

Large-scale multi-omics in Hispanic/Latino populations functionally implicates genes for cardiometabolic traits

Corresponding Author: Dr Jennifer Below

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

Petty et al. present a large GWAS for diabetes in Hispanic/ Latino populations, which are underrepresented in current studies. In this novel work they identify novel genes and integrate other omics to their genetic findings. However, some key aspects need to be addressed to enhance this paper.

Major

1. a) The authors indicate that the participants "clustered with the 1000 Genomes Admixed American super population (AMR) in genetic principal component analysis," but they do not show the PCA plot. The authors need to show the plot of the clustering of the participants in this GWAS with the AMR.

b) It is anticipated that there will also be people who self-report as Hispanic/Latino but do not cluster well with AMR. These will need to be mentioned

2. a) The authors state that "Association testing for the four lipid phenotypes (HDL-C, LDL-C, TC, and TG) were conducted via linear regression SNPTTEST, EFACTS/EMMAX, SOLAR, GENESIS, PLINK, or R". It's not clear what type of transformation was applied to the quantitative phenotypes in view of the varied software used in the study level analysis. This information is not clarified even in the supplementary Table 2. The authors need to clarify this?

b) In the event that different studies employ different transformations, as is suspected, Please explain why MR MEGA was used to aggregate the GWAS study level findings?

3. The proteomic analysis needs to be better described. Its a bit confusing as it is. For instance the authors indicate that "we measured the Olink Explore 3072 panel on 528 stored plasma samples from 271 individuals in CCHC." Does this mean each individual had more that one sample or when these for multiple time points

4. The authors show results for Mendelian Randomisation but the precise methods used for this are missing in the methods section

Reviewer #2

(Remarks to the Author)

The research design yields valuable insights by conducting a large-scale multi-omics investigation within a non-white demographic, with a particular focus on Hispanic/Latino individuals. This study stands out as one of the few GWAS endeavors targeting diverse populations, involving the recruitment of 63,185 samples from 19 distinct studies. The findings unveil novel genes while also pinpointing variants unique to Hispanic/Latino and African-ancestry populations. Subsequent to the GWAS, the researchers conducted transcriptome-wide association analysis based on summary statistics. They utilized newly devised islet expression prediction models and meticulously pursued identified genes through differential abundance analyses, utilizing directly measured blood transcriptomic and proteomic data from an independent Hispanic cohort. The methodology appears robust and is elucidated thoroughly, encompassing both GWAS and post-GWAS procedures. The major limitation of this endeavour is to follow it up with Euro-centric GTEx based prediction models and the need for ethnically diverse new datasets is clearly discussed in the manuscript.

Minor comments:

1. Could the authors please specify the dataset utilized for measuring "gene expression in pancreatic islet cells" as the

InsPIRE consortium, which is first mentioned in the Discussion section. Could the authors please comment on the percentages of Hispanic/Latinos in the InsPIRE consortium?

2.

3. It remains unclear from the abstract whether the research incorporates new measurement endeavors such as Olink in CCHC, in addition to in-silico inquiries. The authors may wish to rebalance the abstract and underscore this point.

4. The meaning of "MVP" is not provided in the text. Please clarify its definition.

Reviewer #3

(Remarks to the Author)

Petty et al reported GWAS meta-analyses and "functionally-oriented" gene analysis of T2D and 4 lipid traits in Hispanic/Latino populations, identifying novel loci as well as replicating previously known signals. The discovery of novel loci near and away from previously discovered genetic associations based on studies that included only European or predominantly-European multi-ancestral studies underscores the value of recently admixed populations in discovery and refining of genetic risk for T2D and dyslipidemia. The newly developed islet model and RNAseq and proteomic resources in Hispanics are important contributions to the field. However, to what extent the RNAseq and proteomics-based analyses validated the present study's main findings after appropriate correction for multiple testing is unclear. Moreover, the added value of the pancreatic islet resource on top of GTEx in the interpretation of the signals identified has not been clearly demonstrated. The paper requires improvement in presentation and methodological details. Main comments are below:

1) The presentation of the paper needs a lot of improvement. There are several instances where the Methods Section lacks details of specific approaches, the results and cited tables do not provide the stated information, and tables are unclear, incomplete, or inconsistent with the text in the Methods Section. I tried to list some of these below (see minor comments), but the authors need to thoroughly review the paper. Although I am not asking the authors to change this, it is obvious that by combining 5 traits in one paper, the Results Section looks overly repetitive and descriptive. Presentation of MR, DEG, and DEP analyses results of five traits in one paragraph was ineffective in communicating the key finding for any one of the five traits.

2) Why was MR-MEGA chosen for meta-analysis of a multi-way admixed Hispanic/Latino population across the 16 studies? MR-MEGA assumes ancestry-related allelic effect heterogeneity as captured by the axes of genetic variation included in the model, and has been effectively demonstrated when samples from different ancestry groups are meta-analyzed. Can the authors clarify the rationale for the choice of MR-MEGA as a tool for pooling GWAS results coming from the same population group? Was the authors' assumption that there may be "...study-level differences in ancestry (line 452)" based on assessment of the data? Can they demonstrate whether the first or second axis of genetic variation included in the meta-regression capture an ancestry cline in the admixed Hispanics/Latinos? Why were only 2 axes chosen?

3) The authors acknowledged that the novel MOB3B variant did not replicate, including in Suzuki et al.'s Hispanic analysis. Is it in a consistent effect direction? In Suppl Table 3, the effect direction for this novel association is equally split between the cohorts included in meta-analysis, with increased T2D risk in 7 cohorts and reduced risk in 7 cohorts. The heterogeneity correlated with ancestry is also significant, the p-value being the smallest of all the reported T2D-associated SNP associations. Given these clues, I would urge the authors to explore whether the Hispanic cohorts that were included in this study but not in Suzuki et al could explain this. Depicting where the different cohorts fall on the 2 axes of genetic variation may also help examine this in the context of both cohort differences and ancestry cline.

4) The Methods Section mentions that the functionally-oriented results that were novel would be tested for gene expression-phenotype (DEG) and protein-phenotype (DEP) associations. a) The results section does not mention any finding based on MR and the DEG and DEP findings are not clearly presented. b) Suppl Table 6 has the DEG and DEP association findings, but the table includes known as well as novel genes, which does not match what was mentioned in the methods section about specific follow up of novel findings. The Table does not have multiple testing-corrected p-values, making it difficult to assess whether any of those associations could be considered significant or not. These need to be addressed. c) Along this line, the Discussion Section states "multi-tiered robust evidence". However, the results do not indicate that the SNP analysis results and gene-oriented results yielded overlapping loci. The evidence from proteomics and RNAseq do not seem to strongly support the gene-oriented results if multiple-testing correction were applied. Thus, I would not call the evidence "multi-tiered" or "robust" unless more evidence is provided.

5) It was mentioned that the DEG and DEP analyses were performed to validate findings of the functionally-oriented analyses. How do the authors see this approach in the lens of previous work showing that differentially expressed genes are more likely disease-induced rather than disease-causing (Porcu et al Nat Commun. 2021 Sep 24;12(1):5647. PMID: 34561431)? Would it not be more reasonable to perform this follow up transcriptome and protein analyses on genes showing evidence of causality guided by MR? Could a more integrated analysis be done on potentially overlapping samples with both RNAseq and protein data?

6) The authors mention "ANXA4 for T2D" and others as example of genes offering multi-omics support for the gene-oriented results. The results in the supplementary table do not suggest that either the S-PrediXcan or proteomics findings are significant nominally or after multiple-testing correction. Can the authors provide evidence to support these?

7) The Discussion Section highlighted some novel associations, which is worthy. However, the S-PrediXcan analysis across the 5 traits yielded loci that are predominantly known in previous studies. As stated in the Introduction Section, multi-ancestral GWAS enables deeper interrogation of known loci. Could the authors comment on whether the functionally-oriented findings add to our understanding of the biology of known signals of T2D or lipids, e.g. by narrowing down the likely causal genes, transcript regulation, tissue-specificity etc.

8) What number of tissues, genes, and phenotypes were considered and how did the authors come up with an FDR of 0.01 for the tissue-agnostic GReX analysis?

9) Provide details about the specific MR test used, approach, and citation. How was reverse causality tested and validated?

10) Provide a specific and clear comparison between Hispanic/Latinos and other US-populations (either non-Hispanics or

another race/ethnic group in the US) to support the stated point that “Many cardiometabolic diseases are more prevalent in Hispanics/Latinos (lines 186-188)” and “...in a population with significant cardiometabolic health disparities”. Comparisons using prevalence figures based on all US populations is ambiguous.

Minor comments:

- Suppl Table 2 has four columns specific about male- and female-stratified analysis. The text does not talk about sex-stratified analysis. Please clarify. Also there are values "NA" and "none", which makes it confusing. Does "none" mean no covariate was adjusted for and "NA" means the whole cohort was not included?
- Suppl Tables 3-6 contain beta and SE. Clarify how these were obtained as MR-MEGA does not provide these statistics.
- Clarify whether the single variant-based analyses and gene-based analysis with S-PrediXcan identified overlapping signals or not.
- The discussion section has several redundant information, and should be significantly reduced.
- Suppl Table 3: The column "Neff" is empty for the "DIAMANTE South Asian results"
- Suppl Table 4: The title should include "BMI-adjusted". Otherwise, the titles of STable 3 and 4 are identical and confusing.
- Across Suppl Tables 4, 7, 8, 9, 10, the 'fine mapping' results are presented confusingly under the columns "known variants plus or minus 1 Mb" and "known variants in credible set(s)". "yes" and "known" look the same. Do you mean to say "novel" instead of "known", and what is "no", "NA"? Clear description is needed.
- Headers need description for many Supplementary Tables, in particular suppl Table 6.
- Suppl Table 6: Where the gene-based results are "known", provide information about the known locus and its associated PMID.
- Line 191: Edit "...which is has a prevalence of..."
- Line 222: Provide citation to the statement "while much of the genetic architecture of cardiometabolic disease risk is shared across populations and environmental contexts,"
- Line 276: The description needs correction because BMI-adjusted genetic effect on T2D does not mean that the effect is driven by mediation by BMI, rather suggests BMI-independence.
- Line 280-281: States “Sentinel variants, or the variant with the lowest p-value, were identified for each region with $p < 5 \times 10^{-8}$ ”. What is the operational definition used to assign signals into a “region”?
- Line 287-291: The flow can be improved by re-ordering sentences - define "known" and "novel" signal, followed by "novel signal in a known locus"
- Line 297-299: Why was the "African ancestry group (MEDIA)" not included in the replication effort?
- Line 297: The last part of the sentence "...top-level multi-ancestry meta-analysis" is unclear.
- Line 380: The first sentence is not clear because it lumps the T2D GWAS results from the model without BMI (which is the mean model according to the methods) and the BMI-adjusted model. Present the number of associations from the main model, then state that the number of loci that remained and the number of new loci that emerged after BMI-adjustment. This aligns with the goals of the BMI-adjustment as stated in the methods (with edits – see my other comment).
- Line 386-388: It says 2 T2D SNPs are "novel signals in known loci", and references suppl Table 5. However, there is no information in the referenced table to know which SNPs are novel. The text is also unclear which SNPs these are, why there are distinct from the previous loci, and does not provide appropriate citation.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

Petty et al. GWAS for diabetes in Hispanic/ Latino populations, is novel and a critical step to increase the representation of this population group in GWAS. The authors have addressed well the concerns and critiques I raised earlier.

Reviewer #2

(Remarks to the Author)

Thank you for answering the comments and suggestions.

Reviewer #3

(Remarks to the Author)

We appreciate the authors for addressing my comments and for improving the content and flow of the manuscript considerably.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

RESPONSES TO REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

Petty et al. present a large GWAS for diabetes in Hispanic/ Latino populations, which are underrepresented in current studies. In this novel work they identify novel genes and integrate other omics to their genetic findings. However, some key aspects need to be addressed to enhance this paper.

Thank you for your time and constructive critique of our manuscript.

Major

1. a) The authors indicate that the participants "clustered with the 1000 Genomes Admixed American super population (AMR) in genetic principal component analysis," but they do not show the PCA plot. The authors need to show the plot of the clustering of the participants in this GWAS with the AMR.

b) It is anticipated that there will also be people who self-report as Hispanic/Latino but do not cluster well with AMR. These will need to be mentioned.

Thank you for this opportunity to better clarify our inclusion criteria. We have changed the wording of the quoted sentence to explain that we respected all individual's self-identification of Hispanic and/or Latino group. We had mistakenly thought MVP had used genetically informed population grouping, but upon rereading their papers they did indeed use self-reported ethnicity. To accurately describe the methods we have also revised the text:

"All participants in the contributing studies self-identify as Hispanic and/or Latino."

2. a) The authors state that "Association testing for the four lipid phenotypes (HDL-C, LDL-C, TC, and TG) were conducted via linear regression SNPTTEST, EPACTS/EMMAX, SOLAR, GENESIS, PLINK, or R". It's not clear what type of transformation was applied to the quantitative phenotypes in view of the varied software used in the study level analysis. This information is not clarified even in the supplementary Table 2. The authors need to clarify this?

Thank you for the opportunity to clarify, the residual values were inverse rank normal transformed for all lipids analyses. We have now revised the text to include this point:

"Association testing for the four lipid phenotypes (HDL-C, LDL-C, TC, and log-transformed TG) were conducted via linear regression SNPTTEST,⁴⁷ EPACTS/EMMAX,⁴⁸ SOLAR,⁴⁹ GENESIS,⁵⁰ PLINK,⁵¹ or R.⁵² Residual lipid values were calculated, adjusting for age, sex, and any necessary study-specific covariates (e.g., study location or batch). The residual values were inverse rank normal transformed."

b) In the event that different studies employ different transformations, as is suspected, Please explain why MR MEGA was used to aggregate the GWAS study level findings?

All studies used an inverse rank normal transformation, which is now clarified in the text. Often in GWAS of admixed populations, there is heterogeneity in allelic effects that is correlated with ancestry. MR-MEGA tests for SNP associations and quantifies the extent of heterogeneity in allelic effects that is correlated with ancestry using a matrix of mean pairwise allele frequency differences between each GWAS. Our meta-analysis of results from rank-normalized trait data follows established approaches, for example see GIANT,¹ GLGC.²

3. The proteomic analysis needs to be better described. Its a bit confusing as it is. For instance the authors indicate that "we measured the Olink Explore 3072 panel on 528 stored plasma samples from 271 individuals in CCHC." Does this mean each individual had more than one sample or when these for multiple time points

Thank you for highlighting these points. In the CCHC proteomics data, all lipid and diabetes measures were taken at the same study visit as the plasma sample was collected. Most participants in this dataset have two study visits and have paired phenotyping and proteomics data from two time points. A Normalized Protein eXpression (NPX) matrix was generated in the full dataset, which included more than one sample for many of the participants. From this NPX matrix, a single sample per person was extracted for use in our proteomics analysis. We agree referencing the additional samples invites confusion, and have attempted to rephrase the section of the paper to clarify:

"To further explore significant S-PrediXcan findings in measured proteomic data, we measured the Olink Explore 3072 panel on 528 stored plasma samples from 271 individuals in CCHC. Normalized Protein eXpression (NPX) was generated in the full dataset, which included multiple measures for many individuals. Our proteomic analyses restricted to one time point per individual and adjusted for age, sex, and five genetic PCs in logistic (T2D) or linear regression (lipids), using the phenotypes measured at the time of specimen collection."

4. The authors show results for Mendelian Randomisation but the precise methods used for this are missing in the methods section

Thank you very much for catching this oversight. We now include the following methodological information describing our Mendelian Randomization work:

"Mendelian Randomization: We assessed evidence of causality for genes identified in S-PrediXcan analyses and found to be significant after study-wide multiple test correction. The GTEx v8 tissue-specific eQTLs of target genes were used as instrumental variables for MR. We performed LD clumping to select eligible instrumental variables in each tissue separately with the LD panels from the 1000 Genome Admixed Americans (AMR), and the R package, "bigsnpr".^{3,4} We used the median weighted MR method from the R package "MendelianRandomization,"⁵ which offers unbiased and reliable estimation even when half of instrumental variables violate the assumptions of

MR and in the presence of genetic pleiotropy.^{5,6}”

Reviewer #2 (Remarks to the Author):

The research design yields valuable insights by conducting a large-scale multi-omics investigation within a non-white demographic, with a particular focus on Hispanic/Latino individuals. This study stands out as one of the few GWAS endeavors targeting diverse populations, involving the recruitment of 63,185 samples from 19 distinct studies. The findings unveil novel genes while also pinpointing variants unique to Hispanic/Latino and African-ancestry populations.

Subsequent to the GWAS, the researchers conducted transcriptome-wide association analysis based on summary statistics. They utilized newly devised islet expression prediction models and meticulously pursued identified genes through differential abundance analyses, utilizing directly measured blood transcriptomic and proteomic data from an independent Hispanic cohort. The methodology appears robust and is elucidated thoroughly, encompassing both GWAS and post-GWAS procedures. The major limitation of this endeavour is to follow it up with Euro-centric GTEx based prediction models and the need for ethnically diverse new datasets is clearly discussed in the manuscript.

Thank you for your time and consideration. We appreciate your shared recognition that we are constrained by Euro-centric nature of extant data, as we, the biomedical research community, work to address this.

Minor comments:

1. Could the authors please specify the dataset utilized for measuring "gene expression in pancreatic islet cells" as the InsPIRE consortium, which is first mentioned in the Discussion section. Could the authors please comment on the percentages of Hispanic/Latinos in the InsPIRE consortium?

Thank you for pointing out this oversight. We now name and cite the source of the islet data in the introduction and better describe this resource in the methods section of the paper:

“The InsPIRE consortium was formed to aggregate human islet RNA-Seq data and genetic data to identify eQTLs and characterize genetic regulation of gene expression in a tissue central to T2D pathogenesis.⁷ Here we leveraged a subset 254 participants of the InsPIRE dataset that were made available to us upon request. Given that samples collected in the US were all described as “Caucasian” and the remainder of sampling occurred in Europe (Geneva, Edmonton, and Oxford) we expect the proportion of the islet sample that is H/L is minimal.”

2. It remains unclear from the abstract whether the research incorporates new

measurement endeavors such as Olink in CCHC, in addition to in-silico inquiries. The authors may wish to rebalance the abstract and underscore this point.

Thank you for pointing out this shortcoming. We now emphasize the addition of our newly developed resource of transcriptomic and proteomic data in the abstract:

“Next, we performed functionally oriented gene-based analyses, identifying genes with multi-omic evidence of impact on T2D or lipid/lipoprotein traits, including genetically regulated expression, Mendelian randomization, and transcriptomic, and proteomic association analyses in a newly developed resource of Hispanic/Latino multi-omics data.”

3. The meaning of "MVP" is not provided in the text. Please clarify its definition.

We have added the full name, “Million Veteran Program” at first usage of the acronym.

Reviewer #3 (Remarks to the Author):

Petty et al reported GWAS meta-analyses and “functionally-oriented” gene analysis of T2D and 4 lipid traits in Hispanic/Latino populations, identifying novel loci as well as replicating previously known signals. The discovery of novel loci near and away from previously discovered genetic associations based on studies that included only European or predominantly-European multi-ancestral studies underscores the value of recently admixed populations in discovery and refining of genetic risk for T2D and dyslipidemia. The newly developed islet model and RNAseq and proteomic resources in Hispanics are important contributions to the field. However, to what extent the RNAseq and proteomics-based analyses validated the present study’s main findings after appropriate correction for multiple testing is unclear. Moreover, the added value of the pancreatic islet resource on top of GTEx in the interpretation of the signals identified has not been clearly demonstrated. The paper requires improvement in presentation and methodological details. Main comments are below:

Thank you for your careful consideration of our manuscript. These comments have improved the impact and clarity of our work. We address the concerns about multiple testing correction of the omics annotations in greater detail below, but briefly, these analyses were intended to aid in functional interpretation of study-wide significant genes implicated by S-PrediXcan analyses, corrected for the number of tissues, genes, and traits at a conservative FDR threshold of 0.01. We have adjusted the language in the paper to better reflect the intention of these annotations. For a detailed list of changes please see response to R3 point 4, below.

We agree there was a missed opportunity to frame the utility of available islet data for improved functional characterization of T2D GWAS signals. We have added the following two sections in the text to clarify our motivation for developing and using novel islet models and the findings from these analyses.

“Because pancreatic islet cells are a central tissue in T2D pathophysiology and exhibit a tissue-specific expression profile (e.g., 40-73% of islet eQTLs replicate in GTEx), especially at T2D relevant genes such as *INS*,⁷ in addition to leveraging extant GTEx prediction models^{8,9} we constructed new models from extant pancreatic islet RNA sequence and associated genomic data.⁷ The InsPIRE consortium was formed to aggregate human islet RNA-Seq data and genetic data to identify eQTLs and characterize genetic regulation of gene expression in a tissue central to T2D pathogenesis.⁷”

“Pancreatic islet cells are not a specific tissue type included in GTEx, yet are a central tissue in T2D pathophysiology that exhibits tissue-specific expression at key T2D genes, including *INS*.^{7,10,11} Therefore, we integrated pancreatic islet cell data from the InsPIRE consortium to generate new S-PrediXcan models...”

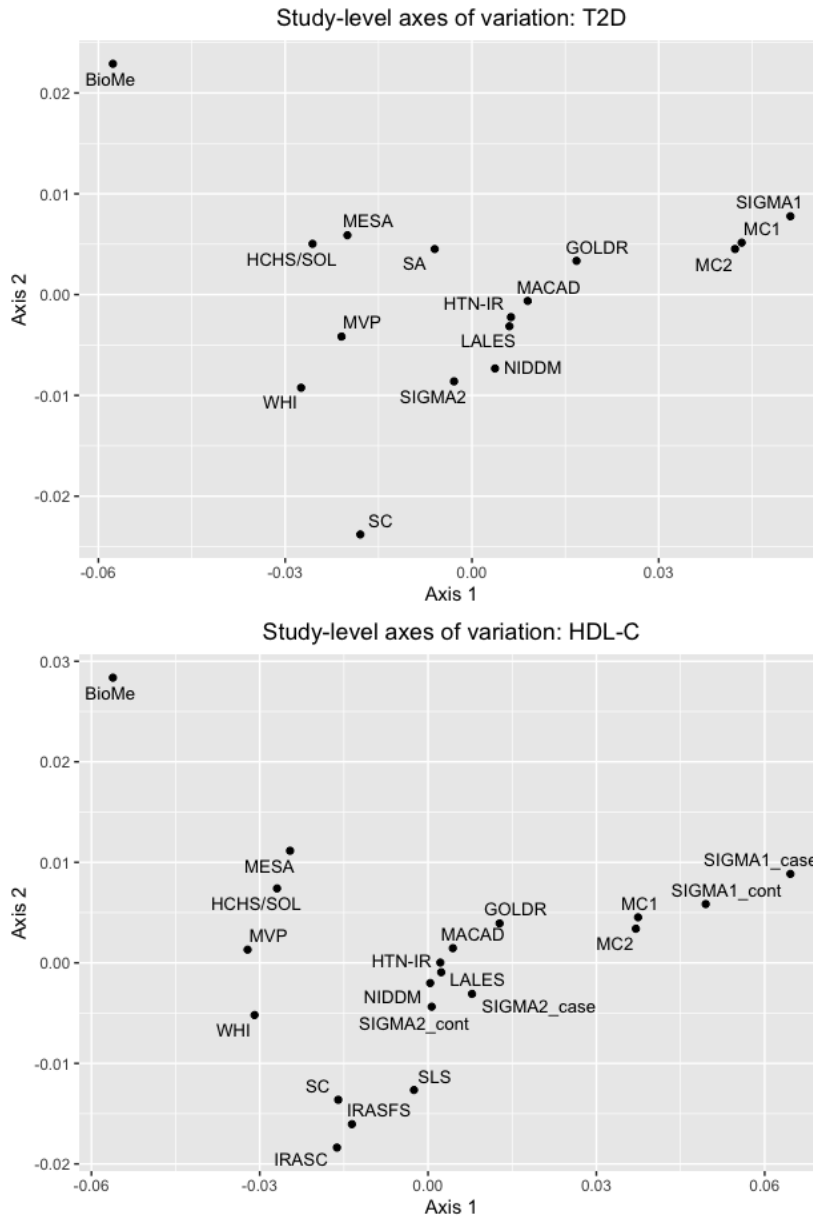
1) The presentation of the paper needs a lot of improvement. There are several instances where the Methods Section lacks details of specific approaches, the results and cited tables do not provide the stated information, and tables are unclear, incomplete, or inconsistent with the text in the Methods Section. I tried to list some of these below (see minor comments), but the authors need to thoroughly review the paper. Although I am not asking the authors to change this, it is obvious that by combining 5 traits in one paper, the Results Section looks overly repetitive and descriptive. Presentation of MR, DEG, and DEP analyses results of five traits in one paragraph was ineffective in communicating the key finding for any one of the five traits.

Thank you for this helpful comment. In addition to commenting on specific changes in response to comments below we have:

- Provided greater details of proteomic analyses in the methods
- Clarified the motivation for our transcriptomic and proteomic analyses
- Added a section on Mendelian Randomization to the methods
- Added a section on the generation of LD scores to the methods
- Added a description of the islet data resource to the methods
- Provided additional motivation for generation and utilization of islet-specific S-PrediXcan models
- Clarified the meaning of novel signals in known loci
- Improved description of sentinel variant identification in single variant analyses
- Improved labeling on all tables and figures
- Expanded results sections for MR and differential transcript and protein abundance annotations
- Highlighted an example of the utility of S-PrediXcan analyses for generating genetic support for biological mechanisms involved in cardiometabolic risk

2) Why was MR-MEGA chosen for meta-analysis of a multi-way admixed Hispanic/Latino population across the 16 studies? MR-MEGA assumes ancestry-related allelic effect heterogeneity as captured by the axes of genetic variation included in the model, and has been effectively demonstrated when samples from different ancestry groups are meta-analyzed. Can the authors clarify the rationale for the choice of MR-

MEGA as a tool for pooling GWAS results coming from the same population group? Was the authors' assumption that there may be "...study-level differences in ancestry (line 452)" based on assessment of the data? Can they demonstrate whether the first or second axis of genetic variation included in the meta-regression capture an ancestry cline in the admixed Hispanics/Latinos? Why were only 2 axes chosen?



All participants in our study self-identify as Hispanic/Latino. Hispanic/Latino is an ethnic group label, which in the US and across Latin America represents multiple unique beliefs, characteristics, and cultures. Hispanic/Latino populations have complex genetic ancestry that reflects recent continental admixture among Native American (heterogeneous and dependent on geographical location), European (mainly, but not only, from the Iberian Peninsula and Southern Europe), and East and West African (at

different geographic locations and times due to the transatlantic slave trade) populations. The majority of our studies recruited participants who predominantly identify as Mexican or Mexican American, however some of our studies are more diverse (e.g., HCHS/SOL, MVP) or from other geographic regions (e.g., SLS). The genomes of Mexicans and Mexican Americans are predominantly admixed between AMR and EUR haplotypes, usually with <5% AFR haplotypes, motivating our use of 2 axes of variation in our MR-MEGA analyses. We have added the text

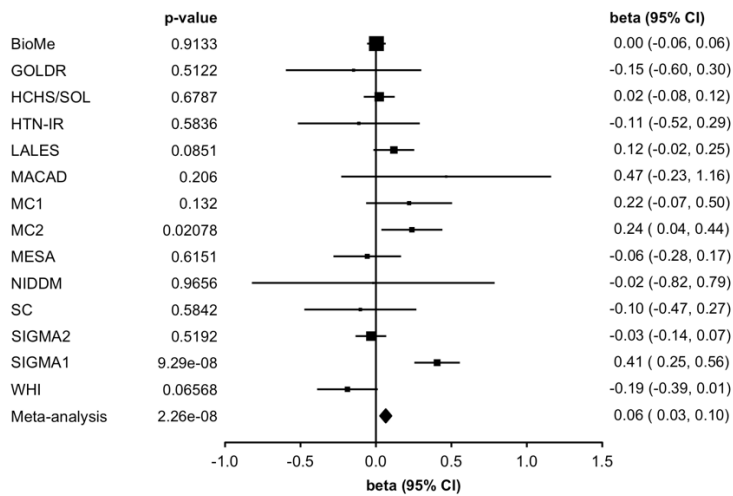
“Two axes of variation were selected to capture major source of admixture within Hispanic/Latino populations (AMR, EUR, and AFR).”

3) The authors acknowledged that the novel MOB3B variant did not replicate, including in Suzuki et al.’s Hispanic analysis. Is it in a consistent effect direction? In Suppl Table 3, the effect direction for this novel association is equally split between the cohorts included in meta-analysis, with increased T2D risk in 7 cohorts and reduced risk in 7 cohorts. The heterogeneity correlated with ancestry is also significant, the p-value being the smallest of all the reported T2D-associated SNP associations. Given these clues, I would urge the authors to explore whether the Hispanic cohorts that were included in this study but not in Suzuki et could explain this. Depicting where the different cohorts fall on the 2 axes of genetic variation may also help examine this in the context of both cohort differences and ancestry cline.

Thank you for this comment. We did look at our study-specific findings. We found this association signal was primarily driven by SIGMA1, which Suzuki et al. have summary statistics from a later release of the data. Given this was near our alpha, our inclination is to acknowledge there will always be type I error. We have added the following text to the paper to better explains this.

“The signal at rs12344703 is primarily driven by effects observed in SIGMA1, however a later release of data from this project did not identify significant effects, suggesting this finding may be type 1 error.”

T2D rs12344703



4) The Methods Section mentions that the functionally-oriented results that were novel would be tested for gene expression-phenotype (DEG) and protein-phenotype (DEP) associations. a) The results section does not mention any finding based on MR and the DEG and DEP findings are not clearly presented.

The analyses of gene expression-phenotype and protein-phenotype analyses were used as a tool for annotating gene-based results from S-PrediXcan. These findings are presented as additional columns in the S-PrediXcan tables. Our genome wide multiomics analyses of T2D and lipid traits are ongoing.

b) Suppl Table 6 has the DEG and DEP association findings, but the table includes known as well as novel genes, which does not match what was mentioned in the methods section about specific follow up of novel findings. The Table does not have multiple testing-corrected p-values, making it difficult to assess whether any of those associations could be considered significant or not. These need to be addressed.

Thank you for pointing out this inconsistency. We have been very conservative in applying multiple testing correction for single variant (Bonferroni) and gene-based tests (Benjamini Hochberg FDR of 0.01 across all tissues, traits, and genes). Due to the abundance of potential approaches of multiple testing correction (RNAseq, proteomics, and MR, many of which are highly correlated and could each potentially be stratified or subset by novelty and tissue), we have opted for transparency in number of tests performed rather than a prescribed penalization strategy for additional lines of evidence. We therefore present raw p-values and number of tests performed for these analyses, allowing readers to interpret significance as they see fit (see modified supplementary tables 6, 11-14). We believe there is value in including known genes in these annotations, as they have greatly varying levels of functional support in the literature. We appreciate your concerns and have removed all claims of significance from our

functional annotations and added a sentence to the discussion regarding the importance of further substantiating these findings.

“All available study-wide significant S-PrediXcan genes were tested ($N_{\text{transcriptomics}}=53$; 127; 113; 152; 138 and $N_{\text{proteomics}}=9$; 34; 31; 39; 37 tests performed for T2D, HDL-C, LDL-C, TC, and TG respectively).”

“Our study reiterates the importance of large-scale studies of non-European-ancestry populations for genetic discovery, presenting novel results and insight into trait biology through multiple lines of transcriptomic and proteomic data, in spite of significantly smaller sample sizes relative to current studies in European-ancestry populations. Our study also demonstrates how precision medicine advances can be made when we integrate GWAS with functional characterization through omics data, revealing new candidate genes and functional variants.”

c) Along this line, the Discussion Section states “multi-tiered robust evidence”. However, the results do not indicate that the SNP analysis results and gene-oriented results yielded overlapping loci. The evidence from proteomics and RNAseq do not seem to strongly support the gene-oriented results if multiple-testing correction were applied. Thus, I would not call the evidence “multi-tiered” or “robust” unless more evidence is provided.

We agree, and we have adjusted the language in the paper to ensure we are not overstating our findings. We do not have an expectation that single variant and gene-based results will implicate all of the same genes, motivating the S-PrediXcan analyses. The S-PrediXcan analyses aggregate evidence of effects across SNP sets, allowing functional discoveries that would be missed by single variant approaches. We then explored the study wide S-PrediXcan in whole blood transcriptome and proteome.

5) It was mentioned that the DEG and DEP analyses were performed to validate findings of the functionally-oriented analyses. How do the authors see this approach in the lens of previous work showing that differentially expressed genes are more likely disease-induced rather than disease-causing (Porcu et al Nat Commun. 2021 Sep 24;12(1):5647. PMID: 34561431)? Would it not be more reasonable to perform this follow up transcriptome and protein analyses on genes showing evidence of causality guided by MR? Could a more integrated analysis be done on potentially overlapping samples with both RNAseq and protein data?

Here we leverage our transcriptomic and proteomic analyses to functionally annotate study-wide significant loci identified through S-PrediXcan. These are valuable analyses because the S-PrediXcan models were built in ostensibly healthy participants of GTEx, while our multiomics analyses capture difference in molecular abundance associated with the relevant phenotype. Joint use of transcriptomic and proteomic data as discovery methods for identifying molecular features associated with disease requires larger samples with omics measures from the same individuals at the same time point to be adequately powered. Here our overlap of measures from the same people at the

same timepoint are low, however we agree that this work would be of interest, and we are actively working on additional typing to pursue this area of research in the future.

6) The authors mention "ANXA4 for T2D" and others as example of genes offering multi-omics support for the gene-oriented results. The results in the supplementary table do not suggest that either the S-PrediXcan or proteomics findings are significant nominally or after multiple-testing correction. Can the authors provide evidence to support these?

We appreciate this feedback, as it gives us a chance to clarify which results we are referring to here. For determining significance in Mendelian randomization results, as expression is highly correlated for most tissues and by chance a specific tissue may have a variant set that better reflects GWAS-significant variants, we used results for each gene for the tissue with the most significant MR findings. That is to say, we conducted MR for all tissues that were significant in the discovery S-PrediXcan analyses and present the MR results with the smallest p-value in MR; this may be a different tissue that was most significant in the discovery analyses. To reduce confusion, we have restructured the tables.

7) The Discussion Section highlighted some novel associations, which is worthy. However, the S-PrediXcan analysis across the 5 traits yielded loci that are predominantly known in previous studies. As stated in the Introduction Section, multi-ancestral GWAS enables deeper interrogation of known loci. Could the authors comment on whether the functionally-oriented findings add to our understanding of the biology of known signals of T2D or lipids, e.g. by narrowing down the likely causal genes, transcript regulation, tissue-specificity etc.

We are very grateful to the reviewer for this suggestion, which prompted us to more deeply explore our functionally oriented S-PrediXcan results. These further explorations led us to identify a gene implicated in our functionally oriented analyses with compelling evidence for a clinically actionable role in type 2 diabetes that has been missed in prior GWAS that had identified this locus. We have now added the following text to our discussion describing the potentially important genetic evidence for *SGLT-1* particularly in populations with Amerindian ancestry.

“The S-PrediXcan analyses provide an opportunity to functionally annotate known loci, adding to our understanding of the biology of these signals by narrowing the likely causal gene(s) at the locus. In the known T2D locus on chromosome 22, we identified significant signals in both our GWAS and S-PrediXcan analyses. In our GWAS, rs16989540, in intron 27 of *DEPDC5*, was significantly associated with T2D status; Open Targets Genetics predicted *YWHAH* to be the causal gene for this signal. This locus has been previously reported in three prior studies, two of which had sample overlap with the present study and in a third study in a Maya population.¹²⁻¹⁴ In all three studies, the signal was mapped to either *DEPDC5* or *YWHAH*. However, S-PrediXcan functionally implicates a different gene in this locus, *SLC5A1*, in small intestine. The role of this gene, also known as *SGLT1*, in type 2 diabetes is supported by an abundance of

non-GWAS evidence. Consistent with our observed direction of effect in small intestine, measured expression of this gene in the small intestine has been shown to be increased in people with type 2 diabetes.^{15,16} Indeed, the FDA recently approved a medication, called sotagliflozin, that targets SGLT-1 and SGLT-2, reducing blood glucose and treating heart failure in T2D patients.^{17,18} It is notable that all three GWAS that reported a signal at this locus included individuals with a large proportion of AMR ancestry, suggesting this might be a key population in which to explore drug efficacy. This is just one example of how functionally oriented analyses, using expression data from multiple tissues, can inform our interpretation of GWAS results and identify clinically actionable targets.”

8) What number of tissues, genes, and phenotypes were considered and how did the authors come up with an FDR of 0.01 for the tissue-agnostic GReX analysis?

We used all GTEx tissues with S-PrediXcan Joint-Tissues Imputations models (49), and for T2D our new islet models. The FDR correction was done across all tissues, genes, and phenotypes. We chose the threshold of 0.01 to be conservative. All results of the S-PrediXcan models will be made publicly available upon publication.

9) Provide details about the specific MR test used, approach, and citation. How was reverse causality tested and validated?

Thank you for noting the lack of details for the MR methods. This was also noted by Reviewer 1; in response to their comment 4, we have now provided these details, also shown here.

“Mendelian Randomization: We assessed evidence of causality for genes identified in S-PrediXcan analyses and found to be significant after study-wide multiple test correction. The GTEx v8 tissue-specific eQTLs of target genes were used as instrumental variables for MR. We performed LD clumping to select eligible instrumental variables in each tissue separately with the LD panels from the 1000 Genome Admixed Americans (AMR), and the R package, "bigsnpr".^{3,4} We performed the median weighted MR method from the R package "MendelianRandomization,"⁵ which offers unbiased and reliable estimation even when half of instrumental variables violate the assumptions of MR and in the presence of genetic pleiotropy.^{5,6}”

Regarding reverse causality, our Mendelian randomization analyses were performed on genes discovered in the S-PrediXcan analyses. Since S-PrediXcan utilizes genetic prediction of gene expression via eQTL-based models, the relationship from eQTL to trait is rooted in genetic variation. Therefore, the association is, by definition, directional, which is why we did not test for reverse causality.

10) Provide a specific and clear comparison between Hispanic/Latinos and other US-populations (either non-Hispanics or another race/ethnic group in the US) to support the stated point that “Many cardiometabolic diseases are more prevalent in Hispanics/Latinos (lines 186-188)” and "...in a population with significant

cardiometabolic health disparities". Comparisons using prevalence figures based on all US populations is ambiguous.

Thank you for this feedback, we have updated the text, and changed the comparator group to Non-Hispanic White populations.

"Many cardiometabolic diseases are more prevalent in Hispanic/Latino populations compared to Non-Hispanic White populations; for example, recent prevalence estimates of type 2 diabetes (T2D) and low high-density lipoprotein cholesterol (HDL-C) in Hispanic/Latino populations are 15.5% and 21.9%, respectively, compared to 13.6% and 16.6% in Non-Hispanic White populations.^{19,20}"

Minor comments:

- Suppl Table 2 has four columns specific about male- and female-stratified analysis. The text does not talk about sex-stratified analysis. Please clarify. Also there are values "NA" and "none", which makes it confusing. Does "none" mean no covariate was adjusted for and "NA" means the whole cohort was not included?

Thank you for pointing out this oversight. The stratified columns have been removed, as those meta-analyses were not completed. Your interpretation of NA vs. none is correct. Since the "none" values were removed with stratified columns, no additional changes have been made.

- Suppl Tables 3-6 contain beta and SE. Clarify how these were obtained as MR-MEGA does not provide these statistics.

The beta and SE columns reflect the beta_0 and SE_0 results from MR-MEGA. We have added a note, also shown below, in the table headers that clarifies this.

*** Beta and SE are beta_0 and SE_0 results from MR-MEGA."

- Clarify whether the single variant-based analyses and gene-based analysis with S-PrediXcan identified overlapping signals or not.

We have now annotated tables and are adding a Miami plot for better comparison of our single variant and S-PrediXcan findings in Supplementary Figure 1.

- The discussion section has several redundant information, and should be significantly reduced.

Thank you for this advice. We have now shortened the discussion and reduced redundancy.

- Suppl Table 3: The column "Neff" is empty for the "DIAMANTE South Asian results"

Neff is not available for the DIAMANTE South Asian results, so we have revised these columns to present total N in the header for each group for consistency.

- Suppl Table 4: The title should include "BMI-adjusted". Otherwise, the titles of STable 3 and 4 are identical and confusing.

Thank you for catching this oversight. The title is now "Supplementary Table 4. T2D adjusted for BMI single variant meta-regression lookups for T2D single variant findings (in Supplementary Table 3)." We have also revised our presentation of results in Supplementary Table 4 to more clearly align with the intention of this analysis, as noted as well in our response to the Line 276 comment.

- Across Suppl Tables 4, 7, 8, 9, 10, the 'fine mapping' results are presented confusingly under the columns "known variants plus or minus 1 Mb" and "known variants in credible set(s)". "yes" and "known" look the same. Do you mean to say "novel" instead of "known", and what is "no", "NA"? Clear description is needed.

We have adjusted the wording for these columns. It is now "yes" or "no; novel locus" for the "plus or minus 1Mb known variants". For "known variants in credible set(s)" it is yes, no, or NA in the case of novel loci.

- Headers need description for many Supplementary Tables, in particular suppl Table 6.

Apologies for the lack of legends in these tables. We have gone through each table and ensured there are complete legends.

- Suppl Table 6: Where the gene-based results are "known", provide information about the known locus and its associated PMID.

We appreciate this suggestion. We have added several new pieces of information to this table, namely:

- PMID for any GWAS that names the gene or
- PMID for a GWAS with a genome-wide significant variant within 1Mb of the gene, as this is the window used in generating PrediXcan models
- Chromosome and position for each gene, and sorted results by genomic coordinates
- Limited MR results to the significant tissue with the more lowest p-value for MR.

Some discrepancies arose in the process of adding PubMed IDs. We therefore regenerated Figure 1. This did not impact any of the genes highlighted in the manuscript.

- Line 191: Edit "...which is has a prevalence of..."

Thank you, we have removed the extra word.

- Line 222: Provide citation to the statement "while much of the genetic architecture of cardiometabolic disease risk is shared across populations and environmental contexts,"

We have added the following references for this statement: ^{21,22}

Bien SA, Pankow JS, Haessler J, Lu Y, Pankratz N, Rohde RR, Tamuno A, Carlson CS, Schumacher FR, Buzkova P, et al. Transethnic insight into the genetics of glycaemic traits: fine-mapping results from the Population Architecture using Genomics and Epidemiology (PAGE) consortium. *Diabetologia*. 2017;60:2384-2398. doi: 10.1007/s00125-017-4405-1

Wojcik GL, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, Highland HM, Patel YM, Sorokin EP, Avery CL, et al. Genetic analyses of diverse populations improves discovery for complex traits. *Nature*. 2019;570:514-518. doi: 10.1038/s41586-019-1310-4

- Line 276: The description needs correction because BMI-adjusted genetic effect on T2D does not mean that the effect is driven by mediation by BMI, rather suggests BMI-independence.

Thank you for pointing out this confusing language. We have revised the description of the purpose of this analysis this in the manuscript, shown below. We have also revised our presentation of results in Supplementary Table 4 to more clearly align with the intention of this analysis.

“we performed an additional meta-analysis adjusting for BMI to explore if effects at our top findings are modified by BMI.”

- Line 280-281: States “Sentinel variants, or the variant with the lowest p-value, were identified for each region with $p < 5 \times 10^{-8}$ ”. What is the operational definition used to assign signals into a “region”?

The region was defined as 1MB in either direction, to maintain consistency with the cis region with S-PrediXcan. The list of sentinel variants was created by iteratively selecting the most significant variant and removing all variants within 1MB until there were no additional significant variants.

- Line 287-291: The flow can be improved by re-ordering sentences - define "known" and "novel" signal, followed by "novel signal in a known locus"

We agree this ordering improves clarity and have made the suggested change.

- Line 297-299: Why was the "African ancestry group (MEDIA)" not included in the replication effort?

The MEDIA analyses have not been published and thus are not available to us for replication at this time.

- Line 297: The last part of the sentence "...top-level multi-ancestry meta-analysis" is unclear.

We have modified the text to improve clarity, shown below.

"Five groups were assembled, including our Hispanic/Latino group, an African ancestry group (MEDIA), an East Asian ancestry group (AGEN-T2D),²³ a European ancestry group (DIAGRAM),²⁴ a South Asian ancestry group (SA-T2D),²⁵ and the multi-ancestry meta-analysis of these five groups.²⁶"

- Line 380: The first sentence is not clear because it lumps the T2D GWAS results from the model without BMI (which is the main model according to the methods) and the BMI-adjusted model. Present the number of associations from the main model, then state that the number of loci that remained and the number of new loci that emerged after BMI-adjustment. This aligns with the goals of the BMI-adjustment as stated in the methods (with edits – see my other comment).

Thank you for pointing out this unclear statement. We have revised the text to read:

"Twenty genome-wide significant loci were associated with T2D (Supplementary Tables 3; Supplementary Figures 1a, 2a).; all six signals of the signals remained significant after adjusting for BMI (Supplementary Table 4, Supplementary Figure 1b)."

- Line 386-388: It says 2 T2D SNPs are "novel signals in known loci", and references suppl Table 5. However, there is no information in the referenced table to know which SNPs are novel. The text is also unclear which SNPs these are, why there are distinct from the previous loci, and does not provide appropriate citation.

Thank you, we agree that this was unclear as previously presented. We have now added a column to Supplementary Table 5 that gives the "novel, "novel in known", or "known" classification for each signal. We have also clarified the designations in Supplementary Tables 3 and 7-10, where two columns, "known variants plus or minus 1 Mb" and "known variants in credible set(s)" give further information on novelty status for each signal, as mentioned in our response to the "Across Suppl Tables 4, 7, 8, 9, 10" comment. "Novel in known" signals are those for which there are no known variants in the 95% credible set, but there are known variants within a Mb of our sentinel variant.

References

1. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, Powell C, Vedantam S, Buchkovich ML, Yang J, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015;518:197-206. doi: 10.1038/nature14177
2. Graham SE, Clarke SL, Wu KH, Kanoni S, Zajac GJM, Ramdas S, Surakka I, Ntalla I, Vedantam S, Winkler TW, et al. Author Correction: The power of genetic diversity in genome-wide association studies of lipids. *Nature*. 2023;618:E19-E20. doi: 10.1038/s41586-023-06194-2
3. Prive F, Aschard H, Ziyatdinov A, Blum MGB. Efficient analysis of large-scale genome-wide data with two R packages: bigstatsr and bigsnpr. *Bioinformatics*. 2018;34:2781-2787. doi: 10.1093/bioinformatics/bty185
4. Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, et al. A global reference for human genetic variation. *Nature*. 2015;526:68-74. doi: 10.1038/nature15393
5. Yavorska OO, Burgess S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int J Epidemiol*. 2017;46:1734-1739. doi: 10.1093/ije/dyx034
6. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genetic epidemiology*. 2016;40:304-314. doi: 10.1002/gepi.21965
7. Vinuela A, Varshney A, van de Bunt M, Prasad RB, Asplund O, Bennett A, Boehnke M, Brown AA, Erdos MR, Fadista J, et al. Genetic variant effects on gene expression in human pancreatic islets and their implications for T2D. *Nat Commun*. 2020;11:4912. doi: 10.1038/s41467-020-18581-8
8. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, Eyer AE, Denny JC, Consortium GT, Nicolae DL, et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet*. 2015;47:1091-1098. doi: 10.1038/ng.3367
9. Zhou D, Jiang Y, Zhong X, Cox NJ, Liu C, Gamazon ER. A unified framework for joint-tissue transcriptome-wide association and Mendelian randomization analysis. *Nat Genet*. 2020;52:1239-1246. doi: 10.1038/s41588-020-0706-2
10. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, Payne AJ, Steinthorsdottir V, Scott RA, Grarup N, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet*. 2018;50:1505-1513. doi: 10.1038/s41588-018-0241-6
11. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*. 2020;369:1318-1330. doi: 10.1126/science.aaz1776
12. Dominguez-Cruz MG, Munoz ML, Totomoch-Serra A, Garcia-Escalante MG, Burgueno J, Valadez-Gonzalez N, Pinto-Escalantes D, Diaz-Badillo A. Pilot

- genome-wide association study identifying novel risk loci for type 2 diabetes in a Maya population. *Gene*. 2018;677:324-331. doi: 10.1016/j.gene.2018.08.041
13. Vujkovic M, Keaton JM, Lynch JA, Miller DR, Zhou J, Tcheandjieu C, Huffman JE, Assimes TL, Lorenz K, Zhu X, et al. Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nat Genet*. 2020;52:680-691. doi: 10.1038/s41588-020-0637-y
 14. Mahajan A, Spracklen CN, Zhang W, Ng MCY, Petty LE, Kitajima H, Yu GZ, Rueger S, Speidel L, Kim YJ, et al. Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nat Genet*. 2022;54:560-572. doi: 10.1038/s41588-022-01058-3
 15. Dyer J, Wood IS, Palejwala A, Ellis A, Shirazi-Beechey SP. Expression of monosaccharide transporters in intestine of diabetic humans. *Am J Physiol Gastrointest Liver Physiol*. 2002;282:G241-248. doi: 10.1152/ajpgi.00310.2001
 16. Fiorentino TV, Suraci E, Arcidiacono GP, Cimellaro A, Mignogna C, Presta I, Andreozzi F, Hribal ML, Perticone F, Donato G, et al. Duodenal Sodium/Glucose Cotransporter 1 Expression Under Fasting Conditions Is Associated With Postload Hyperglycemia. *J Clin Endocrinol Metab*. 2017;102:3979-3989. doi: 10.1210/jc.2017-00348
 17. Avgerinos I, Karagiannis T, Kakotrichi P, Michailidis T, Liakos A, Matthews DR, Tsapas A, Bekiari E. Sotagliflozin for patients with type 2 diabetes: A systematic review and meta-analysis. *Diabetes Obes Metab*. 2022;24:106-114. doi: 10.1111/dom.14555
 18. Administration UFaD. INPEFA (sotagliflozin) tablets, for oral use. 2023.
 19. Prevention CfDCa. National Diabetes Statistics Report, 2021. Services USDoHaH; 2024.
 20. Carroll MD, Fryar CD. Total and High-density Lipoprotein Cholesterol in Adults: United States, 2015–2018. NCHS Data Brief, no 363. Hyattsville, MD: Statistics NCfH; 2020.
 21. Bien SA, Pankow JS, Haessler J, Lu Y, Pankratz N, Rohde RR, Tamuno A, Carlson CS, Schumacher FR, Buzkova P, et al. Transethnic insight into the genetics of glycaemic traits: fine-mapping results from the Population Architecture using Genomics and Epidemiology (PAGE) consortium. *Diabetologia*. 2017;60:2384-2398. doi: 10.1007/s00125-017-4405-1
 22. Wojcik GL, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, Highland HM, Patel YM, Sorokin EP, Avery CL, et al. Genetic analyses of diverse populations improves discovery for complex traits. *Nature*. 2019;570:514-518. doi: 10.1038/s41586-019-1310-4
 23. Spracklen CN, Horikoshi M, Kim YJ, Lin K, Bragg F, Moon S, Suzuki K, Tam CHT, Tabara Y, Kwak SH, et al. Identification of type 2 diabetes loci in 433,540 East Asian individuals. *Nature*. 2020;582:240-245. doi: 10.1038/s41586-020-2263-3
 24. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, Payne AJ, Steinthorsdottir V, Scott RA, Grarup N. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nature genetics*. 2018;50:1505.

25. Loh M, Zhang W, Ng HK, Schmid K, Lamri A, Tong L, Ahmad M, Lee JJ, Ng MCY, Petty LE, et al. Identification of genetic effects underlying type 2 diabetes in South Asian and European populations. *Commun Biol.* 2022;5:329. doi: 10.1038/s42003-022-03248-5
26. Suzuki K, Hatzikotoulas K, Southam L, Taylor HJ, Yin X, Lorenz KM, Mandla R, Huerta-Chagoya A, Melloni GEM, Kanoni S, et al. Genetic drivers of heterogeneity in type 2 diabetes pathophysiology. *Nature.* 2024. doi: 10.1038/s41586-024-07019-6