3E-17
African American	21153663	MCV	T/C	0.11	0.34	0.04	3.0E-18
African American	23446634	RBC	T/C	0.11	-0.42	0.04	4.3E-21
Hispanic	28453575	RBC	T/C	0.02	-0.13	0.01	1.8E-18
Hispanic	28453575	RDW	T/C	0.02	-0.04	0.003	1.5E-29

[a]EAF: Effect Allele Frequency
[b]HCT: Hematocrit
[c]HB: hemoglobin
[d]MCH: mean corpuscular hemoglobin
[e]MCV: Mean Corpuscular Volume
[f]RBC: Red Blood Cell count
[g]RDW: Red cell Distribution Width

Table S10: Association of the most associated single nucleotide variants in the regions detected at the sub-genome-wide level ($P<5\times10^{-7}$) in the meta-analysis of HbA1c in non-diabetic individuals in TOPMed cohorts with fasting glucose (FG) in the TOPMed Diabetes working group (N=26,883) and erythrocytic traits (HCT, HB, MCV, MCH, MCHC, RBC, RDW) in the TOPMed Hematology working group (N=25,080)

				HbA1c		**FG[c]**		**HB[d]**		**HCT[e]**		**MCV[f]**	
MarkerID[a]	**rsID**	**Closest Gene**	**Ref/Alt**	**Beta**	***P value***	**Beta**	***P value***	**Score[b]**	***P value***	**Score[b]**	***P value***	**Score[b]**	***P value***
1:69561493	rs546782425	*LRRC7*	G/A	-0.37	4.10E-07	-0.03	5.10E-01	-13.65	3.90E-02	-4.37	5.60E-02	1.74	1.30E-01
2:168900844	rs1402837	*G6PC2*	C/T	0.03	1.40E-07	0.04	3.40E-17	-2.77	9.80E-01	-3.85	9.00E-01	-11.70	4.40E-01
3:5940489	rs541283713	*EDEM1*	G/T	-0.48	8.90E-08	-0.04	4.90E-01	-3.06	6.40E-01	-2.05	3.60E-01	-0.09	9.30E-01
3:73834413	rs73103979	*PDZRN3*	C/T	-0.03	8.90E-08	-0.002	7.60E-01	81.14	3.40E-01	17.81	5.30E-01	14.51	3.10E-01
4:100766435	rs184930483	*EMCN*	T/C	-0.33	6.50E-08	-0.13	7.20E-02	--	--	--	--	--	--
5:73310829	rs148873180	*FOXD1*	G/A	-0.32	4.70E-07	-0.01	7.60E-01	-17.70	4.50E-02	-5.77	5.90E-02	-0.91	5.50E-01
7:117730753	rs77974976	*CTTNBP2*	G/T	-0.35	2.40E-07	0.02	6.80E-01	1.52	8.30E-01	0.59	8.10E-01	2.76	1.50E-02
7:117730990	--	*CTTNBP2*	TG/T	-0.35	2.40E-07	0.02	6.80E-01	1.52	8.30E-01	0.59	8.10E-01	2.76	1.50E-02
7:117734487	rs115069847	*CTTNBP2*	A/G	-0.34	3.90E-07	0.03	6.00E-01	0.71	9.20E-01	0.33	8.90E-01	2.82	1.40E-02
8:57797437	rs182571425	*FAM110B*	T/G	-0.35	6.60E-08	-0.02	6.90E-01	8.95	3.00E-01	3.34	2.50E-01	1.18	3.50E-01
8:59954108	rs148165085	*CA8*	C/G	-0.40	1.60E-07	-0.02	6.40E-01	-3.57	6.30E-01	-1.35	6.00E-01	0.43	7.10E-01
10:53414779	rs185736000	*PCDH15*	G/A	-0.48	1.10E-07	-0.007	9.30E-01	-18.85	1.00E-03	-6.37	1.30E-03	0.22	8.00E-01
10:108997093	rs551601853	*XPNPEP1*	A/G	-0.55	3.10E-08	-0.01	0.85	2.98	0.56	1.63	0.36	-0.26	0.76
11:5078417	rs77362408	*OR52E2*	C/G	-0.15	3.10E-07	0.03	2.80E-01	-56.71	9.60E-04	-22.80	1.20E-04	-11.35	7.90E-05
11:5492424	rs73392143	*HBG2/HBE1*	G/T	-0.20	1.20E-07	0.02	5.00E-01	-45.16	8.40E-04	-18.46	7.80E-05	-5.04	2.40E-02
11:6014322	rs76128436	*OR56A4*	C/T	-0.19	1.80E-07	0.01	6.30E-01	-21.24	1.30E-01	-10.67	2.70E-02	-1.81	4.20E-01
17:32533283	rs140882821	*MYO1D*	G/A	-0.19	2.80E-07	-0.04	2.00E-01	-12.32	3.30E-01	-3.20	4.70E-01	2.30	2.90E-01
17:78138386	rs144120533	*TMC8*	C/T	-0.26	2.70E-07	-0.08	1.30E-01	0.30	9.70E-01	0.55	8.60E-01	4.02	1.10E-02
18:2770634	--	*SMCHD1*	G/GT	0.17	1.40E-07	0.04	8.00E-02	30.21	5.90E-02	10.26	6.20E-02	0.65	8.10E-01
18:34585879	rs187431535	*DTNA*	T/A	0.25	9.20E-08	0.02	6.60E-01	-16.25	1.70E-01	-3.10	4.40E-01	-1.30	4.60E-01
18:34712254	rs183229631	*DTNA*	G/A	0.25	3.40E-07	0.01	8.20E-01	-13.07	2.70E-01	-1.92	6.30E-01	-1.23	4.80E-01
21:37519947	rs11701836	*DYRK1A*	A/G	-0.03	4.40E-07	-0.004	3.80E-01	125.48	1.80E-01	26.06	4.10E-01	-9.34	5.40E-01
22:37066896	rs855791	*TMPRSS6*	G/A	0.03	1.10E-07	0.005	2.90E-01	-1006.23	1.50E-23	-243.33	7.70E-13	-172.78	2.30E-24
X:153542357	rs143745197	*ATP2B3*	G/A	-0.31	2.30E-07	-0.06	1.70E-01	-2.48	7.60E-01	-1.40	6.20E-01	3.42	1.70E-02
X:154593327	rs184539426	*CTAG1A*	G/T	-0.23	1.40E-07	-0.04	2.50E-01	-10.06	3.10E-01	-3.41	3.20E-01	1.82	2.90E-01
X:154877653	rs189305788	*F8*	C/A	-0.40	2.40E-07	-0.09	9.10E-02	-10.14	1.30E-01	-2.25	3.30E-01	2.10	7.00E-02

Table S10 (continued):

				MCH[g]		MCHC[h]		RBC[i]		RDW[j]		
MarkerID[a]	**rsID**	**Closest Gene**	**Ref/Alt Allele**	**Score[b]**	***P value***	**Score[b]**	***P value***	**Score[b]**	***P value***	**Score[b]**	***P value***	**Class**
1:69561493	rs546782425	*LRRC7*	G/A	4.88	8.60E-02	-3.55	6.40E-01	-42.23	4.50E-03	-7.04	1.10E-01	Erythrocytic
2:168900844	rs1402837	*G6PC2*	C/T	-28.09	4.60E-01	-50.20	5.20E-01	295.90	9.40E-02	18.65	6.90E-01	Glycemic
3:5940489	rs541283713	*EDEM1*	G/T	1.10	6.90E-01	11.35	9.00E-02	-16.87	2.30E-01	-2.05	6.40E-01	Unclassified
3:73834413	rs73103979	*PDZRN3*	C/T	38.39	2.70E-01	71.65	2.90E-01	-135.03	4.00E-01	-41.56	3.30E-01	Unclassified
4:100766435	rs184930483	*EMCN*	T/C	--	--	--	--	--	--	--	--	Unclassified
5:73310829	rs148873180	*FOXD1*	G/A	-2.42	5.40E-01	-4.85	6.30E-01	-26.00	2.00E-01	-0.71	8.90E-01	Unclassified
7:117730753	rs77974976	*CTTNBP2*	G/T	5.60	4.40E-02	4.20	5.60E-01	-40.12	5.90E-03	-0.58	8.90E-01	Unclassified
7:117730990	--	*CTTNBP2*	TG/T	5.60	4.40E-02	4.20	5.60E-01	-40.12	5.90E-03	-0.58	8.90E-01	Unclassified
7:117734487	rs115069847	*CTTNBP2*	A/G	5.67	4.40E-02	3.65	6.20E-01	-41.33	5.00E-03	-0.79	8.60E-01	Unclassified
8:57797437	rs182571425	*FAM110B*	T/G	1.25	6.60E-01	-0.98	8.70E-01	-8.07	5.40E-01	-1.82	6.50E-01	Unclassified
8:59954108	rs148165085	*CA8*	C/G	0.18	9.50E-01	-5.88	4.30E-01	2.58	8.70E-01	1.13	8.00E-01	Unclassified
10:53414779	rs185736000	*PCDH15*	G/A	0.32	8.90E-01	0.39	9.50E-01	-21.08	8.40E-02	1.75	5.80E-01	Erythrocytic
10:108997093	rs551601853	*XPNPEP1*	A/G	-0.35	0.87	-4.87	0.4	8.82	0.44	-4.14	0.18	Unclassified
11:5078417	rs77362408	*OR52E2*	C/G	-24.54	8.90E-04	8.94	6.40E-01	39.43	3.10E-01	29.01	8.60E-03	Erythrocytic
11:5492424	rs73392143	*HBG2/HBE1*	G/T	-12.84	2.60E-02	8.86	5.60E-01	-14.88	6.20E-01	12.08	1.60E-01	Erythrocytic
11:6014322	rs76128436	*OR56A4*	C/T	-3.24	5.80E-01	15.99	3.00E-01	-28.05	3.60E-01	9.26	2.80E-01	Unclassified
17:32533283	rs140882821	*MYO1D*	G/A	3.39	5.40E-01	-0.40	9.80E-01	-13.13	6.50E-01	-14.39	9.30E-02	Unclassified
17:78138386	rs144120533	*TMC8*	C/T	9.74	1.10E-02	4.00	5.80E-01	-21.21	2.10E-01	-8.72	9.80E-02	Unclassified
18:2770634	--	*SMCHD1*	G/GT	2.95	6.70E-01	9.26	5.90E-01	33.09	3.50E-01	2.04	8.30E-01	Unclassified
18:34585879	rs187431535	*DTNA*	T/A	-4.56	2.90E-01	-6.49	4.50E-01	-9.72	6.20E-01	1.23	8.30E-01	Unclassified
18:34712254	rs183229631	*DTNA*	G/A	-3.79	3.70E-01	-5.50	5.10E-01	-11.17	5.60E-01	-1.79	7.60E-01	Unclassified
21:37519947	rs11701836	*DYRK1A*	A/G	-23.17	5.40E-01	85.42	2.80E-01	185.59	3.00E-01	-80.01	9.40E-02	Unclassified
22:37066896	rs855791	*TMPRSS6*	G/A	-493.31	1.80E-32	-611.28	8.40E-14	218.78	2.60E-01	361.15	1.50E-12	Erythrocytic
X:153542357	rs143745197	*ATP2B3*	G/A	9.92	5.50E-03	4.22	6.50E-01	-42.99	2.20E-02	-14.61	3.80E-03	Erythrocytic
X:154593327	rs184539426	*CTAG1A*	G/T	3.53	4.20E-01	-4.71	6.80E-01	-43.98	5.70E-02	-15.30	2.10E-02	Unclassified
X:154877653	rs189305788	*F8*	C/A	3.25	2.70E-01	-8.98	2.40E-01	-43.17	5.70E-03	-16.90	1.60E-04	Erythrocytic

[a]MarkerID defined as Chromosome:Position. Positions are indicated on Build 38.
[b]from a pooled ancestry score test
[g]MCH: mean corpuscular hemoglobin [h]MCHC: mean corpuscular hemoglobin concentration
[i]RBC: red blood cell count [j]RDW: red cell distribution width

Table S11: Gene-based and burden tests performed in the regions detected at the genome-wide level ($P<2\times10^{-8}$) in the meta-analysis of HbA1c in non-diabetic individuals in TOPMed cohorts and classified as erythrocytic

SKAT								
	Ancestry	**Chr**	**Ensembl gene id**	**Number of variants[a]**	**n.sample.alt[b]**	**Q**	***P value***	**Annotation**
HK1	Europeans	10	ENSG00000156515	37	48	255403.79	0.59	missense
				2	2	11847.45	0.33	high confidence loss of function
				30	173	659814.94	0.68	synonymous
HBG2	African Americans	11	ENSG00000196565	5	54	37503.20	0.73	missense
				2	3	1572.69	0.84	high confidence loss of function
				3	7	9253.68	0.69	synonymous
G6PD	African Americans	X	ENSG00000160211	15	76	4320297.73	9.72E-11	missense
				--	--	--	--	high confidence loss of function
				11	34	92231.24	0.82	synonymous

Burden										
	Ancestry	**Chr**	**Ensembl gene id**	**Number of variants[a]**	**n.sample.alt[b]**	**burden.skew[c]**	**Beta**	**SE**	***P value***	**Annotation**
HK1	Europeans	10	ENSG00000156515	37	48	11.19	0.001	0.002	0.70	missense
				2	2	55.45	-0.015	0.010	0.13	high confidence loss of function
				30	173	6.66	0.001	0.001	0.42	synonymous
HBG2	African Americans	11	ENSG00000196565	5	54	7.45	-0.002	0.003	0.52	missense
				2	3	32.20	0.004	0.010	0.69	high confidence loss of function
				3	7	21.04	-0.004	0.007	0.57	synonymous
G6PD	African Americans	X	ENSG00000160211	15	76	7.41	-0.007	0.002	4.69E-05	missense
				--	--	--	--	--	--	high confidence loss of function
				11	34	11.23	0.003	0.002	0.19	synonymous

[a]variants with MAF < 1% were included in the test
[b]The number of samples with an observed alternate allele at any variant in the aggregate set
[c]The skewness of the burden value for all samples

Table S12: Analysis of HbA1c associations of the top single nucleotide variants in *HGB2* and *G6PD* before and after conditioning on the sickle cell trait variant (rs334) in non-diabetic African-Americans in TOPMed cohorts

MarkerID[a]	Closest Gene	Ref/Alt Alleles	EAF[b]	Beta	SE	*P value*
Main analysis						
11:5464062 (rs1039215)	*HBG2/HBE1*	A/G	0.03	-0.21	0.04	2.04E-09
X:154533025 (rs76723693)	*G6PD*	A/G	0.01	-0.44	0.07	2.16E-09
X:154536002 (rs1050828)	*G6PD*	C/T	0.12	-0.41	0.01	8.38E-205
Adjusted on rs334						
11:5464062 (rs1039215)	*HBG2/HBE1*	A/G	0.03	-0.14	0.04	1.22E-03
X:154533025 (rs76723693)	*G6PD*	A/G	0.01	-0.45	0.07	8.29E-10
X:154536002 (rs1050828)	G6PD	C/T	0.12	-0.41	0.01	5.75E-210

[a]MarkerID defined as Chromosome:Position. Positions are indicated on Build 38.
[b]EAF: Effect Allele Frequency

Table S13: Analysis of HbA1c associations of haplotypes constructed with the distinct single nucleotide variants in the *HBB-HBG2* locus

	rs334	**rs1039215**	**haplotype frequency**	**Beta**	**SE**	***P value***
reference haplotype	T	A	0.95	--	--	--
haplo.1	A	A	0.02	-0.12	0.04	5.24E-03
haplo.2	**A**	**G**	**0.02**	**-0.21**	**0.04**	**5.54E-08**
haplo.3	T	G	0.006	-0.18	0.08	1.82E-02

Table S14: Distribution of HbA1c in the 3,123 African-Americans from TOPMed according to genotypes at the two *G6PD* variants (rs1050828 & rs76723693) and at the *HBG2* variant (rs1039215)

Genotype	Variable	Obs	Mean	SD	Min	Max
rs1050828=0 & rs1039215=0	hba1c	2,372	5.65	0.39	3.7	6.4
rs1050828=0 & rs1039215=1	hba1c	138	5.42	0.44	3.8	6.4
rs1050828=1 & rs1039215=0	hba1c	375	5.28	0.46	3.6	6.4
rs1050828=1 & rs1039215=1	hba1c	15	5.27	0.61	4.5	6.2
rs1050828=2 & rs1039215=0	hba1c	152	4.78	0.40	3.6	6.3
rs1050828=2 & rs1039215=1	hba1c	12	4.60	0.41	4.0	5.5
rs76723693=0 & rs1039215=0	hba1c	2,897	5.56	0.46	3.6	6.4
rs76723693=0 & rs1039215=1	hba1c	167	5.35	0.50	3.8	6.4
rs76723693=1 & rs1039215=0	hba1c	19	5.09	0.39	4.3	6.0
rs76723693=1 & rs1039215=1	hba1c	1	4.90	--	4.9	4.9
rs76723693=2 & rs1039215=0	hba1c	5	4.62	0.13	4.5	4.8

Table S15: Estimated number of African Americans adults with type 2 diabetes in the United States whose diagnosis would be missed due to the common and rare glycose-6-phosphate dehydrogenase (*G6PD*) variants if screened with HbA1c

AFRICAN AMERICAN					
rs1050828					
Sex	**Adjusted HbA1c thresholds[a]**	**Estimated proportion with HbA1c above the adjusted threshold* but below the standard clinical threshold of 6.5 %-units (%) for diabetes diagnosis**	**Genotype**	**Expected proportion with genotype (%)**	**Estimated proportion with missed diabetes diagnosis (%)**
Men	5.62-6.49 %-units	13.59 [11.42-15.75]	T	12.00	1.63
Women	5.62-6.49 %-units	14.19 [11.38-17.00]	TT	1.40	0.20
Women	6.16-6.49 %-units	2.31 [0.13-3.32]	CT	21.10	0.49
				Total (%)	**2.32**
				Total (million)[b]	**0.70**
rs76723693					
Sex	**Adjusted HbA1c thresholds**	**Estimated proportion with HbA1c above the adjusted threshold* but below the standard clinical threshold of 6.5 %-units (%) for diabetes diagnosis**	**Genotype**	**Expected proportion with genotype (%)**	**Estimated proportion with missed diabetes diagnosis (%)**
Men	5.52-6.49 %-units	17.25 [15.03-19.47]	G	0.50	0.09
Women	5.52-6.49 %-units	18.32 [14.86-21.78]	GG	2.5E-04	$4.6x10^{-4}$
Women	6.04-6.49 %-units	3.58 [2.33-4.82]	AG	1.00	0.04
				Total (%)	**0.13**
				Total (million)[b]	**0.04**
rs1050828 & rs76723693					
				Total (%)	**2.45**
				Total (million)[b]	**0.74**

[a]Adjusted HbA1c thresholds to define diabetes were calculated by subtracting the sex and genotype specific effect of each variant from the standard clinical threshold of 6.5 %-units.

[b]The estimated number of African American adults in the United States is 30.14 million (2016 United States Census Bureau).

We restricted the NHANES analytic sample to 1,227 adults aged ≥ 18 years, who self-identified as Non-Hispanic Black.

In African Americans, the mean age was 44.2 (42.9, 45.6) years, 55.4% (52.5, 58.3) were women, and mean HbA1c, excluding those with physician-diagnosed diabetes by self-report or medication use, was 5.56 (5.52, 5.59) %-units. Estimated proportion with diagnosed diabetes = 14.56 (11.73, 17.9) %. Estimated proportion with undiagnosed diabetes by HbA1c ≥ 6.5%-units = 2.56 (1.89, 3.24) %.

Table S16: Estimated number of Hispanic adults with type 2 diabetes in the United States whose diagnosis would be missed due to the common and rare glycose-6-phosphate dehydrogenase (*G6PD*) variants if screened with HbA1c

HISPANIC					
rs1050828					
Sex	**Adjusted HbA1c thresholds[a]**	**Estimated proportion with HbA1c above the adjusted threshold* but below the standard clinical threshold of 6.5 %-units (%) for diabetes diagnosis**	**Genotype**	**Expected proportion with genotype (%)**	**Estimated proportion with missed diabetes diagnosis (%)**
Men	5.62-6.49 %-units	11.14 [9.28-13.01]	T	2.00	0.22
Women	5.62-6.49 %-units	10.56 [8.70-12.42]	TT	4.0E-02	$4.0x10^{-3}$
Women	6.16-6.49 %-units	0.11 [0.54-1.68]	CT	3.90	0.04
				Total (%)	**0.26**
				Total (million)[b]	**0.10**

[a]Adjusted HbA1c thresholds to define diabetes were calculated by subtracting the sex and genotype specific effect of each variant from the standard clinical threshold of 6.5 %-units.

[b]The estimated number of Hispanic adults in the United States is 39.33 million (2016 United States Census Bureau).

We restricted the NHANES analytic sample to 1,768 adults aged ≥ 18 years, who self-identified as Mexican American/Other Hispanic.

In Mexican Americans/Other Hispanics, the mean age was 41.3 (39.4, 43.1) years, 50.4 (48.4, 52.3)% were women, and mean HbA1c was 5.50 (5.45, 5.56) %-units. Estimated proportion with physician-diagnosed diabetes = 13.46 (10.09, 16.83) &. Estimated proportion with undiagnosed diabetes by HbA1c ≥ 6.5%-units = 2.33 (1.39, 3.28) %.

Table S17: Association results of the TOPMed sub-genome signals in the UK Biobank

					African-ancestry (N=5,311)				European-ancestry (N=398,122)			
MarkerID[a]	**rsID**	**Gene**	**Alternate Allele**	**Imputation Info Score**	**Allele Frequency**	**Beta**	**SE**	***P value***	**Allele Frequency**	**Beta**	**SE**	***P value***
1:69561493	rs546782425	*LRRC7*	A	0.70	0.01	0.019	0.048	7.00E-01	--	--	--	--
2:168900844	rs1402837	*G6PC2*	T	1.00	--	--	--	--	0.23	0.023	0.001	4.46E-116
3:5940489	rs541283713	*EDEM1*	T	0.72	0.004	0.022	0.078	7.83E-01	--	--	--	--
3:73834413	rs73103979	*PDZRN3*	T	1.00	--	--	--	--	0.25	0.000	0.001	9.25E-01
4:100766435	rs184930483	*EMCN*	--	--	--	--	--	--	--	--	--	--
5:73310829	rs148873180	*FOXD1*	A	0.89	0.01	-0.036	0.041	3.74E-01	--	--	--	--
7:117730753	rs77974976	*CTTNBP2*	T	0.84	0.008	-0.123	0.051	1.57E-02	--	--	--	--
7:117730990	--	*CTTNBP2*	--	--	--	--	--	--	--	--	--	--
7:117734487	rs115069847	*CTTNBP2*	G	0.83	0.008	-0.086	0.051	8.94E-02	--	--	--	--
8:57797437	rs182571425	*FAM110B*	G	0.86	--	--	--	--	0.003	-0.003	0.008	7.05E-01
8:59954108	rs148165085	*CA8*	G	0.91	0.008	0.038	0.051	4.56E-01	--	--	--	--
10:53414779	rs185736000	*PCDH15*	C	0.59	--	--	--	--	--	--	--	--
10:108997093	rs551601853	*XPNPEP1*	G	0.64	5.28E-04	-0.073	0.249	7.69E-01	--	--	--	--
11:5078417	rs77362408	*OR52E2*	G	0.95	0.02	0.037	0.035	2.84E-01	--	--	--	--
11:5492424	rs73392143	*HBG2/HBE1*	T	0.98	0.01	-0.017	0.042	6.86E-01	--	--	--	--
11:6014322	rs76128436	*OR56A4*	T	0.94	0.02	0.004	0.033	9.04E-01	--	--	--	--
17:32533283	rs140882821	*MYO1D*	A	0.73	0.02	-0.029	0.037	4.42E-01	--	--	--	--
17:78138386	rs144120533	*TMC8*	T	1.00	--	--	--	--	0.003	-0.145	0.008	1.10E-79
18:2770634	--	*SMCHD1*	--	--	--	--	--	--	--	--	--	--
18:34585879	rs187431535	*DTNA*	A	0.92	--	--	--	--	0.005	-0.006	0.006	3.62E-01
18:34712254	rs183229631	*DTNA*	A	0.94	--	--	--	--	0.005	-0.004	0.006	5.62E-01
21:37519947	rs11701836	*DYRK1A*	G	0.99	0.33	0.000	0.009	9.65E-01	0.26	0.000	0.001	8.67E-01
22:37066896	rs855791	*TMPRSS6*	G	0.99	0.87	0.005	0.013	7.19E-01	0.56	-0.024	0.001	7.33E-167
X:153542357	rs143745197	*ATP2B3*	A	0.94	0.007	-0.210	0.044	2.39E-06	--	--	--	--
X:154593327	rs184539426	*CTAG1A*	T	0.55	0.007	-0.045	0.060	4.54E-01	--	--	--	--
X:154877653	rs189305788	*F8*	A	0.64	0.004	-0.183	0.074	1.30E-02	--	--	--	--

Table S17 (continued):

					Asian-ancestry (N=1,331)			
MarkerID[a]	**rsID**	**Gene**	**Alternate Allele**	**Imputation Info Score**	**Allele Frequency**	**Beta**	**SE**	***P value***
1:69561493	rs546782425	*LRRC7*	A	0.70	--	--	--	--
2:168900844	rs1402837	*G6PC2*	T	1.00	--	--	--	--
3:5940489	rs541283713	*EDEM1*	T	0.72	--	--	--	--
3:73834413	rs73103979	*PDZRN3*	T	1.00	--	--	--	--
4:100766435	rs184930483	*EMCN*	C	0.77	0.02	-0.034	0.047	4.69E-01
5:73310829	rs148873180	*FOXD1*	A	0.89	--	--	--	--
7:117730753	rs77974976	*CTTNBP2*	T	0.84	--	--	--	--
7:117730990	--	*CTTNBP2*	--	--	--	--	--	--
7:117734487	rs115069847	*CTTNBP2*	G	0.83	--	--	--	--
8:57797437	rs182571425	*FAM110B*	G	0.86	--	--	--	--
8:59954108	rs148165085	*CA8*	G	0.91	--	--	--	--
10:53414779	rs185736000	*PCDH15*	C	0.59	--	--	--	--
10:108997093	rs551601853	*XPNPEP1*	G	0.64	--	--	--	--
11:5078417	rs77362408	*OR52E2*	G	0.95	--	--	--	--
11:5492424	rs73392143	*HBG2/HBE1*	T	0.98	--	--	--	--
11:6014322	rs76128436	*OR56A4*	T	0.94	--	--	--	--
17:32533283	rs140882821	*MYO1D*	A	0.73	--	--	--	--
17:78138386	rs144120533	*TMC8*	T	1.00	--	--	--	--
18:2770634	--	*SMCHD1*	--	--	--	--	--	--
18:34585879	rs187431535	*DTNA*	A	0.92	--	--	--	--
18:34712254	rs183229631	*DTNA*	A	0.94	--	--	--	--
21:37519947	rs11701836	*DYRK1A*	G	0.99	0.39	-0.029	0.014	4.14E-02
22:37066896	rs855791	*TMPRSS6*	G	0.99	--	--	--	--
X:153542357	rs143745197	*ATP2B3*	A	0.94	--	--	--	--
X:154593327	rs184539426	*CTAG1A*	T	0.55	--	--	--	--
X:154877653	rs189305788	*F8*	A	0.64	--	--	--	--

[a]MarkerID defined as Chromosome:Position. Positions are indicated on Build 38.

Table S18: Effect of the *G6PD*-rs76723693 and the *HBG2*-rs1039215 variants in the whole sample of non-diabetic African-American individuals from TOPMed cohorts and in the two largest African-American cohorts (JHS and MESA)

MarkerID[a]	Ref/Alt Allele	N	EAF[b]	AC[c]	Beta	SE	*P value*	Cohort
11:5464062	A/G	2,356	0.03	139.00	-0.204	0.040	3.74E-07	JHS
11:5464062	A/G	667	0.02	28.00	-0.202	0.078	9.54E-03	MESA
11:5464062	A/G	3,123	0.03	172.00	-0.210	0.035	2.04E-09	JHS+MESA+ARIC
X:154533025	A/G	2,356	0.006	26.28	-0.419	0.091	4.35E-06	JHS
X:154533025	A/G	667	0.004	5.24	-0.484	0.130	2.01E-04	MESA
X:154533025	A/G	3,123	0.005	32.85	-0.436	0.073	2.16E-09	JHS+MESA+ARIC

[a]MarkerID defined as Chromosome:Position. Positions are indicated on Build 38.
[b]EAF: Effect Allele Frequency
[c]AC: Allele Count

Supplemental Methods

Cohort Description

1. **Cohorts included in the HbA1c and RBC analyses**

The Amish study

The Amish Complex Disease Research Program (http://www.medschool.umaryland.edu/endocrinology/Amish-Research-Program/) includes a set of large community-based studies focused largely on cardiometabolic health carried out in the Old Order Amish (OOA) community of Lancaster County, Pennsylvania. Over 7,000 Amish have been recruited to date. This Amish community is a founder population who immigrated to Pennsylvania from Western Europe in the early 1700's, later expanding into other regions of the U.S. The Amish cohort participating in the TOPMed Consortium comprises 1,120 subjects ≥ 18 years of age from large multigenerational families who were recruited for specific protocols between 2001 and 2006. Subjects have been extensively phenotyped for a range of cardiometabolic traits, including anthropometry, lipids, blood pressure, glucose and related measures, vascular imaging, and a range of other phenotypes. DNA samples have been collected and serum and plasma samples biobanked. The TOPMed Program has provided WGS data to complement GWAS array data already collected in >5,000 Amish study participants. Due to their ancestral history, the OOA are enriched for rare exonic variants that arose in the population from a single founder (or small number of founders) and propagated through genetic drift. Many of these variants have large effect sizes, and identifying them can lead to new biological insights about health and disease. A major goal of the TOPMed WGS sequencing efforts is to identify functional variants that underlie some of the large effect associations observed in this unique population.

Complete Blood Counts analyses were performed within 24 h for all individuals and consisted of RBC, HB, HCT, MCV, MCH, MCHC, RDW and WBC. RBC, HB, and MCV were directly measured, whereas HCT, MCH, and MCHC were mathematically derived from directly measured erythrocyte traits (MCV × RBC, HB/RBC, and MCH/MCV, respectively).[1]

The Atherosclerosis Risk in Communities Study

The ARIC study is a population-based prospective cohort study of cardiovascular disease sponsored by the National Heart, Lung, and Blood Institute (NHLBI). ARIC included 15,792 individuals, predominantly European American and African American, aged 45-64 years at baseline (1987-89) and chosen by probability sampling from four US communities. Cohort members completed three additional triennial follow-up examinations, a fifth exam in 2011-2013, and a sixth exam in 2016-2017. The ARIC study has been described in detail previously.[2] Complete Blood Counts was measured using automated hematology analyzers: Coulter S + IV (calibration S - Cal, Beckman Coulter, Inc, Fullerton, CA) at 2 sites, Coulter S + III and Coulter S + IV (calibration S- Cal) at 1 site, and Technicon H-6000 (calibration Fisher, Technicon Corporation, Tarrytown, NY) at 1 site.

The Framingham Heart Study

The Framingham Heart Study (FHS) is a single-site, community-based, prospective cohort study that was initiated in 1948 to investigate risk factors for cardiovascular disease. The population of Framingham was almost entirely white in 1948. The FHS comprises three

generations of participants: the original cohort followed since 1948 (Original or Gen1);[3] their offspring and spouses of the offspring, followed since 1971 (Offspring or Gen2);[4] and children from the largest offspring families enrolled in 2002 (Gen3).[5] The Original cohort enrolled 5,209 men and women who comprised two-thirds of the adult population then residing in Framingham, MA, USA. Survivors continue to receive biennial examinations. The Offspring cohort comprises 5,124 persons (including 3,514 biological offspring) who have been examined approximately once every 4 to 8 years and have completed 9 examinations. The first examination of the Gen3 was completed in July 2005 and included 4,095 participants, a third examination is currently ongoing. All cohorts continue under active surveillance for cardiovascular events. All participants provided written informed consent. This study was approved by the Institutional Review Board of the Boston University Medical Center.

The determination of hematological phenotypes in the FHS has been detailed previously.[6, 7] Hematology testing was performed on a Baker 9000 Hematology Analyzer (Baker Instruments Corp.) for the Original cohort and on a Beckman Coulter HmX Hematology Analyzer (Beckman Coulter, Inc.) for the Offspring and Gen3 cohorts.

The Jackson Heart Study

The Jackson Heart Study (JHS) was designed to study the reasons for the greater prevalence of cardiovascular disease among African Americans and to find new approaches for reducing this health disparity. JHS recruited 5,306 AA participants from urban and rural areas of the three counties (Hinds, Madison and Rankin) that comprise the Jackson, Mississippi metropolitan area from 2000-2004. Recruitment was limited to non-institutionalized adult African Americans. Participants were recruited in four ways: (1) randomly sampling households from a commercial listing; (2) a structured volunteer sample designed to mirror the eligible population; (3) current enrollment in the Atherosclerosis Risk in Communities (ARIC) study; and (4) a nested family cohort. Unrelated participants were between 35 and 84 years old, while members of the family cohort were $\geq$ 21 years old at baseline. A range of measures, including health behaviors, medication use, anthropometry, blood pressure, assessments of kidney function and diabetes, and CVD biomarkers, were assessed at the baseline JHS visit. The JHS has been described in detail previously.[8, 9] Blood cell count metrics were assessed using standard methods (Beckman Coulter automated hematology analyzer), as described previously.[10]

The Multi-Ethnic Study of Atherosclerosis

The Multi-Ethnic Study of Atherosclerosis (MESA) is a study of the characteristics of subclinical cardiovascular disease and the risk factors that predict progression to clinically overt cardiovascular disease or progression of the subclinical disease. MESA consisted of a diverse, population-based sample of an initial 6,814 asymptomatic men and women aged 45-84. 38 percent of the recruited participants were white, 28 percent African American, 22 percent Hispanic, and 12 percent Asian, predominantly of Chinese descent. Participants were recruited from six field centers across the United States: Wake Forest University, Columbia University, Johns Hopkins University, University of Minnesota, Northwestern University and University of California - Los Angeles. Each participant received an extensive physical exam and determination of coronary calcification, ventricular mass and function, flow-mediated endothelial vasodilation, carotid intimal-medial wall thickness and presence of echogenic lucencies in the carotid artery, lower extremity vascular insufficiency,

arterial wave forms, electrocardiographic (ECG) measures, standard coronary risk factors, sociodemographic factors, lifestyle factors, and psychosocial factors. Selected repetition of subclinical disease measures and risk factors at follow-up visits allowed study of the progression of disease. Participants are being followed for identification and characterization of cardiovascular disease events, including acute myocardial infarction and other forms of coronary heart disease (CHD), stroke, and congestive heart failure; for cardiovascular disease interventions; and for mortality. The first examination took place over two years, from July 2000 - July 2002. It was followed by four examination periods that were 17-20 months in length. Participants have been contacted every 9 to 12 months throughout the study to assess clinical morbidity and mortality.

MESA Family

In the MESA Family Study, the goal is to locate and identify genes contributing to the genetic risk for cardiovascular disease (CVD), by looking at the early changes of atherosclerosis within families (mainly siblings). 2128 individuals from 594 families, yielding 3,026 sib pairs divided between African Americans and Hispanic-Americans, were recruited by utilizing the existing framework of MESA. MESA Family studied siblings of index subjects from the MESA study and from new sibpair families (with the same demographic characteristics) and is determining the extent of genetic contribution to the variation in coronary calcium (obtained via CT Scan) and carotid artery wall thickness (B-mode ultrasound) in the two largest non-majority U.S. populations. The MESA Family cohort was recruited from the six MESA Field Centers. MESA Family participants underwent the same examination as MESA participants during May 2004 - May 2007. The MESA Study has been described in detail previously.[11] Complete Blood Counts was performed at LabCorp using automated cell counter with mixed technologies.

2. Cohorts included in the RBC analyses only

The Cardiovascular Health Study

CHS is a population-based cohort study of risk factors for coronary heart disease and stroke in adults ≥65 years conducted across four field centers.[12] The original predominantly European ancestry cohort of 5,201 persons was recruited in 1989-1990 from random samples of the Medicare eligibility lists; subsequently, an additional predominantly African-American cohort of 687 persons was enrolled for a total sample of 5,888. CHS participants selected for inclusion in the TOPMed sequencing program included African-Americans participants, cases of idiopathic venous thromboembolism, myocardial infarction, coronary heart disease or stroke and a random sample of "healthy elderly". Blood samples were drawn from all participants at their baseline examination and DNA was subsequently extracted from available samples. Details of measurement of hematologic traits have been previously described.[13]

CHS was approved by institutional review committees at each field center and individuals in the present analysis had available DNA and gave informed consent including consent to use of genetic information for the study of cardiovascular disease.

The Genetic Studies of Atherosclerosis Risk

GeneSTAR: The Genetic Study of Atherosclerosis Risk (GeneSTAR) is a longitudinal family-based study designed to explore environmental, phenotypic, and genetic causes of premature cardiovascular disease. European- and African-American participants were

recruited from European- and African-American families (n=891) identified from probands hospitalized for a coronary disease event prior to 60 years of age in any of 10 Baltimore, Maryland area hospitals between 1983 and 2006. Apparently healthy siblings of the probands, offspring of the siblings and probands, and the co-parents of the offspring were screened for traditional coronary disease and stroke risk factors between 2003 and 2006 as part of a platelet function study of two weeks of 81 mg/day of aspirin.[14, 15]. All measures described here were obtained prior to the commencement of aspirin. Exclusion criteria included: 1) any coronary heart disease or vascular thrombotic event, 2) any bleeding disorder or hemorrhagic event, 3) current use of any anticoagulants or anti-platelet agents, 4) current use of chronic or acute nonsteroidal anti-inflammatory agents that could not be discontinued, 5) recent active gastrointestinal disorder, 6) current pharmacotherapy for a gastrointestinal disorder, 7) pregnancy or risk of pregnancy during the trial, 8) recent menorrhagia, 9) known aspirin intolerance or allergic side effects, 10) serious medical disorders, (eg, autoimmune diseases, cancer or HIV-AIDS), 11) current chronic or acute use of glucocorticosteroid therapy or any drug that may interfere with the measured outcomes, 12) serious psychiatric disorders, and, 13) inability to independently make a decision to participate. Of 3003 aspirin study participants, 1786 were selected for TOPMed based on 1) complete platelet function phenotyping and 2) largest family size. Complete Cell Counts were obtained using an automated cell counter (ACT-Diff, Beckman Coulter).

The San Antonio Family Study

The San Antonio Family Study results from the amalgamation of two San Antonio-based genetic studies. The first is the longitudinal San Antonio Family Heart Study (SAFHS) which began in 1991 and was designed to primarily investigate the genetics of cardiovascular disease and its risk factors in Mexican Americans. The SAFHS included 1,431 individuals in 42 extended families at baseline.[16] With some additional recruiting, it has now been expanded to 1,662 individuals in 47 families. Ascertainment occurred by way of the random selection of an adult Mexican American proband, without regard to presence or absence of disease and almost exclusively from Mexican American census tracts in San Antonio. The second component study is the San Antonio Family Gall Bladder Study (Dr. Duggirala, PI) which included 907 individuals from 39 families ascertained similarly to the SAFHS but with the requirement that the original proband also be diabetic.[17] This is a very weak form of ascertainment in Mexican Americans, where lifetime prevalence of diabetes approaches 30%. In fact, 20 years after the initiation of the SAFS, the prevalence for major diseases such as heart disease, diabetes, and obesity are not significantly different between these two component studies. Finally, we have expanded these pedigrees in recent years by examining a set of 498 children than are part of these families. This expansion was part of Dr. Duggirala's San Antonio Family Assessment of Metabolic Risk Factors in Youth (SAFARI) study. Additionally, we have seen 112 of these children subsequently as adults. Combined, these studies have 3,099 individuals primarily from 73 families. Our study is a mixed longitudinal design. Subjects have been seen between 1 and 4 times with an average of 1.95 examinations. For this study, samples and data analyzed were from the San Antonio Family Heart Study component. Blood cell counts were obtained within 2hrs of blood collection from an EDTA tube using a Beckman Coulter AcT diff2 Hematology Analyzer.

The Women's Health Initiative

The Women's Health Initiative (WHI) is a long-term, prospective, multi-center cohort study investigating post-menopausal women's health in the US.[18] WHI was funded by the National Institutes of Health and the National Heart, Lung, and Blood Institute to study strategies to prevent heart disease, breast cancer, colon cancer, and osteoporotic fractures in women 50-79 years of age. WHI involves 161,808 women recruited between 1993 and 1998 at 40 centers across the US. The study consists of two parts: the WHI Clinical Trial which was a randomized clinical trial of hormone therapy, dietary modification, and calcium/Vitamin D supplementation, and the WHI Observational Study, which focused on many of the inequities in women's health research and provided practical information about incidence, risk factors, and interventions related to heart disease, cancer, and osteoporotic fractures. Fasting blood samples were drawn and analyzed for RBC traits at designated clinical laboratories using an automated electronic cell counter at the baseline examination.[19] Participants with HCT>80, HGB>30, RBC=0 or HCT/HGB>7 were excluded. In total, 9912, 9909 and 1285 participants were included for HGB, HCT and RDW analyses, respectively, and 1286 participants were included for MCH, MCHC, MCV and RBC counts analyses.

Supplemental Acknowledgements

Whole genome sequencing:

WGS for "NHLBI TOPMed: Genetics of Cardiometabolic Health in the Amish (phs000956.v1.p1) was performed at the Broad Institute of MIT and Harvard (3R01HL121007-01S1). WGS for "NHLBI TOPMed: Trans-Omics for Precision Medicine Whole Genome Sequencing Project: ARIC (phs001211.v1.p1) was performed at the Broad Institute of MIT and Harvard and at the Baylor Human Genome Sequencing Center (3R01HL092577-06S1 (Broad, AFGen), HHSN268201500015C (Baylor, VTE), 3U54HG003273-12S2 (Baylor, VTE)). WGS for "NHLBI TOPMed: Whole Genome Sequencing and Related Phenotypes in the Framingham Heart Study (phs000974.v1.p1) was performed at the Broad Institute of MIT and Harvard (3R01HL092577-06S1 (AFGen)). WGS for "NHLBI TOPMed: The Jackson Heart Study (phs000964.v1.p1) was performed at the University of Washington Northwest Genomics Center (HHSN268201100037C). WGS for "NHLBI TOPMed: MESA and MESA Family AA-CAC (phs001416) was performed at the Broad Institute of MIT and Harvard (3U54HG003067-13S1 (MESA, TOPMed supplement to NHGRI), HHSN268201500014C (Broad, AA_CAC)). WGS for "NHLBI TOPMed: Cardiovascular Health Study (phs001368.v1.p1)" was performed at Baylor Human Genome Sequencing Center (HHSN268201500015C, VTE portion of CHS). WGS for "NHLBI TOPMed: GeneSTAR (Genetic Study of Atherosclerosis Risk) (phs001218.v1.p1)" was performed at Macrogen Corp and at the Broad Institute of MIT and Harvard (HHSN268201500014C (Broad, AA_CAC)). WGS for "NHLBI TOPMed: Women's Health Initiative (WHI) (phs001237.v1.p1)" was performed at the Broad Institute of MIT and Harvard (HHSN268201500014C). WGS for "NHLBI TOPMed: San Antonio Family Heart Study (WGS) (phs001215.v1.p1)" was performed at Illumina Genomic Services (3R01HL113323-03S1).

The Amish study

NIH grant R01 HL121007

The Atherosclerosis Risk in Communities Study

Whole genome sequencing (WGS) for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the National Heart, Lung and Blood Institute (NHLBI). WGS for “NHLBI TOPMed: Atherosclerosis Risk in Communities (ARIC)” (phs001211) was performed at the Baylor College of Medicine Human Genome Sequencing Center (HHSN268201500015C and 3U54HG003273-12S2) and the Broad Institute for MIT and Harvard (3R01HL092577-06S1). Centralized read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1). Phenotype harmonization, data management, sample-identity QC, and general study coordination, were provided by the TOPMed Data Coordinating Center (3R01HL-120393-02S1). We gratefully acknowledge the studies and participants who provided biological samples and data for TOPMed.

The Atherosclerosis Risk in Communities study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services (contract numbers HHSN268201700001I,

HHSN268201700002I, HHSN268201700003I, HHSN268201700004I and HHSN268201700005I). The authors thank the staff and participants of the ARIC study for their important contributions.

The Cardiovascular Health Study

This research was supported by contracts HHSN268201200036C, HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086, and grants R01HL120393, U01HL080295 and U01HL130114 from the National Heart, Lung, and Blood Institute (NHLBI), with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided by R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org.

The Framingham Heart Study

This research was conducted in part using data and resources from the Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine. The Framingham Heart Study (FHS) acknowledges the support of contracts NO1-HC-25195 and HHSN268201500001I from the National Heart, Lung and Blood Institute, contract 75N92019D00031 and grant supplement R01 HL092577-06S1 for this research. This work was also supported in part by grant U01DK078616.

We also acknowledge the dedication of the FHS study participants without whom this research would not be possible.

The Genetic Studies of Atherosclerosis Risk

GeneSTAR was supported by the National Institutes of Health/National Heart, Lung, and Blood Institute (U01 HL72518, HL087698, HL112064, HL11006, HL118356) and by a grant from the National Institutes of Health/National Center for Research Resources (M01-RR000052) to the Johns Hopkins General Clinical Research Center. We would like to thank the participants and families of GeneSTAR and our dedicated staff for all their sacrifices.

The Jackson Heart Study

The Jackson Heart Study (JHS) is supported and conducted in collaboration with Jackson State University (HHSN268201800013I), Tougaloo College (HHSN268201800014I), the Mississippi State Department of Health (HHSN268201800015I/HHSN26800001) and the University of Mississippi Medical Center (HHSN268201800010I, HHSN268201800011I and HHSN268201800012I) contracts from the National Heart, Lung, and Blood Institute (NHLBI) and the National Institute for Minority Health and Health Disparities (NIMHD). The authors also wish to thank the staffs and participants of the JHS.

The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services.

The Multi-Ethnic Study of Atherosclerosis

Whole genome sequencing (WGS) for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the National Heart, Lung and Blood Institute (NHLBI). WGS for "NHLBI TOPMed: Multi-Ethnic Study of Atherosclerosis (MESA)" (phs001416.v1.p1)

was performed at the Broad Institute of MIT and Harvard (3U54HG003067-13S1). Centralized read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1). Phenotype harmonization, data management, sample-identity QC, and general study coordination, were provided by the TOPMed Data Coordinating Center (3R01HL-120393-02S1). MESA and the MESA SHARe project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

The San Antonio Family Study

Collection of the San Antonio Family Study data was supported in part by National Institutes of Health (NIH) grants R01 HL045522, DK047482, DK053889, MH078143, MH078111 and MH083824; and whole genome sequencing of SAFS subjects was supported by U01 DK085524 and R01 HL113323. This work was conducted in part in facilities constructed under the support of NIH grant C06 RR020547. We are very grateful to the participants of the San Antonio Family Study for their continued involvement in our research programs.

The Women's Health Initiative

The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201600018C, HHSN268201600001C, HHSN268201600002C, HHSN268201600003C, and HHSN268201600004C. The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: http://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Long%20List.pdf

The UK Biobank

This research has been conducted using the UK Biobank Resource under Application Number 42614.

References

1. Hinckley, J.D., Abbott, D., Burns, T.L., Heiman, M., Shapiro, A.D., Wang, K., Di Paola, J. (2013). Quantitative trait locus linkage analysis in a large Amish pedigree identifies novel candidate loci for erythrocyte traits. Mol. Genet. Genomic Med. *1*, 131-141.

2. Anonymous (1989). The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. Am. J. Epidemiol. *129*, 687-702.

3. Dawber, T.R., Kannel, W.B. (1966). The Framingham study. An epidemiological approach to coronary heart disease. Circulation *34*, 553-555.

4. Feinleib, M., Kannel, W.B., Garrison, R.J., McNamara, P.M., Castelli, W.P. (1975). The Framingham Offspring Study. Design and preliminary data. Prev. Med. *4*, 518-525.

5. Splansky, G.L., Corey, D., Yang, Q., Atwood, L.D., Cupples, L.A., Benjamin, E.J., D'Agostino RB, S., Fox, C.S., Larson, M.G., Murabito, J.M. et al. (2007). The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. Am. J. Epidemiol. *165*, 1328-1335.

6. Yang, Q., Kathiresan, S., Lin, J.P., Tofler, G.H., O'Donnell, C.J. (2007). Genome-wide association and linkage analyses of hemostatic factors and hematological phenotypes in the Framingham Heart Study. BMC Med. Genet. *8 Suppl 1*, S12-2350-8-S1-S12.

7. Sloan, A., Gona, P., Johnson, A.D. (2015). Cardiovascular correlates of platelet count and volume in the Framingham Heart Study. Ann. Epidemiol. *25*, 492-498.

8. Taylor, H.A.,Jr, Wilson, J.G., Jones, D.W., Sarpong, D.F., Srinivasan, A., Garrison, R.J., Nelson, C., Wyatt, S.B. (2005). Toward resolution of cardiovascular health disparities in African Americans: design and methods of the Jackson Heart Study. Ethn. Dis. *15*, S6-4-17.

9. Wilson, J.G., Rotimi, C.N., Ekunwe, L., Royal, C.D., Crump, M.E., Wyatt, S.B., Steffes, M.W., Adeyemo, A., Zhou, J., Taylor, H.A.,Jr et al. (2005). Study design for genetic analysis in the Jackson Heart Study. Ethn. Dis. *15*, S6-30-37.

10. Carpenter, M.A., Crow, R., Steffes, M., Rock, W., Heilbraun, J., Evans, G., Skelton, T., Jensen, R., Sarpong, D. (2004). Laboratory, reading center, and coordinating center data management methods in the Jackson Heart Study. Am. J. Med. Sci. *328*, 131-144.

11. Bild, D.E., Bluemke, D.A., Burke, G.L., Detrano, R., Diez Roux, A.V., Folsom, A.R., Greenland, P., Jacob, D.R.,Jr, Kronmal, R., Liu, K. et al. (2002). Multi-Ethnic Study of Atherosclerosis: objectives and design. Am. J. Epidemiol. *156*, 871-881.

12. Fried, L.P., Borhani, N.O., Enright, P., Furberg, C.D., Gardin, J.M., Kronmal, R.A., Kuller, L.H., Manolio, T.A., Mittelmark, M.B., Newman, A. (1991). The Cardiovascular Health Study: design and rationale. Ann. Epidemiol. *1*, 263-276.

13. Ganesh, S.K., Zakai, N.A., van Rooij, F.J., Soranzo, N., Smith, A.V., Nalls, M.A., Chen, M.H., Kottgen, A., Glazer, N.L., Dehghan, A. et al. (2009). Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. Nat. Genet. *41*, 1191-1198.

14. Bray, P.F., Mathias, R.A., Faraday, N., Yanek, L.R., Fallin, M.D., Herrera-Galeano, J.E., Wilson, A.F., Becker, L.C., Becker, D.M. (2007). Heritability of platelet function in families with premature coronary artery disease. J. Thromb. Haemost. *5*, 1617-1623.

15. Faraday, N., Yanek, L.R., Mathias, R., Herrera-Galeano, J.E., Vaidya, D., Moy, T.F., Fallin, M.D., Wilson, A.F., Bray, P.F., Becker, L.C. et al. (2007). Heritability of platelet responsiveness to aspirin in activation pathways directly and indirectly related to cyclooxygenase-1. Circulation *115*, 2490-2496.

16. Mitchell, B.D., Kammerer, C.M., Blangero, J., Mahaney, M.C., Rainwater, D.L., Dyke, B., Hixson, J.E., Henkel, R.D., Sharp, R.M., Comuzzie, A.G. et al. (1996). Genetic and environmental contributions to cardiovascular risk factors in Mexican Americans. The San Antonio Family Heart Study. Circulation *94*, 2159-2170.

17. Duggirala, R., Blangero, J., Almasy, L., Dyer, T.D., Williams, K.L., Leach, R.J., O'Connell, P., Stern, M.P. (1999). Linkage of type 2 diabetes mellitus and of age at onset to a genetic location on chromosome 10q in Mexican Americans. Am. J. Hum. Genet. *64*, 1127-1140.

18. Anonymous (1998). Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. Control. Clin. Trials *19*, 61-109.

19. Chen, Z., Tang, H., Qayyum, R., Schick, U.M., Nalls, M.A., Handsaker, R., Li, J., Lu, Y., Yanek, L.R., Keating, B. et al. (2013). Genome-wide association analysis of red blood cell traits in African Americans: the COGENT Network. Hum. Mol. Genet. *22*, 2529-2538.