Electronic Supplementary Materials

Methods

Study population

The multi-ethnic PAGE Study [1] is a consortium funded by the National Institutes of Health to examine the genetic architecture of common complex diseases and traits in diverse populations. PAGE data used in this analysis included participants enrolled in the following cohort studies:

Atherosclerosis Risk in Communities Study (ARIC)

ARIC is a multi-centre prospective epidemiologic cohort study funded by the National Heart, Lung, and Blood Institute (NHLBI) to investigate the causes of atherosclerosis and its clinical outcomes, and variation in cardiovascular risk factors, medical care, and disease by race, gender, location, and date [2]. In total, 15,792 participants ages 45-64 of primarily European American and African American descent were recruited between 1987 and 1989 from four communities in the United States: Washington County, MD; Forsyth County, NC; Jackson, MS; and Minneapolis, MN. At study baseline (1987-1989), participants received standardized physical examinations and interviewer-administered questionnaires. Semi-annual telephone follow-up calls are ongoing to maintain contact and assess health status of the cohort.

The BioMe^{TM} Biobank Program (BioMe)

The Charles Bronfman Institute for Personalized Medicine at Mount Sinai Medical Center (MSMC), BioMe^{TM} BioBank (BioMe) is an EMR-linked bio-repository drawing from Mount Sinai Medical Center consented patients which were drawn from a population of over 70,000 inpatients and 800,00 outpatients annually from diverse local communities in upper Manhattan (https://icahn.mssm.edu/research/ipm/programs/biome-biobank). Data on

anthropometrics, demographics, and medication use were derived from participants' EMR and a medical history questionnaire administered at baseline [3].

The Coronary Artery Risk Development in Young Adults Study (CARDIA)

CARDIA is a multi-centre prospective cohort study funded by the NHLBI to study the development and distribution of cardiovascular diseases and their risk factors [4]. A total of 5,115 participants ages 18-30 years (52% African American, 55% women) were recruited in 1985-1986 from four communities in the United States: Birmingham, AL; Chicago, IL; Minneapolis, MN; and Oakland, CA. The participants were selected so that there would be approximately the same number of people in subgroups of race, gender, education (high school or less/more than high school) and age (18-24 and 25-30) in each of these 4 centres. Participants were asked to participate in follow-up examinations during 1987-1988 (Year 2), 1990-1991 (Year 5), 1992-1993 (Year 7), 1995-1996 (Year 10), 2000-2001 (Year 15), and 2005-2006 (Year 20), 2010-2011 (Year 25), and 2015-2016 (Year 35). Data have been collected on factors believed to be related to heart disease, including blood pressure, cholesterol and other lipids, and glucose as well as physical measurements such as weight and skinfold fat, as well as lifestyle factors, behavioural and psychological variables, and medical and family history.

The Multiethnic Cohort Study (MEC)

MEC is a prospective cohort study funded by the National Cancer Institute (NCI) to study diet and cancer in the United States [5]. In total, 215,251 participants living in Hawaii and California (primarily Los Angeles County) ages 45-75 years were recruited between 1993 and 1996 (16.3% African American, 22.0% Latino, 26.4% Japanese-American, 6.5% Native Hawaiian, 22.9% white). Upon recruitment, participants completed a self-administered questionnaire on demographic, dietary, and lifestyle traits. Biological specimens were also

collected from over 70,000 MEC members in 2001-2005. There were seven ancillary MEC sub-studies included in this analysis: the Slim Initiative in Genomic Medicine for the Americas (MEC-Sigma), a type 2 diabetes study in Hispanic/Latino adults; MEC-AAPC, -JAPC, and -LAPC, studies of prostate cancer in African American, Asian, and Hispanic/Latino men, respectively; and MEC-AABC, -JABC, and LABC, studies of breast cancer in African American, Asian, and Hispanic/Latina women, respectively.

The Hispanic Community Health Study/Study of Latinos (HCHS/SOL)

HCHS/SOL is a multi-centre, community-based cohort study of Hispanic/Latinos in the United States to identify the prevalence of and risk factors for multiple chronic diseases, including cardiovascular diseases, lung, kidney and liver diseases [6]. Over 16,000 participants ages 18-74 years were recruited from four communities in the United States between 2008 and 2011: Bronx, NY; Chicago, IL; Miami, FL; and San Diego, CA. These recruitment sites were selected so that the overall sample would include at least 2,000 people from each of the following origin designations: Mexican, Puerto Rican and Dominican, Cuban, and Central and South American. Households were selected via a two-stage sampling within census block groups [7]. At baseline (2008-2011), participants received standardized examinations and interviewer-administered questionnaires. A re-examination of the HCHS/SOL cohort was conducted during 2015-2017, and annual telephone follow-up interviews are ongoing since study inception to determine health outcomes of interest. For this study, HCHS/SOL participants were genotyped on the multi-ethnic genotyping array (MEGA).

The Women's Health Initiative Study (WHI)

WHI is a long-term prospective study to investigate causes of morbidity and mortality among postmenopausal women in the United States [8]. Between 1993 and 1998, 161,808

women ages 50-79 years were recruited from 40 clinical centres and enrolled in randomized clinical trials or an observational cohort study. Women in the observational study received a standardized examination at baseline and interviewer-administered questionnaires. The following ancillary studies were included in this analysis: the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO), the Modification of PM-Mediate Arrhythmogenesis in Populations study (MOPMAP), the Genomics and Randomized Trials Networks (GARNET), the Hip Fracture GWAS (HIPFX), the Long Life Study (LLS), the Women's Health Initiative Memory Study (WHIMS), and the Women's Health Initiative-SNP Health Association Resource (WHI-SHARe). The GECCO study aims to investigate the genetic susceptibility of colorectal cancer, including common and rare loci, gene-environment interactions, and survival. The MOPMAP study aims to investigate the susceptibility to arrhythmogenic effects of particulate matter air pollution contributed by common genetic and environmental variation. The GARNET study is a series of genome-wide association studies of treatment response in randomized clinical trials, aiming to identify genetic variants associated with response to treatments for conditions of clinical or public health significance. The HIPX study was designed to perform epidemiological studies of hip fracture in women. The LLS study included 7,875 women from the WHI Extension II Medical Records Cohort (MRC). The LLS consisted of a one-time in-person visit (between March 2012 and May 2013) with a blood draw, a brief clinical assessment, and an assessment of functional status. The WHIMS study is a trial to examine the effect of oestrogen therapy in preventing and slowing the progression of dementia. The WHI-SHARe study is part of NHLBI's SNP Health Association Resource (SHARe) project, aiming to enhance the statistical power for research specific to groups defined by race and ethnicity and to discover or replicate genes associated with quantitative traits (such as blood pressure and blood lipids) in

these groups. The participants in GECCO, MOPMAP, GARNET, HIPFX, LLS and WHIMS are self-reported European Americans, while the participants in WHI-SHARe are self-reported African Americans.

Trait measurement

Genotyping and imputation

Genotyping platform, quality control, and imputation methods are specified for each PAGE substudy and replication study in **ESM Table 1**.

Continuous BMI measurement

In ARIC, CARDIA, HCHS/SOL, and WHI, BMI was calculated from height and weight measured at time of study enrolment in a clinic setting. In WHI only, measurements collected 1 or 3 years after enrolment were substituted for 140 participants missing enrolment height and/or weight. In MEC and Bio*Me*, self-reported height and weight were used to calculate baseline BMI.

Smoking status measurement

Self-reported smoking status was harmonised across studies as current versus former/never.

Race/ethnicity

In all studies, self-reported race/ethnicity was collected via epidemiological questionnaires at baseline visits.

Replication

Replication of novel loci for fasting glucose, fasting insulin, and HbA_{1c} was performed in the four following studies, using a common analysis plan:

Jackson Heart Study (JHS)

The JHS is a longitudinal, population-based cohort designed for prospective research into the epidemiology and determinants of cardiovascular disease (CVD) in African American populations from the Jackson, Mississippi metropolitan area [9]. The design and sampling of JHS initially only included participants from the Jackson cohort of the ARIC study, random and family components. To reach recruitment goals and address community concerns regarding broad participation, an additional structured volunteer sample was added later [10]. Study baseline data collection began in late 2000 and was completed in early 2004, and a total of 5,306 male and female participants were recruited. The baseline examination included a home interview, a clinic visit, laboratory tests, complete blood cell counts, and a physical examination. JHS participants who were included in the ARIC primary analysis were also excluded from replication analyses.

Cameron County Hispanic Cohort (CCHC)

The CCHC is a randomly-ascertained, community-based cohort of Mexican Americans recruited from border communities on the Texas-Mexico border [11]. Established in 2004, this large cohort study, currently numbering around 5,000 individuals, documents sociodemographic, clinical, behavioural, and biologic characteristics of Cameron County Mexican Americans. Households were randomly identified by US census tract/block and members were invited to participate in the study. The baseline examination included physical examination, collection of biospecimens, and completion of questionnaires.

Reasons for Geographical And Racial Differences in Stroke (REGARDS) Study

The REGARDS cohort is a national, population-based, longitudinal study of 30,239 African American and white adults aged ≥ 45 years recruited from the 48 contiguous United States between 2003 and 2007 [12]. The REGARDS cohort was devised to study racial and

geographic differences in stroke mortality in the United States. Participants were community-dwellers who self-identified as white or African American and were enrolled in 2003-2007 via a stratified random sampling approach that balanced race, sex, and geographic location, with planned oversampling of the Southeastern United States, where the burden of cardiovascular disease is high. All participants completed a 45-minute baseline telephone interview ascertaining details of medical history, followed by an in-home visit collecting blood and urine samples as well as physiologic data (blood pressure, height and weight, electrocardiogram).

Multi-Ethnic Study of Atherosclerosis (MESA)

MESA was initiated in 2000 to investigate subclinical cardiovascular disease and the risk factors that predict progression to clinically overt cardiovascular disease in a multi-ethnic US population free of clinically-recognized cardiovascular disease at study baseline [13]. The population-based cohort recruited 6,814 men and women of European American (38%), African American (28%), Hispanic (22%), and Chinese American (12%) descent, 45-84 years of age from six field centres (Winston-Salem, NC; St. Paul, MN; Chicago, IL; Los Angeles, CA; New York, NY; Baltimore, MD). The baseline examination, which was designed to be the most comprehensive of all examinations, included a physical examination, biospecimen collection, completion of questionnaires, and several heart monitoring procedures.

China Health and Nutrition Survey (CHNS)

Publicly available summary statistics on fasting insulin associations in the CHNS were also included for replication [14]. The CHNS is a nationwide, longitudinal survey examining health, sociological, economic, and demographic questions in a Chinese population across 9 diverse provinces [15]. A stratified probability sample with multistage, random cluster design

was used to select 228 communities stratified by income and urbanicity, and a total of 4,560 households were subsequently selected from within each stratum.

Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC)

Publicly available data on glycaemic traits were also contributed by MAGIC investigators and were downloaded from www.magicinvestigators.org [16]. MAGIC is a collaborative effort to combine data from multiple GWAS to identify additional loci that impact glycaemic and metabolic traits. As described in Lagou et al [16], the summary statistics used for this replication analysis were derived from a meta-analysis of 38 GWAS, including up to 80,512 individuals genotyped using either Illumina or Affymetrix genome-wide SNP arrays, 27 studies with up to 47,150 individuals genotyped using the iSELECT Metabochip array (~197 K SNPs), 8 studies including up to 27,173 individuals genotyped from custom variant sets, and 4 studies including up to 13,613 individuals from four family-based studies. All participants were of European ancestry.

Because self-reported European American participants from the ARIC study were also included in the Lagou et al, we excluded them from the Lagou et al MAGIC meta-analysis results through the following steps. First, we converted our ARIC EA fasting insulin measures from μU/mL to pmol/L (1 μU/mL = 6 pmol/L) to match the scale used in the Lagou et al. MAGIC analyses, and then performed association analyses on natural log-transformed fasting insulin among non-diabetic participants, adjusting for age, study site, and ancestral principal components to align with the analyses performed in MAGIC. We subsequently used the R package MetaSubtract version 1.60 (https://cran.r-project.org/web/packages/MetaSubtract/) [17] to remove the ARIC EA cohort results from the MAGIC meta-analysis, using our PAGE ARIC EA association results as a proxy for the ARIC EA cohort results included in MAGIC. Because

the MAGIC meta-analysis results specified separate genomic control values (λ) for men and women ($\lambda_{FI, M}$ = 0.997, $\lambda_{FI, F}$ = 1.009) [16], we ran MetaSubtract using each λ value, and the final re-derived meta-analysis results were nearly identical. All reported Lagou et al MAGIC replication results were derived using the male λ value, which resulted in slightly more conservative *p* value estimates. Genomic control correction was not applied to the re-derived Lagou et al MAGIC meta-analysis results excluding ARIC EA. Finally, we rescaled the MAGIC association results to the inverse rank normalized residual scale, for consistency with all other replication analyses, before including in the meta-analysis of replication results.

References

[1] Matise TC, Ambite JL, Buyske S, et al. (2011) The Next PAGE in understanding complex traits: design for the analysis of Population Architecture Using Genetics and Epidemiology (PAGE) Study. American journal of epidemiology 174(7): 849-859. 10.1093/aje/kwr160

[2] The ARIC Investigators (1989) The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. . American journal of epidemiology 129(4): 687-702

[3] Gottesman O, Kuivaniemi H, Tromp G, et al. (2013) The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. Genet Med 15(10): 761-771. 10.1038/gim.2013.72

[4] Friedman GD, Cutter GR, Donahue RP, et al. (1988) CARDIA: study design, recruitment, and some characteristics of the examined subjects. J Clin Epidemiol 41(11): 1105-1116. 10.1016/0895-4356(88)90080-7

[5] Kolonel LN, Henderson BE, Hankin JH, et al. (2000) A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. American journal of epidemiology 151(4): 346-357. 10.1093/oxfordjournals.aje.a010213

[6] Sorlie PD, Aviles-Santa LM, Wassertheil-Smoller S, et al. (2010) Design and implementation of the Hispanic Community Health Study/Study of Latinos. Annals of epidemiology 20(8): 629-641. 10.1016/j.annepidem.2010.03.015

[7] Lavange LM, Kalsbeek WD, Sorlie PD, et al. (2010) Sample design and cohort selection in the Hispanic Community Health Study/Study of Latinos. Annals of epidemiology 20(8): 642-649. 10.1016/j.annepidem.2010.05.006

[8] The Women's Health Initiative Study Group (1998) Design of the Women's Health Initiative clinical trial and observational study. Control Clin Trials 19(1): 61-109. 10.1016/s0197-2456(97)00078-0

[9] Taylor HA, Jr., Wilson JG, Jones DW, et al. (2005) Toward resolution of cardiovascular health disparities in African Americans: design and methods of the Jackson Heart Study. Ethn Dis 15(4 Suppl 6): S6-4-17

[10] Fuqua SR, Wyatt SB, Andrew ME, et al. (2005) Recruiting African-American research participation in the Jackson Heart Study: methods, response rates, and sample description. Ethn Dis 15(4 Suppl 6): S6-18-29

[11] Fisher-Hoch SP, Rentfro AR, Salinas JJ, et al. (2010) Socioeconomic status and prevalence of obesity and diabetes in a Mexican American community, Cameron County, Texas, 2004-2007. Prev Chronic Dis 7(3): A53

[12] Howard VJ, Cushman M, Pulley L, et al. (2005) The reasons for geographic and racial differences in stroke study: objectives and design. Neuroepidemiology 25(3): 135-143. 10.1159/000086678

[13] Bild DE, Bluemke DA, Burke GL, et al. (2002) Multi-Ethnic Study of Atherosclerosis: objectives and design. Am J Epidemiol 156(9): 871-881. 10.1093/aje/kwf113

[14] Spracklen CN, Shi J, Vadlamudi S, et al. (2018) Identification and functional analysis of glycemic trait loci in the China Health and Nutrition Survey. PLoS genetics 14(4): e1007275. 10.1371/journal.pgen.1007275

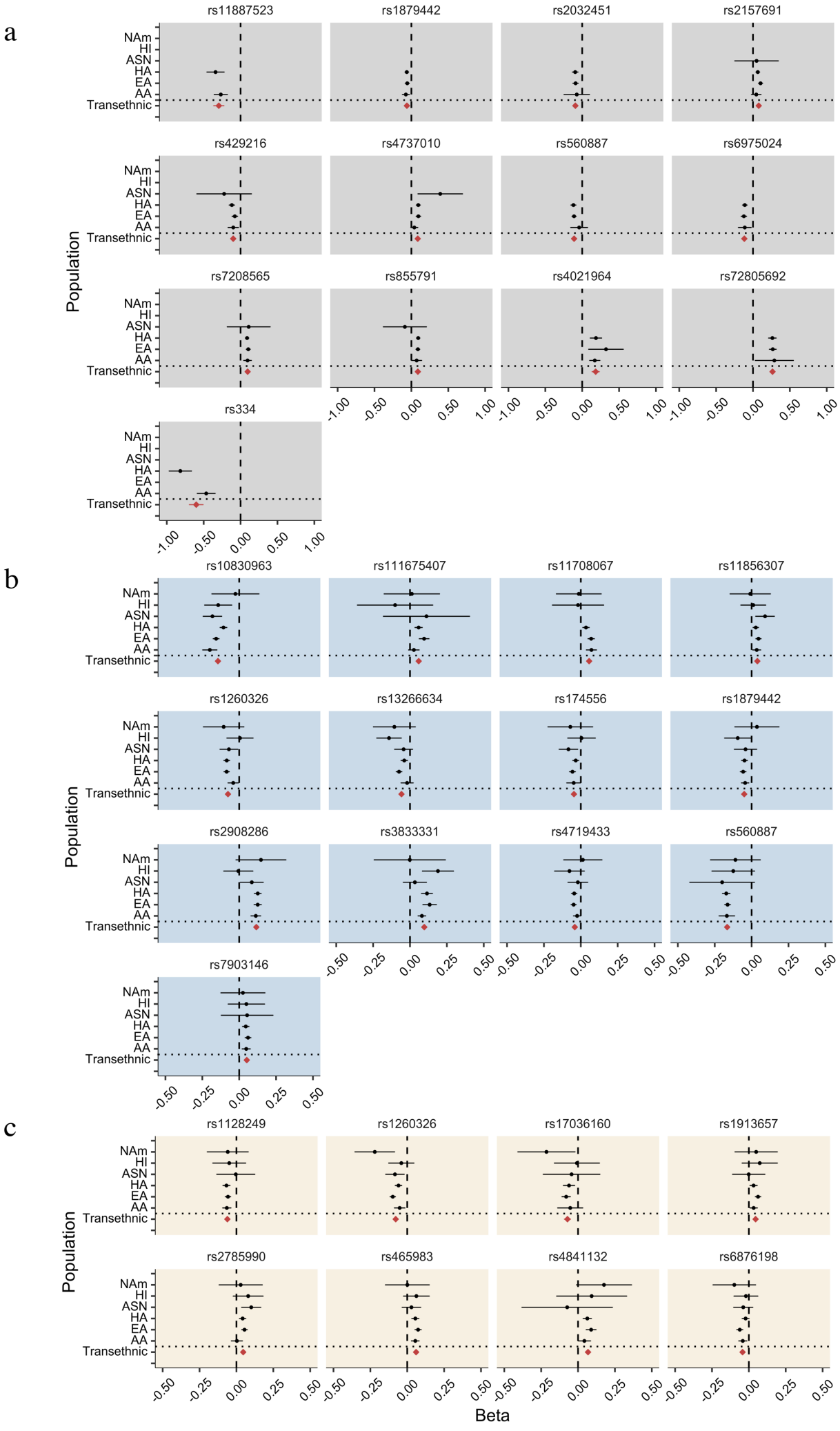
[15] Popkin BM, Du S, Zhai F, Zhang B (2010) Cohort Profile: The China Health and Nutrition Survey--monitoring and understanding socio-economic and health change in China, 1989-2011. International journal of epidemiology 39(6): 1435-1440. 10.1093/ije/dyp322

[16] Lagou V, Magi R, Hottenga JJ, et al. (2021) Sex-dimorphic genetic effects and novel loci for fasting glucose and insulin variability. Nature communications 12(1): 24. 10.1038/s41467-020-19366-9
[17] Nolte IM (2020) Metasubtract: an R-package to analytically produce leave-one-out meta-analysis GWAS summary statistics. Bioinformatics 36(16): 4521-4522. 10.1093/bioinformatics/btaa570

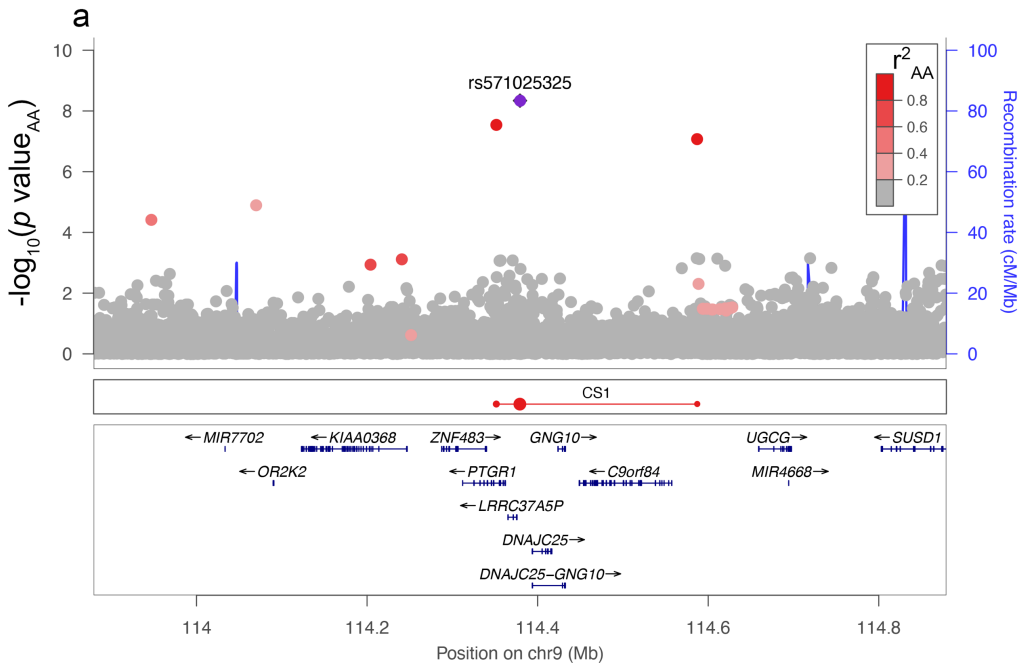
Electronic Supplementary Tables

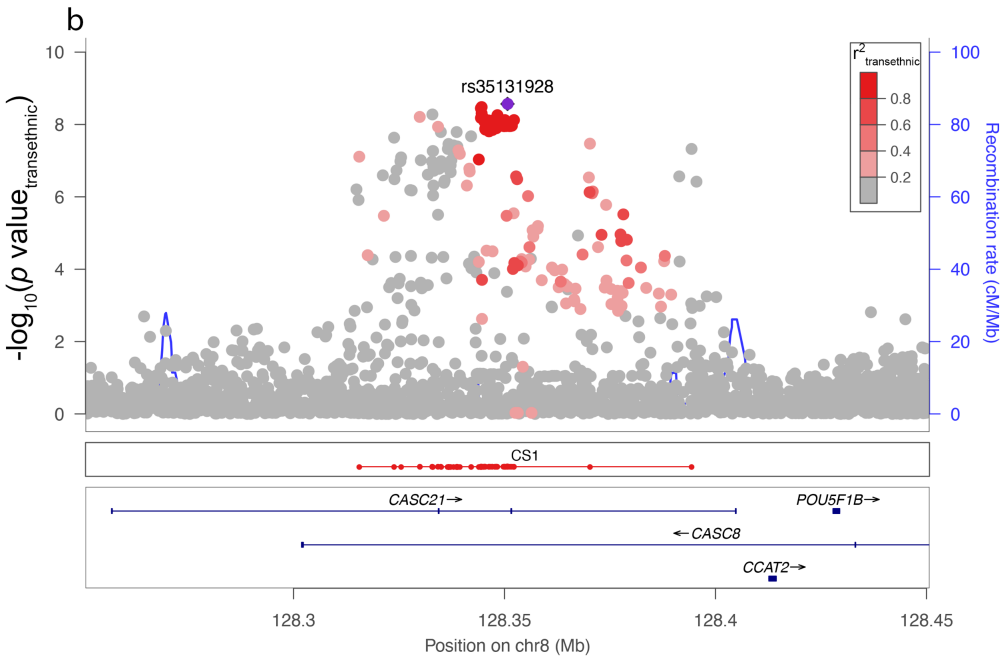
See Excel file for ESM Tables 1-11.

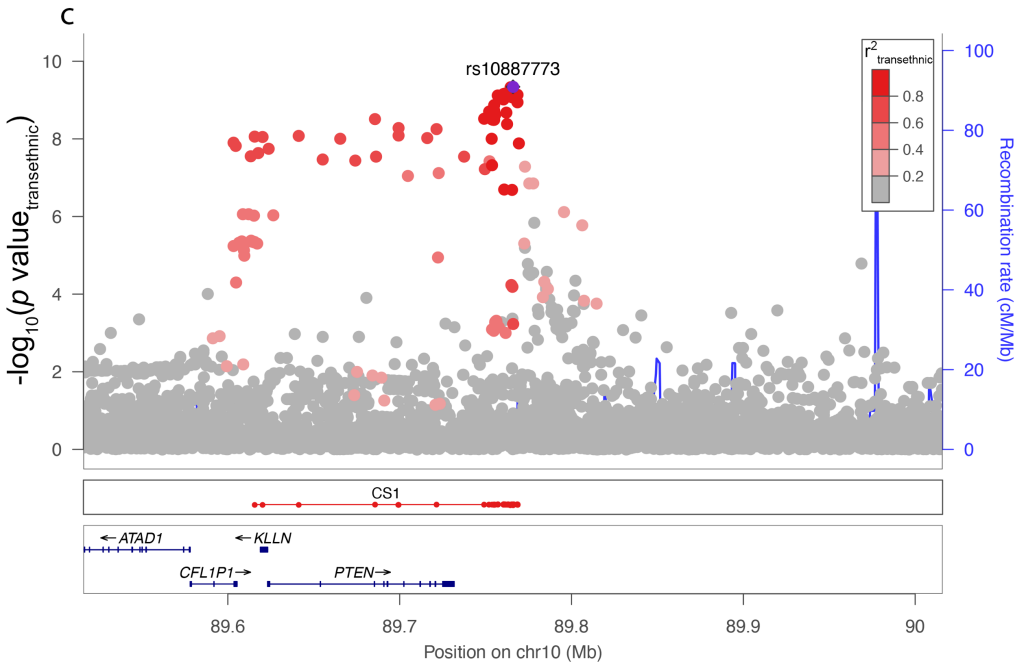
Electronic Supplementary Figures

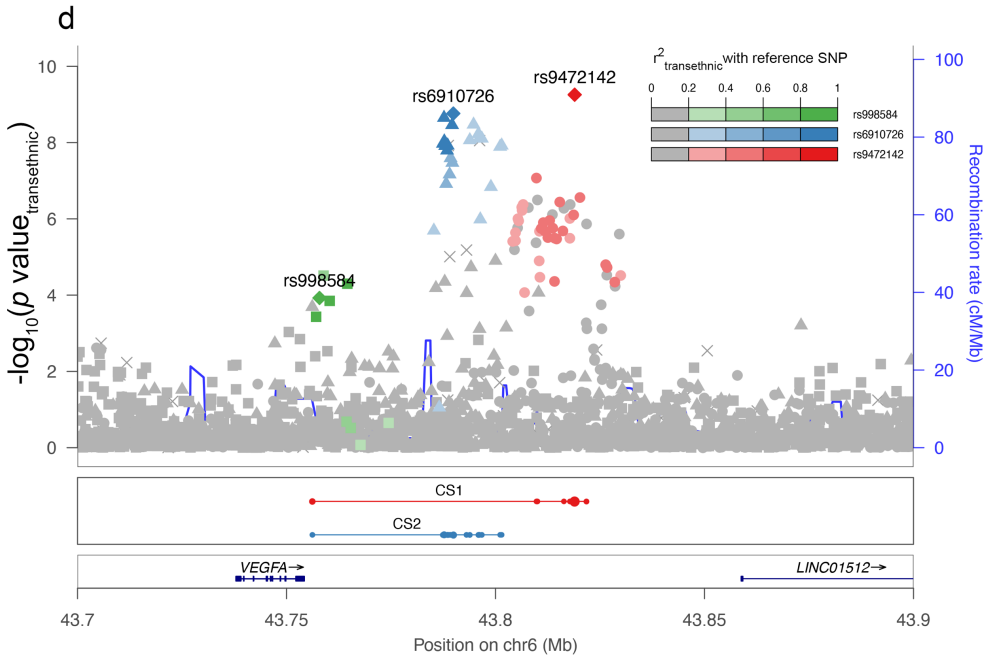

a
rs11887523
rs1879442
rs2032451
rs2157691
rs429216
rs4737010
rs560887
rs6975024
rs7208565
rs855791
rs4021964
rs72805692
rs334
b
rs10830963
rs111675407
rs11708067
rs11856307
rs1260326
rs13266634
rs174556
rs1879442
rs2908286
rs3833331
rs4719433
rs560887
rs7903146
c
rs1128249
rs1260326
rs17036160
rs1913657
rs2785990
rs465983
rs4841132
rs6876198
NAm
HI
ASN
HA
EA
AA
Transethnic
Population
Beta

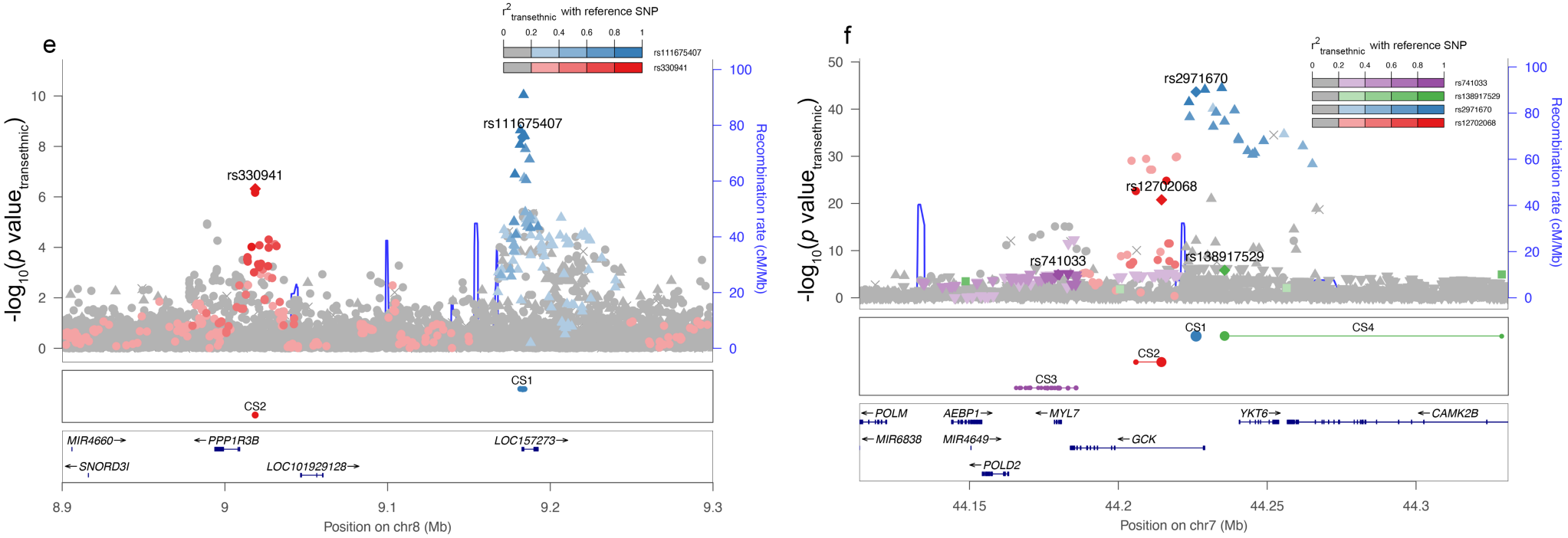
ESM Figure 1. Forest plot of all known significant ($p < 5.0 \times 10^{-9}$) GWAS results in the PAGE Study. If effective N < 30, population-specific meta-analyses were not computed. **(a)** HbA_{1c} known significant loci. HbA_{1c} data was unavailable for HI and NAm participants. **(b)** Fasting glucose known significant loci. **(c)** Fasting insulin known significant loci. Abbreviations: AA: African-American, EA: European American, HA: Hispanic/Latino, ASN: Asian, HI: Hawaiian, NAm: Native American.


a
rs571025325
r2 AA
0.8
0.6
0.4
0.2
-log10(p value_AA)
Recombination rate (cM/Mb)
CS1
MIR7702
KIAA0368
ZNF483
GNG10
UGCG
SUSD1
OR2K2
PTGR1
C9orf84
MIR4668
LRRC37A5P
DNAJC25
DNAJC25–GNG10
114
114.2
114.4
114.6
114.8
Position on chr9 (Mb)


b
rs35131928
r2 transethnic
0.8
0.6
0.4
0.2
-log10(p value_transethnic)
Recombination rate (cM/Mb)
CS1
CASC21
POU5F1B
CASC8
CCAT2
128.3
128.35
128.4
128.45
Position on chr8 (Mb)


c
rs10887773
r2 transethnic
0.8
0.6
0.4
0.2
-log10(p value_transethnic)
Recombination rate (cM/Mb)
CS1
ATAD1
KLLN
CFL1P1
PTEN
89.6
89.7
89.8
89.9
90
Position on chr10 (Mb)


d
rs9472142
rs6910726
rs998584
r2 transethnic with reference SNP
0
0.2
0.4
0.6
0.8
1
rs998584
rs6910726
rs9472142
-log10(p value_transethnic)
Recombination rate (cM/Mb)
CS1
CS2
VEGFA
LINC01512
43.7
43.75
43.8
43.85
43.9
Position on chr6 (Mb)



ESM Figure 2. LocusZoom plots of novel primary and independent secondary loci in the PAGE Study, and fine-mapping credible set top variants. For loci containing multiple credible sets (CS), the top variant from each credible set is displayed as a reference variant; the size of the dot indicating variants in each CS corresponds to the variants' posterior probabilities of being the causal variant. **(a)** Novel fasting glucose *LRRC37A5P* locus; all CS1 variants are shown. **(b)** Novel fasting insulin *CASC8|CASC21* locus; all CS1 variants are shown. **(c)** Novel fasting insulin *PTEN* locus; all CS1 variants are shown. **(d)** Novel fasting insulin *VEGFA* locus; top CS1 and CS2 variants are shown (chromosome position region is limited to improve visibility, so not all top CS variants are displayed), as well as Chen et al. 2021's identified *VEGFA* top variant rs998584. **(e)** Fasting insulin *PPP1R3B* locus containing novel independent secondary signal; top CS1 and CS2 variants are shown. **(f)** Fasting glucose *GCK* locus containing novel independent secondary signal; top CS1, CS2, CS3, and CS4 variants are shown. Transethnic LD was generated from unrelated subset of AA, HA, ASN, HI, NAm participants in the PAGE Study.