TITLE

Whole Genome Sequence Association Analysis of Fasting Glucose and Fasting Insulin Levels in Diverse Cohorts from the NHLBI TOPMed Program

AUTHORS & AFFILIATIONS

Daniel DiCorpo,^{1,2} Sheila M Gaynor,^{1,3} Emily M Russell,⁴ Kenneth E Westerman,^{5,6,7} Laura M Raffield,⁸ Timothy D Majarian,⁶ Peitao Wu,² Chloé Sarnowski,² Heather M Highland,⁹ Anne Jackson,¹⁰ NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, Natalie R Hasbani,¹¹ Paul S de Vries,¹² Jennifer A Brody,^{13,14} Bertha Hidalgo,¹⁵ Xiuqing Guo,¹⁶ James A Perry,^{17,18} Jeffrey R O'Connell,^{17,18} Samantha Lent,² May E Montasser,¹⁷ Brian E Cade,^{19,20,21} Deepti Jain,²² Heming Wang,^{19,20,21} Ricardo D'Oliveira Albanus,²³ Arushi Varshney,²³ Lisa R Yanek,²⁴ Leslie Lange,²⁵ Nicholette D Palmer,²⁶ Marcio Almeida,²⁷ Juan M Peralta,²⁷ Stella Aslibekyan,²⁸ Abigail S Baldrige,²⁹ Alain G Bertoni,³⁰ Lawrence F Bielak,³¹ Chung-Shiuan Chen,³² Yii-Der Ida Chen,¹⁶ Won Jung Choi,³³ Mark O Goodarzi,³⁴ James S Floyd,^{35,36} Marguerite R Irvin,¹⁵ Rita R Kalyani,²⁴ Tanika N Kelly,³² Seonwook Lee,³³ Ching-Ti Liu,² Douglas Loesch,^{37,38,18} JoAnn E Manson,^{7,39} James S Pankow,⁴⁰ Laura J Rasmussen-Torvik,²⁹ Alexander P Reiner,⁴¹ Elizabeth Selvin,⁴² Jennifer A Smith,^{31,43} Daniel E Weeks,^{44,4} Huichun Xu,^{17,18} Jie Yao,¹⁶ Wei Zhao,³¹ Stephen Parker,^{23,45} Alvaro Alonso,⁴⁶ Donna K Arnett,⁴⁷ John Blangero,²⁷ Eric Boerwinkle,¹¹ Adolfo Correa,⁴⁸ L. Adrienne Cupples,^{2,49} Joanne E Curran,²⁷ Ravindranath Duggirala,²⁷ Jiang He,³² Susan R Heckbert,^{35,41} Sharon LR Kardia,³¹ Ryan W Kim,³³ Charles Kooperberg,⁵⁰ Simin Liu,⁵¹ Rasika A Mathias,²⁴ Stephen T McGarvey,⁵² Braxton D Mitchell,^{17,18,53} Alanna C Morrison,¹¹ Patricia A Peyser,³¹ Bruce M Psaty,^{13,41,14,54} Susan Redline,^{19,20,55} Alan R Shuldiner,⁵⁶ Kent D Taylor,¹⁶ Ramachandran S Vasan,^{49,57,58} Karine A Viaud-Martinez,⁵⁹ Jose C Florez,^{6,21,60,7} James G Wilson,⁶¹ Robert Sladek,^{62,63} Stephen S Rich,⁶⁴ Jerome I Rotter,¹⁶ Xihong Lin,³ Josée Dupuis,² James B Meigs,^{21,7,65} Jennifer Wessel,^{66,67,68} Alisa K Manning^{6,7,5}

¹These authors contributed equally to this work, ²Department of Biostatistics, Boston University School of Public Health, Boston, MA, 02118, USA, ³Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, 02115, USA, ⁴Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, 15261, USA, ⁵Clinical and Translational Epidemiology Unit, Mongan Institute, Massachusetts General Hospital, Boston, MA, 02114, USA, ⁶Metabolism Program, The Broad Institute of MIT and Harvard, Cambridge, MA, 02124, USA, ⁷Department of Medicine, Harvard Medical School, Boston, MA, 02115, USA, ⁸Department of Genetics, University of North Carolina, Chapel Hill, NC, 27599, USA, ⁹Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27514, USA, ¹⁰Department of Biostatistics, University of Michigan, Ann Arbor, MI, 48109, USA, ¹¹Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, 77030, USA, ¹²Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, 77030, USA, ¹³Cardiovascular Health Research Unit, University of Washington, Seattle, WA, 98101, USA, ¹⁴Department of Medicine, University of Washington, Seattle, WA, 98101, USA, ¹⁵Ryals School of Public Health, University of Alabama at Birmingham, Birmingham, AL, 35294, USA, ¹⁶The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, 90502, USA, ¹⁷Division of Endocrinology, Diabetes, and Nutrition, University of Maryland School of Medicine, Baltimore, MD, 21201, USA, ¹⁸Program for Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, MD, 21201, USA, ¹⁹Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA, 02115, USA, ²⁰Division of Sleep Medicine, Harvard Medical School, Boston, MA, 02115, USA, ²¹Program in Medical and Population Genetics, The Broad Institute of MIT and Harvard, Cambridge, MA, 02124, USA, ²²Department of Biostatistics, University of Washington, Seattle, WA, 98195, USA, ²³Department of Computational Medicine & Bioinformatics, University of Michigan, Ann Arbor, MI, 48109, USA, ²⁴GeneSTAR Research Program, Johns Hopkins University School of Medicine, Baltimore, MD, 21287, USA, ²⁵Department of Medicine, Anschutz Medical Campus, University of Colorado Denver, Aurora, CO, 80045, USA, ²⁶Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, 27157, USA, ²⁷Department of Human Genetics and South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley School of Medicine, Brownsville and Edinburg, TX, 78539, USA, ²⁸23andMe, Sunnyvale,

CA, 94086, USA, ²⁹Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, 60611, USA, ³⁰Department of Epidemiology & Prevention, Wake Forest School of Medicine, Winston-Salem, NC, 27157, USA, ³¹Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, 48109, USA, ³²Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, 70112, USA, ³³Psomagen, Inc., Rockville, MD, 20850, USA, ³⁴Department of Medicine, Division of Endocrinology, Diabetes and Metabolism, Cedars-Sinai Medical Center, Los Angeles, CA, 90048, USA, ³⁵Cardiovascular Health Research Unit, University of Washington, Seattle, WA, 98195, USA, ³⁶Department of Medicine, University of Washington, Seattle, WA, 98195, USA, ³⁷Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, 21201, USA, ³⁸Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, 21201, USA, ³⁹Brigham and Women's Hospital, Boston, MA, 02115, USA, ⁴⁰Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN, 55454, USA, ⁴¹Department of Epidemiology, University of Washington, Seattle, WA, 98195, USA, ⁴²Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, 21287, USA, ⁴³Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, USA, ⁴⁴Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, 15261, USA, ⁴⁵Department of Human Genetics, University of Michigan, MI, 48109, USA, ⁴⁶Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, 30322, USA, ⁴⁷College of Public Health, University of Kentucky, Lexington, KY, 40506, USA, ⁴⁸Department of Medicine, University of Mississippi Medical Center, Jackson, MS, 39211, USA, ⁴⁹National Heart Lung and Blood Institute and Boston University's Framingham Heart Study, Framingham, MA, 01702, USA, ⁵⁰Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, 98109, USA, ⁵¹Center for Global Cardiometabolic Health (CGCH), Boston, MA, 02215, USA, ⁵²International Health Institute and Department of Epidemiology, Brown University School of Public Health, Providence, RI, 02912, USA, ⁵³Geriatrics Research and Education Clinical Center, Baltimore VA Medical Center, Baltimore, MD, 21201, USA, ⁵⁴Department of Health Services, University of Washington, Seattle, WA, 98101, USA, ⁵⁵Division of Pulmonary, Critical Care, and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, MA, 02115, USA, ⁵⁶Program for Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, MD, 21231, USA, ⁵⁷Evans Department of Medicine, Section of Preventive Medicine and Epidemiology, Boston University School of Medicine, Boston, MA, 02118, USA, ⁵⁸Evans Department of Medicine, Whitaker Cardiovascular Institute and Cardiology Section, Boston University School of Medicine, Boston, MA, 02118, USA, ⁵⁹Illumina Laboratory Services, Illumina, Inc., San Diego, CA, 92122, USA, ⁶⁰Center for Genomic Medicine and Diabetes Unit, Massachusetts General Hospital, Boston, MA, 02114, USA, ⁶¹Division of Cardiovascular Medicine, Beth Israel Deaconess Medical Center, Boston, MA, 02115, USA, ⁶²Department of Human Genetics, McGill University, Montreal, Quebec, H3A 0G1, CA, ⁶³Department of Medicine, McGill University, Montreal, Quebec, H3A 0G1, CA, ⁶⁴Center for Public Health Genomics, University of Virginia, Charlottesville, VA, 22908, USA, ⁶⁵Division of General Internal Medicine, Massachusetts General Hospital, Boston, MA, 02114, USA, ⁶⁶Department of Epidemiology, Fairbanks School of Public Health, Indiana University, IN, 46202, USA, ⁶⁷Department of Medicine, School of Medicine, Indiana University, IN, 46202, USA, ⁶⁸Diabetes Translational Research Center, Indiana University, IN, 46202, USA

* Corresponding Author:

Alisa K. Manning

amanning@broadinstitute.org

ABSTRACT

The genetic determinants of fasting glucose (FG) and fasting insulin (FI) have been studied mostly through genome and exome arrays, resulting in over 100 associated variants. We extended this work with a high-coverage whole genome sequencing (WGS) analysis from fifteen cohorts in the NHLBI Trans-Omics for Precision Medicine (TOPMed) program. More than 23,000 non-diabetic individuals from five self-reported race/ethnicities (African, Asian, European, Hispanic and Samoan) were included for each trait. We analyzed 60M variants in race/ethnicity-specific and pooled single variant and rare variant aggregate tests. Twenty-two variants across sixteen gene regions were found significantly associated with FG or FI, eight of which were rare (Minor Allele Frequency, MAF<0.05). Functional annotation from resources including the Diabetes Epigenome Atlas were compiled for each signal (chromatin states, annotation principal components, and others) to elucidate variant-to-function hypotheses. Near the *G6PC2* locus we identified a distinct FG signal at rare variant rs2232326 (MAF=0.01) after conditioning on known common variants. Functional annotations show rs2232326 to be disruptive and likely damaging while being weakly transcribed in islets. A pair of FG-associated variants were identified near the *SLC30A8* locus. These variants, one of which was rare (MAF=0.001) and Asian race/ethnicity-specific, were shown to be in islet-specific active enhancer regions. Other associated regions include rare variants near *ROBO1* and *PTPRT*, and common variants near *MTNR1B*, *GCK*, *GCKR*, *FOXA2*, *APOB*, *TCF7L2*, and *ADCY5*. We provide a catalog of nucleotide-resolution genomic variation spanning intergenic and intronic regions down to a minor allele count of 20, creating a foundation for future sequencing-based investigation of glycemic traits.

Type 2 diabetes (T2D [MIM: 125853]) and insulin resistance are complex genetic conditions resulting from dysregulation of fasting glucose (FG) and fasting insulin (FI)¹. These are traits influenced by a spectrum of common to rare genetic variation²⁻⁷ but most evidence has come from common variants^{8,9} and exome arrays^{2,3,6} as well as exome sequencing² and small samples of low-pass whole genome sequence^{4,10}. Genome-wide association studies (GWAS) of FG and FI have found over 100 associated common variants including those in non-coding and intergenic space^{2-4,6,8,9}. Comprehensive whole-genome sequencing (WGS) association analysis in population cohorts allows for a better understanding of likely causal variants at GWAS loci, discovery and fine-mapping in coding and noncoding regions, and rare variant testing to aggregate protein-coding variants or intergenic variants.

We identified and annotated with functional characterization common and rare variants associated with FG and (log-transformed) FI by performing variant association tests with WGS data from fifteen cohorts in the NHLBI Trans-Omics for Precision Medicine (TOPMed) program (**Table S1**). For all analyses in the present work, appropriate ethics approvals were received to use the data and data access protocols were followed for all data sets. As in prior quantitative traits analysis, we excluded individuals with diabetes (by glycemia or medication), giving N=26,807 (FG) and N=23,211 (FI) non-diabetic individuals from a diverse sample of five self-reported race/ethnicities (African, Asian, European, Hispanic, and Samoan). We used participant's self-reported race/ethnicity to assign individuals to one of five groups for stratified analyses or inclusion as a covariate. Individuals were given a single label to infer their particular ancestry, but each group represents a diverse cross-section of race, culture or admixture. Trait measures were harmonized across cohorts and assays and adjusted for self-reported race/ethnicity, study age, sex, and body mass index (BMI; **Table S2, Table S3, Supplemental Methods**). We assessed 60M variants from the TOPMed Freeze 5b WGS data freeze for each trait using single variant testing (minor allele count, MAC>20) in pooled and race/ethnicity-specific approaches. We further performed rare variant aggregate testing (minor allele frequency, MAF<0.01) using aggregate burden and SKAT tests in both gene centric and genetic region approaches (**Supplemental Methods**). We computed 95% credible sets for each distinct common variant signal conditioned on any other identified signal at the locus.

We identified twenty-two variants significantly associated with FG or FI across sixteen regions (**Table 1**). Five variants at the *MTNR1B*, *G6PC2*, *SLC30A8*, and *APOB* loci contained more than one distinct association signal (as determined with conditional analysis). Eighteen of the associations were identified among all race/ethnic groups while the remaining four were specific to a race/ethnicity group. Many of these identified associations were also rare, with eight having a MAF less than 0.05. Rare variant aggregate testing performed using both gene centric and genetic region approaches identified one significantly associated region with FG at the known *G6PC2* locus. No rare variant aggregate signals were found to be associated with FI (**Table S4-S9**). These results replicate and extend previous GWAS findings as summarized in **Figure 1, Table S10 and Table S11**. We used the Diabetes Epigenome Atlas (DGA) and TOPMed resources to provide functional annotations including chromatin states, annotation principal components (aPCs)¹¹, and expression quantitative trait loci (eQTL) from adipose, pancreas, liver and skeletal muscle (**Figure 1**). Manhattan and QQ plots for single variant analyses of FG and FI are shown in **Figure S1**, regional locus plots with tissue-specific annotations for significant loci in **Figure S2**, results for rare variant aggregate analysis, and Manhattan plots for region-based rare variant aggregate analysis in **Figure S3 and Tables S5-S9**, and associations of significant loci with related traits in **Figure S4 and Table S12**.

At the known FG- and T2D-associated *G6PC2* locus^{2,3}, we identified several previously identified variant associations with FG, including with rare variants (**Figure 2**). In single variant analyses, we identified a rare tertiary variant, rs2232326, after conditioning on the two common GWAS variants rs560887 (primary signal) and rs540524 (secondary signal). As one of the two exonic variants detected in the single variant analysis, rs2232326 is annotated¹¹ by the aPCs as disruptive and likely damaging (high aPC-Protein Function; 31.5, top 0.07% genomewide) and as highly conserved (high aPC-Conservation; 28.8, top 0.13% genomewide). It is weakly transcribed in islets and near the transcription end site (**Figure 2**). The allele frequency of the C allele at the missense variant rs2232326 was less than 0.01 in all race/ethnicity groups except for the Asian group where the frequency was 0.03 (gnomAD: East Asian AF=0.05, Overall AF=0.01). In aggregate gene centric testing of all 75 rare missense variants in *G6PC2*, this previously identified rare (MAF=0.01) variant rs2232326, along with variant rs2232323 (MAF=0.01), contributed the most to the significant association test statistic ($P_{\text{Burden},1,1}=1.4 \times 10^{-10}$, **Table S4**).

Given the multiple distinct signals at *G6PC2*, we performed a haplotype analysis to evaluate the contribution of rare variants and identify allele specific effects. We extended the haplotype analysis of Mahajan et al³ (rs560887, rs138726309, rs2232323, rs492594) to include our secondary (rs540524) and tertiary (rs2232326) signals. Our secondary signal is in moderate linkage ($r^2=0.58$) to the previously haplotyped rs492594 and the effect allele A has a glucose-raising effect in both marginal and conditional analyses (**Table S4, Table 1**). We observed consistent direction of effects as the previous haplotype analysis, demonstrating the reliability of associations identified in the present TOPMed sample. Both haplotypes including the C allele of the tertiary signal at rs2232326, the variant with the largest effect size included in the analysis, had glucose-lowering effects. The largest glucose-lowering effects at *G6PC2* were observed at the two haplotypes each carrying a rare allele: rs2232326 (rs560887-C, rs138726309-C, rs2232323-A, rs492594-C, rs540524-G, rs2232326-C, Beta=-0.15+/-0.00002) and rs2232323 (rs560887-T, rs138726309-C, rs2232323-C, rs492594-G, rs540524-A, rs2232326-T, Beta=-0.11+/-0.00008, Table S4).

We identified a pair of FG-associated signals in islet-specific active enhancer regions at the *SLC30A8* locus. The primary signal is at variant rs35859536, which is an intergenic variant 2.5KB downstream of *SLC30A8*. This variant is highly linked ($R^2>0.95$) to previously identified lead variants rs11558471 and rs3802177 at *SLC30A8*, both of which are in the 3' UTR. This is a known T2D susceptibility locus and has been identified as associated with triglyceride levels¹². Our lead variant is also significantly associated with T2D in TOPMed (**Table S12**)¹³. To evaluate potential causal variants (**Supplemental Methods**) we performed credible set analyses and found rs35859536 has a posterior probability (PP) of 0.48; other variants in the 95% credible set with PP of at least 0.05 were either missense or in the 3' UTR, are highly linked with this lead variant ($R^2>0.97$), and were significantly associated with FG in previous studies^{2, 8, 14}. Our lead variant, along with other previous lead variants, is in an active enhancer 2 region for islets; rs35859536 is also tagged for chromatin interaction for target genes in islet of Langerhans. The secondary FG-associated signal at the *SLC30A8* locus is at variant rs542965166. This intergenic variant is only observed in individuals in the Asian population (Asian EAF=0.007); this race/ethnicity-specificity is replicated in gnomAD¹⁵ where the allele is only observed in East Asians at a rare frequency. This secondary, race/ethnicity-specific signal is not highly linked to other variants in the region, which may indicate that this is a distinct, secondary signal and requires further follow-up in an Asian population.

We identified an FI-associated rare variant, rs539973028, upstream of the *ROBO1* locus. This locus has previously been studied for *SLIT-ROBO* signaling and expression in T2D complication diabetic retinopathy¹⁶. *ROBO1* has been associated with the glycemia-related traits of BMI and waist-to-hip ratio¹⁷⁻¹⁹ and is commonly expressed in muscle and skin²⁰. This variant is only observed in the African population of TOPMed and gnomAD¹⁵. It is intergenic and in a weakly transcribed region in islets.

We identified a pair of distinct rare variant signals associated with FI near the *PTPRT* gene (**Table 1**). The primary signal, rs185250851, is an intronic variant. It is rare in all tested population groups and not observed in Asian individuals, as validated in gnomAD¹⁵. The secondary signal at variant rs78618809 is in an intergenic region. This variant is within the top 5% of variants with respect to aPC-Distance to TSS/TSE, a composite measure of individual annotations indicating low variant distance to the endpoints of the intergenic region. It is rare overall, but observed with low frequency (EAF=0.07) in the African population. This gene has previously been associated with BMI, which has moderate genetic correlation (previously estimated as $\rho_g=0.48$) with FI²¹; ²². The expression of this gene is most commonly associated with variants as eQTLs in pancreas in GTEx²⁰. The multi-ethnic TOPMed sample permits the identification of this signal, which requires a sufficiently diverse sample.

In race/ethnicity-specific analyses, we observed two race/ethnicity-specific rare variant associations with FG in individuals of the Hispanic/Latinx population (**Table 1**). The first signal, rs1328056, is an intronic variant in the *HS6ST2* gene, which has been associated with obesity and impaired glucose metabolism in mice studies²³. The second signal is an intergenic variant near the *ATP5CKMT* gene, which is associated with eosinophil counts, a measure that has been negatively correlated with FG²⁴. We would require further data from individuals from the Hispanic/Latinx population in order to replicate these suggestive signals.

We identified two race/ethnicity-specific rare alleles associated with FI. In the European population, rs775018107 at the *LINC00704/LINC00705* locus was significantly associated with FI (**Table 1**); however, this

signal was not replicated in the UK Biobank (UKBB) (N=26,000, **Table S14**). We also identified a significant FI association in the Samoan population at rs117592405; this intronic variant was not replicated in an independent Samoan cohort (N=1,401, **Table S15**).

Our principal analytic focus was on rare variants, but at common variant loci we identified associations with FG ($P < 5 \times 10^{-8}$) in the pooled analysis including: *MTNR1B*, *GCK*, *GCKR*, *FOXA2*, *APOB*, *TCF7L2*, and *ADCY5* (**Table 1**). *APOB* has robust associations with lipids traits²⁵ and has significant parent-of-origin effects on metabolic traits²⁶. We further identified an additional locus with variants significantly associated with FI (*GCKR*) in the pooled analysis; *GCKR* has previously been reported to be associated with glycemic-related phenotypes.

In this paper, we leveraged high-coverage WGS data in large multi-ethnic population-based cohorts to assemble a comprehensive catalog of nucleotide-resolution genomic variation associated with the key diabetes-related quantitative traits FG and FI. Our analysis covered intergenic and intronic regions to a MAC of 20 in single variant analysis and combines base pair variation with tissue-specific epigenomic annotation to illuminate variant-to-function hypotheses in diabetes pathobiology. We have characterized *G6PC2* in terms of allelic effects and provided functional characterization of rare signal, demonstrating the glucose-lowering effects of rare alleles and islet-specificity of this locus's associations.

A strength of the present analysis is the inclusion of individuals from 15 cohorts, comprised of five major race/ethnicity groups (African, Asian, European, Hispanic/Latinx, and Samoan). Previous genetic studies of glycemic traits have included samples primarily from individuals of European ancestry, but increasingly a larger degree of African ancestry. The most recent meta-analysis by the MAGIC consortium included approximately 30% non-European ancestry individuals, demonstrating that a number of trait-associated loci that would have been undetected in samples exclusively of European ancestry²⁷. While extending the genetic ancestries studied beyond European populations, the MAGIC results were subject to the limitation of imputation by the 1000 Genomes Project reference panel, so most rare and ancestry-specific variation was still not assessed.

This analysis benefits from the availability of whole genome sequencing data provided by the TOPMed Program of the NHLBI's Precision Medicine Initiative^{10; 28}. Previous studies have been limited by reliance on imputation or minimal sample sizes for data with sequencing paired with glycemic phenotypes. The GoT2D study has performed WGS in a limited sample, contributing to the larger DIAMANTE meta-analysis of summary statistics but relying on imputation for complete genotyping of most samples²⁹. The UKBB study includes a large set of primarily European individuals with whole exome sequencing; however, the sample size with measured fasting glycemic traits is limited as described in the validation study (see **Supplemental Note**). One of the most expansive efforts, a MAGIC collaboration^{8; 30; 31}, has performed extensive analyses for glycemic traits, but results rely primary on Exome Chip data and thus have limited coverage of intergenic and intronic regions⁶.

A limitation of this study is the lack of replication of identified signals, particularly those which are race/ethnicity specific. We analyzed independent studies with genetic data to investigate associations significant in TOPMed; we were unable to replicate potentially novel signals in these external cohorts. This may be attributable to limitations in the available replication studies' samples with respect to size and race/ethnicity. To support the understanding of these signals, we consider a set of tissue-specific chromatin states, an effort that would benefit from further tissue-specific characterization across functional measures. This could also help inform the underlying biological mechanisms, towards improving the understanding of glycemic regulation and its role in diabetes.

This multi-ethnic WGS study provides the foundation for future sequencing-based investigation of glycemic traits. Our results from common and rare variant analysis included multiple suggestive hits, indicating the potential for the identification of novel signals given larger sequencing studies and external validation studies. The value of diverse studies like TOPMed is further evidenced by the specificity of such signals to certain populations and cohorts. This value is also demonstrated by the intronic and intergenic location of many such suggested signals. Future TOPMed study phases will permit the continued investigation of these signals. To support future research, all results from this analysis have been made available to the research community through the Type 2 Diabetes Knowledge Portal (Genetic Association Data will be released in January 2021).

SUPPLEMENTAL INFORMATION

Supplemental information includes description of Subject and Methods, four figures and fifteen tables.

ACKNOWLEDGMENTS

TOPMed Acknowledgments.

Molecular data for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the National Heart, Lung and Blood Institute (NHLBI). See the TOPMed Omics Support Table (Table S16a) for study specific omics support information. Core support including centralized genomic read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1; contract HHSN268201800002I). Core support including phenotype harmonization, data management, sample-identity QC, and general program coordination were provided by the TOPMed Data Coordinating Center (R01HL-120393; U01HL-120393; contract HHSN268201800001I). We gratefully acknowledge the studies and participants who provided biological samples and data for TOPMed.

Study-Specific, Resource and Individual Acknowledgements.

See Table S16b for study's, resources and individuals' acknowledgments.

DECLARATION OF INTERESTS

Jose Florez	Dr. Florez has received a consulting honorarium from Goldfinch Bio. and speaker fees from Novo Nordisk
James S. Floyd	Dr. Floyd has consulted for Shionogi Inc
Bruce M. Psaty	Psaty serves on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson.
Stella Aslibekyan	SA holds equity in 23andMe, Inc.

WEB RESOURCES

AMP-T2D Portal, <http://t2d.hugeamp.org/>

GTEEx Portal, <https://gtexportal.org/home/>

TOPMed, <https://www.nhlbiwgs.org/>

OMIM, <http://www.omim.org/>

LocusZoom, <https://statgen.github.io/localzoom/>

Diabetes Genome Atlas, <https://www.t2depigenome.org/>

DATA AND CODE AVAILABILITY

The summary results generated during this study are available at the AMP-T2D Portal, <http://t2d.hugeamp.org/>.

FIGURES

Figure 1. Characterization of significant single-variant signals associated with fasting glucose and fasting insulin in TOPMed. TOPMed features of distinct, significant signal associated with fasting glucose and fasting insulin (log-transformed). P-values (unconditional -log₁₀-transformed) for glycemic and related traits (HbA_{1c} and type 2 diabetes) and effect allele frequency (with respect to the pooled analysis effect allele) across race/ethnicities in TOPMed are reported. Chromatin states at relevant tissues were drawn from two sets of experiments from DGA³²; annotation PCs provide summaries of multi-faceted variant function; variants that are significant eQTLs in relevant tissues are denoted.

Abbreviations: EAF, effect allele frequency for TOPMed sample; EnhA1, Active Enhancer 1; EnhA2, Active Enhancer 2; Het, Heterochromatin; Quies, Quiescent/Low; ReprPC, Repressed PolyComb; ReprPCWk, Weak Repressed PolyComb; TssA, Active TSS; TssFlnk, Flanking TSS; TxWk, Weak Transcription; ZNF/Rpts, ZNF genes & repeats

Figure 2. Regional investigation of three conditionally significant signals associated with fasting glucose in the *G6PC2* locus in TOPMed

Regional association plot of $-\log_{10}$ P values by genomic position for sequential conditional single-variant analyses. The linkage disequilibrium (r^2) between the primary signal (rs560887, 2:168906638:T:C), as defined by the highest posterior probability, and variants in the region for each panel as calculated in TOPMed is indicated in the colors of the points. The chromatin states at four relevant tissues³³ and annotation PCs are provided across the region.

Abbreviations: APC1, APC Epigenetics, APC2, APC Conservation, APC3, APC Protein, APC9, APC Distance to TSS/TSE; EnhA1, Active Enhancer 1; EnhA2, Active Enhancer 2; Het, Heterochromatin; Quies, Quiescent/Low; ReprPC, Repressed PolyComb; ReprPCWk, Weak Repressed PolyComb; TssA, Active TSS; TssFlnk, Flanking TSS; TxWk, Weak Transcription; ZNF/Rpts, ZNF genes & repeats

TABLES

Table 1. Distinct signals at loci associated with glycemic traits FG and FI in TOPMed at genome-wide significance ($p < 5 \times 10^{-8}$).

Trait	Population	Nearest Gene	MarkerID ^a	EA	rsID	Annotation	EAF	P-Value	Beta	SE	Conditioned on
Fasting Glucose	Pooled	<i>MTNR1B</i>	11:92975544:C:G	G	rs10830963	Intronic	0.24	9.1×10^{-46}	0.07	0.01	
			11:92884161:G:A ^b	G	rs73560545	intronic	0.83	5.7×10^{-6}	0.03	0.01	rs10830963
		<i>G6PC2</i>	2:168906638:T:C	C	rs560887	Intronic	0.82	6.8×10^{-37}	0.07	0.01	
			2:168900420:A:G ^b	A	rs540524	2KB upstream	0.66	9.9×10^{-14}	0.04	0.01	rs560887
			2:168907981:T:C ^c	T	rs2232326	Missense	0.99	5.0×10^{-6}	0.15	0.03	rs560887, rs540524
		<i>GCK</i>	7:44189469:C:T	T	rs1799884	2KB upstream	0.18	3.9×10^{-28}	0.06	0.01	
		<i>GCKR</i>	2:27508073:T:C	C	rs1260326	Missense	0.65	6.1×10^{-21}	0.04	0.01	
		<i>FOXA2</i>	20:22581688:A:AG	A	rs3833331	3' UTR	0.85	5.4×10^{-10}	0.04	0.01	
		<i>SLC30A8</i>	8:117179236:C:T	C	rs35859536	2.5KB downstream	0.74	1.0×10^{-9}	0.03	0.01	
			8:117258547:C:T ^b	T	rs542965166	Intronic	0.001	1.9×10^{-6}	0.50	0.11	rs35859536
		<i>APOB</i>	2:21074277:A:G	A	rs478588	30KB upstream	0.29	2.9×10^{-9}	0.03	0.01	
		<i>TCF7L2</i>	10:112998590:C:T	T	rs7903146	Intronic	0.25	2.0×10^{-8}	0.03	0.01	
		<i>ADCY5</i>	3:123335923:A:C	A	rs72964564	Intronic	0.80	2.8×10^{-8}	0.03	0.01	
Fasting Insulin	HS/L	<i>HS6ST3</i>	13:96407609:A:G	G	rs1328056	Intronic	0.02	3.6×10^{-8}	0.33	0.06	
		<i>CTD-2199C04.4</i>	5:10169711:T:C	C	rs13361160	Intergenic	0.45	3.1×10^{-8}	0.10	0.02	
	Pooled	<i>GCKR</i>	2:27508073:T:C	C	rs1260326	Missense	0.67	7.2×10^{-13}	0.03	0.01	
		<i>PTPRT</i>	20:42752773:G:A	A	rs185250851	Intronic	0.002	2.1×10^{-8}	0.30	0.05	
			20:43230137:C:T ^b	T	rs78618809	Intergenic	0.02	5.9×10^{-6}	0.08	0.02	rs185250851
		<i>ROBO1</i>	3:79812347:C:A	C	rs539973028	44KB upstream	0.99	4.7×10^{-8}	0.51	0.09	
	EU	<i>LINC00704, LINC00705</i>	10:4656482:GAAAAT:G	G	rs775018107	ncRNA intronic	0.002	4.5×10^{-8}	0.33	0.06	
	Samoan	<i>RP11/IGSF11</i>	3:118656074:T:G	G	rs117592405	ncRNA intronic	0.01	3.3×10^{-8}	0.80	0.14	

^a chromosome, position(Hg38), reference allele, alternative allele^b indicates secondary signal; ^c indicates tertiary signal for association at significance level $P < 1 \times 10^{-5}$ and MAC > 20 after conditional analysis. Abbreviations: EA=Effect Allele, EAF=Effect Allele Frequency, EU=European, HS/L=Hispanic/Latinx

REFERENCES

1. Saeedi, P., Petersohn, I., Salpea, P., Malanda, B., Karuranga, S., Unwin, N., Colagiuri, S., Guariguata, L., Motala, A.A., Ogurtsova, K., et al. (2019). Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9(th) edition. *Diabetes Res Clin Pract* 157, 107843.
2. Wessel, J., Chu, A.Y., Willems, S.M., Wang, S., Yaghootkar, H., Brody, J.A., Dauriz, M., Hivert, M.F., Raghavan, S., Lipovich, L., et al. (2015). Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. *Nat Commun* 6, 5897.
3. Mahajan, A., Sim, X., Ng, H.J., Manning, A., Rivas, M.A., Highland, H.M., Locke, A.E., Grarup, N., Im, H.K., Cingolani, P., et al. (2015). Identification and functional characterization of G6PC2 coding variants influencing glycemic traits define an effector transcript at the G6PC2-ABCB11 locus. *PLoS Genet* 11, e1004876.
4. Jun, G., Manning, A., Almeida, M., Zawistowski, M., Wood, A.R., Teslovich, T.M., Fuchsberger, C., Feng, S., Cingolani, P., Gaulton, K.J., et al. (2018). Evaluating the contribution of rare variants to type 2 diabetes and related traits using pedigrees. *Proc Natl Acad Sci U S A* 115, 379-384.
5. Manning, A., Highland, H.M., Gasser, J., Sim, X., Tukiainen, T., Fontanillas, P., Grarup, N., Rivas, M.A., Mahajan, A., Locke, A.E., et al. (2017). A Low-Frequency Inactivating AKT2 Variant Enriched in the Finnish Population Is Associated With Fasting Insulin Levels and Type 2 Diabetes Risk. *Diabetes* 66, 2019-2032.
6. Ng, N.H.J., Willems, S.M., Fernandez, J., Fine, R.S., Wheeler, E., and Wessel, J. (2020). Tissue-Specific Alteration of Metabolic Pathways Influences Glycemic Regulation. <https://www.biorxiv.org/content/101101/790618v1>.
7. Sarnowski, C., Leong, A., Raffield, L.M., Wu, P., de Vries, P.S., DiCorpo, D., Guo, X., Xu, H., Liu, Y., Zheng, X., et al. (2019). Impact of Rare and Common Genetic Variants on Diabetes Diagnosis by Hemoglobin A1c in Multi-Ancestry Cohorts: The Trans-Omics for Precision Medicine Program. *Am J Hum Genet* 105, 706-718.
8. Manning, A.K., Hivert, M.F., Scott, R.A., Grimsby, J.L., Bouatia-Naji, N., Chen, H., Rybin, D., Liu, C.T., Bielak, L.F., Prokopenko, I., et al. (2012). A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet* 44, 659-669.
9. Dupuis, J., Langenberg, C., Prokopenko, I., Saxena, R., Soranzo, N., Jackson, A.U., Wheeler, E., Glazer, N.L., Bouatia-Naji, N., Gloyn, A.L., et al. (2010). New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* 42, 105-116.
10. Fuchsberger, C., Flannick, J., Teslovich, T.M., Mahajan, A., Agarwala, V., Gaulton, K.J., Ma, C., Fontanillas, P., Moutsianas, L., McCarthy, D.J., et al. (2016). The genetic architecture of type 2 diabetes. *Nature* 536, 41-47.
11. Li, X., Li, Z., Zhou, H., Gaynor, S.M., Liu, Y., Chen, H., Sun, R., Dey, R., Arnett, D.K., Aslibekyan, S., et al. (2020). Dynamic incorporation of multiple in silico functional annotations empowers rare variant association analysis of large whole-genome sequencing studies at scale. *Nat Genet* 52, 969-983.
12. Richardson, T.G., Sanderson, E., Palmer, T.M., Ala-Korpela, M., Ference, B.A., Davey Smith, G., and Holmes, M.V. (2020). Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: A multivariable Mendelian randomisation analysis. *PLoS Med* 17, e1003062.
13. Wessel, J., Majarian, T.D., Highland, H.M., Raghavan, S., Szeto, M.D., Hasbani, N.R., de Vries, P.S., Brody, J.A., Sarnowski, C., DiCorpo, D., et al. (2020). Rare Non-coding Variation Identified by Large Scale Whole Genome Sequencing Reveals Unexplained Heritability of Type 2 Diabetes. *medRxiv*, 2020.2011.2013.20221812.

14. Nagy, R., Boutin, T.S., Marten, J., Huffman, J.E., Kerr, S.M., Campbell, A., Evenden, L., Gibson, J., Amador, C., Howard, D.M., et al. (2017). Exploration of haplotype research consortium imputation for genome-wide association studies in 20,032 Generation Scotland participants. *Genome Med* 9, 23.
15. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alfoldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581, 434-443.
16. Zhou, W., Wang, H., Yu, W., Xie, W., Zhao, M., Huang, L., and Li, X. (2017). The expression of the Slit-Robo signal in the retina of diabetic rats and the vitreous or fibrovascular retinal membranes of patients with proliferative diabetic retinopathy. *PLoS One* 12, e0185795.
17. Pulit, S.L., Stoneman, C., Morris, A.P., Wood, A.R., Glastonbury, C.A., Tyrrell, J., Yengo, L., Ferreira, T., Marouli, E., Ji, Y., et al. (2019). Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum Mol Genet* 28, 166-174.
18. Kichaev, G., Bhatia, G., Loh, P.R., Gazal, S., Burch, K., Freund, M.K., Schoech, A., Pasaniuc, B., and Price, A.L. (2019). Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *Am J Hum Genet* 104, 65-75.
19. Winkler, T.W., Justice, A.E., Graff, M., Barata, L., Feitosa, M.F., Chu, S., Czajkowski, J., Esko, T., Fall, T., Kilpelainen, T.O., et al. (2015). The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. *PLoS Genet* 11, e1005378.
20. GTEx Consortium. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 369, 1318-1330.
21. Pena, G.G., Dutra, M.S., Gazzinelli, A., Correa-Oliveira, R., and Velasquez-Melendez, G. (2014). Heritability of phenotypes associated with glucose homeostasis and adiposity in a rural area of Brazil. *Ann Hum Genet* 78, 40-49.
22. Gervais, O., Ueno, K., Kawai, Y., Hitomi, Y., Misawa, K., Teraguchi, S., Wang, Y.Y., Tokunaga, K., and Nagasaki, M. (2020). Genomic Heritabilities and Correlations of 17 Traits Related to Obesity and Associated Conditions in the Japanese Population. *G3 (Bethesda)* 10, 2221-2228.
23. Pessentheiner, A.R., Ducasa, G.M., and Gordts, P. (2020). Proteoglycans in Obesity-Associated Metabolic Dysfunction and Meta-Inflammation. *Front Immunol* 11, 769.
24. Zhu, L., Su, T., Xu, M., Xu, Y., Li, M., Wang, T., Sun, J., Zhang, J., Xu, B., Lu, J., et al. (2013). Eosinophil inversely associates with type 2 diabetes and insulin resistance in Chinese adults. *PLoS One* 8, e67613.
25. de Vries, P.S., Brown, M.R., Bentley, A.R., Sung, Y.J., Winkler, T.W., Ntalla, I., Schwander, K., Kraja, A.T., Guo, X., Franceschini, N., et al. (2019). Multiancestry Genome-Wide Association Study of Lipid Levels Incorporating Gene-Alcohol Interactions. *Am J Epidemiol* 188, 1033-1054.
26. Hochner, H., Allard, C., Granot-HersHKovitz, E., Chen, J., Sitlani, C.M., Szadovska, S., Lumley, T., McKnight, B., Rice, K., Enquobahrie, D.A., et al. (2015). Parent-of-Origin Effects of the APOB Gene on Adiposity in Young Adults. *PLoS Genet* 11, e1005573.
27. Chen, J., Spracklen, C.N., Marenne, G., and Varshney, A. (2020). The Trans-Ancestral Genomic Architecture of Glycaemic Traits. <https://www.biorxiv.org/content/101101/20200723217646v1>.
28. Flannick, J., Fuchsberger, C., Mahajan, A., Teslovich, T.M., Agarwala, V., Gaulton, K.J., Caulkins, L., Koesterer, R., Ma, C., Moutsianas, L., et al. (2017). Sequence data and association statistics from 12,940 type 2 diabetes cases and controls. *Sci Data* 4, 170179.
29. Mahajan, A., Taliun, D., Thurner, M., Robertson, N.R., Torres, J.M., Rayner, N.W., Payne, A.J., Steinthorsdottir, V., Scott, R.A., Grarup, N., et al. (2018). Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet* 50, 1505-1513.

30. Scott, R.A., Scott, L.J., Magi, R., Marullo, L., Gaulton, K.J., Kaakinen, M., Pervjakova, N., Pers, T.H., Johnson, A.D., Eicher, J.D., et al. (2017). An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. *Diabetes* 66, 2888-2902.
31. Scott, R.A., Lagou, V., Welch, R.P., Wheeler, E., Montasser, M.E., Luan, J., Magi, R., Strawbridge, R.J., Rehnberg, E., Gustafsson, S., et al. (2012). Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat Genet* 44, 991-1005.
32. Gaulton, K.J., Nammo, T., Pasquali, L., Simon, J.M., Giresi, P.G., Fogarty, M.P., Panhuis, T.M., Mieczkowski, P., Secchi, A., Bosco, D., et al. (2010). A map of open chromatin in human pancreatic islets. *Nat Genet* 42, 255-259.
33. Varshney, A., Scott, L.J., Welch, R.P., Erdos, M.R., Chines, P.S., Narisu, N., Albanus, R.D., Orchard, P., Wolford, B.N., Kursawe, R., et al. (2017). Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. *Proc Natl Acad Sci U S A* 114, 2301-2306.