



# The power of TOPMed imputation for the discovery of Latino-enriched rare variants associated with type 2 diabetes

Alicia Huerta-Chagoya<sup>1,2,3,4</sup> · Philip Schroeder<sup>1,2,5</sup> · Ravi Mandla<sup>1,2,5</sup> · Aaron J. Deutsch<sup>1,2,5,6</sup> · Wanying Zhu<sup>7</sup> · Lauren Petty<sup>7</sup> · Xiaoyan Yi<sup>8</sup> · Joanne B. Cole<sup>1,2,9,10,11</sup> · Miriam S. Udler<sup>2</sup> · Peter Dornbos<sup>1</sup> · Bianca Porneala<sup>12</sup> · Daniel DiCorpo<sup>13</sup> · Ching-Ti Liu<sup>13</sup> · Josephine H. Li<sup>1,2,5,6</sup> · Lukasz Szczerbiński<sup>1,2,5,14,15</sup> · Varinderpal Kaur<sup>5</sup> · Joohyun Kim<sup>16</sup> · Yingchang Lu<sup>16</sup> · Alicia Martin<sup>1,17</sup> · Decio L. Eizirik<sup>8,18</sup> · Piero Marchetti<sup>19</sup> · Lorella Marselli<sup>19</sup> · Ling Chen<sup>2</sup> · Shylaja Srinivasan<sup>20</sup> · Jennifer Todd<sup>21</sup> · Jason Flannick<sup>1,22</sup> · Rose Gubitosi-Klug<sup>23</sup> · Lynne Levitsky<sup>24</sup> · Rachana Shah<sup>25</sup> · Megan Kelsey<sup>26</sup> · Brian Burke<sup>27</sup> · Dana M. Dabelea<sup>28</sup> · Jasmin Divers<sup>29</sup> · Santica Marcovina<sup>30</sup> · Lauren Stalbow<sup>31</sup> · Ruth J. F. Loos<sup>31,32</sup> · Burcu F. Darst<sup>33</sup> · Charles Kooperberg<sup>33</sup> · Laura M. Raffield<sup>34</sup> · Christopher Haiman<sup>35,36</sup> · Quan Sun<sup>37</sup> · Joseph B. McCormick<sup>38</sup> · Susan P. Fisher-Hoch<sup>38</sup> · Maria L. Ordoñez<sup>39</sup> · James Meigs<sup>1,6,12</sup> · Leslie J. Baier<sup>40</sup> · Clicerio González-Villalpando<sup>41</sup> · Maria Elena González-Villalpando<sup>41</sup> · Lorena Orozco<sup>42</sup> · Lourdes García-García<sup>43</sup> · Andrés Moreno-Estrada<sup>44</sup> · Mexican Biobank · Carlos A. Aguilar-Salinas<sup>45</sup> · Teresa Tusie<sup>3,4</sup> · Josée Dupuis<sup>13</sup> · Maggie C. Y. Ng<sup>16</sup> · Alisa Manning<sup>46</sup> · Heather M. Highland<sup>47</sup> · Miriam Cnop<sup>8,48</sup> · Robert Hanson<sup>49</sup> · Jennifer Below<sup>7</sup> · Jose C. Florez<sup>1,2,5,6,9,50</sup> · Aaron Leong<sup>1,2,5,6,12,50</sup> · Josep M. Mercader<sup>1,2,5,9</sup> 

Received: 9 September 2022 / Accepted: 3 February 2023 / Published online: 6 May 2023

© The Author(s) 2023

## Abstract

**Aims/hypothesis** The Latino population has been systematically underrepresented in large-scale genetic analyses, and previous studies have relied on the imputation of ungenotyped variants based on the 1000 Genomes (1000G) imputation panel, which results in suboptimal capture of low-frequency or Latino-enriched variants. The National Heart, Lung, and Blood Institute (NHLBI) Trans-Omics for Precision Medicine (TOPMed) released the largest multi-ancestry genotype reference panel representing a unique opportunity to analyse rare genetic variations in the Latino population. We hypothesise that a more comprehensive analysis of low/rare variation using the TOPMed panel would improve our knowledge of the genetics of type 2 diabetes in the Latino population.

**Methods** We evaluated the TOPMed imputation performance using genotyping array and whole-exome sequence data in six Latino cohorts. To evaluate the ability of TOPMed imputation to increase the number of identified loci, we performed a Latino type 2 diabetes genome-wide association study (GWAS) meta-analysis in 8150 individuals with type 2 diabetes and 10,735 control individuals and replicated the results in six additional cohorts including whole-genome sequence data from the All of Us cohort.

**Results** Compared with imputation with 1000G, the TOPMed panel improved the identification of rare and low-frequency variants. We identified 26 genome-wide significant signals including a novel variant (minor allele frequency 1.7%; OR 1.37,  $p=3.4 \times 10^{-9}$ ). A Latino-tailored polygenic score constructed from our data and GWAS data from East Asian and European populations improved the prediction accuracy in a Latino target dataset, explaining up to 7.6% of the type 2 diabetes risk variance.

---

Jose C. Florez, Aaron Leong and Josep M. Mercader jointly directed this work.

---

A full list of the Mexican Biobank contributors can be found in the electronic supplementary material (ESM).

---

✉ Alicia Huerta-Chagoya  
ahuerta@broadinstitute.org; achagoya@ciencias.unam.mx

✉ Josep M. Mercader  
mercader@broadinstitute.org

Extended author information available on the last page of the article

## Research in context

### What is already known about this subject?

- Latino populations are underrepresented in genetic studies of type 2 diabetes, despite having higher prevalence of the disease
- Underrepresentation extends to the reference panels needed for genotype imputation of population-specific low-frequency/rare variants
- The National Heart, Lung, and Blood Institute (NHLBI) Trans-Omics for Precision Medicine (TOPMed) programme has released the largest multi-ancestry imputation panel, representing a unique opportunity to explore low-frequency/rare variants

### What is the key question?

- Does a more comprehensive analysis of low-frequency/rare variation using the TOPMed panel improve our knowledge type 2 diabetes genetics in the Latino population?

### What are the new findings?

- Compared with the 1000 Genomes (1000G) imputation panel, the TOPMed panel imputed 6.5-fold more variants with <0.1% frequency and, for >0.1% frequency variants, it retrieved >80% of whole-exome sequenced variants
- We identified two novel low-frequency variants associated with  $\geq 1.4$ -fold type 2 diabetes risk in the Latino population; these variants were either better, or only, imputed with the TOPMed panel
- We generated a multi-ancestry polygenic score (PS) that outperformed a European-based PS

### How might this impact on clinical practice in the foreseeable future?

- The novel Latino-enriched variants may reveal therapeutic targets for type 2 diabetes and precision medicine avenues and, although future studies are needed to assess the clinical utility of PSs, the improved PS suggests that, by applying more comprehensive approaches to address diversity, we may improve our ability to identify Latino individuals at high genetic risk for type 2 diabetes

**Conclusions/interpretation** Our results demonstrate the utility of TOPMed imputation for identifying low-frequency variants in understudied populations, leading to the discovery of novel disease associations and the improvement of polygenic scores.

**Data availability** Full summary statistics are available through the Common Metabolic Diseases Knowledge Portal (<https://t2d.hugeamp.org/downloads.html>) and through the GWAS catalog (<https://www.ebi.ac.uk/gwas/>, accession ID: GCST90255648). Polygenic score (PS) weights for each ancestry are available via the PGS catalog (<https://www.pgscatalog.org>, publication ID: PGP000445, scores IDs: PGS003443, PGS003444 and PGS003445).

**Keywords** GWAS meta-analysis · Latino population · Polygenic score · TOPMed imputation · Type 2 diabetes

### Abbreviations

1000G	1000 Genomes	LD	Linkage disequilibrium
AGEN	Asian Genetic Epidemiology Network	LoF	Loss-of-function
AF	Allele frequency	MAF	Minor allele frequency
CADD	Combined Annotation Dependent Depletion	METS	Mexican Metabolic Syndrome (cohort)
CCHC	Cameron County Hispanic Cohort	MGB	Mass General Brigham Biobank
DIAMANTE	Diabetes Meta-Analysis of Trans-Ethnic association studies	MHTG	Mexican Hypertriglyceridemia
FNRS	Fonds National de la Recherche Scientifique	MXBB	Mexican Biobank
GERA	Genetic Epidemiology Research on Aging	Neff	Effective sample size
GTEX	Genotype–Tissue Expression project	NHGRI	National Human Genome Research Institute
GWAS	Genome-wide association study	NHLBI	National Heart, Lung, and Blood Institute
		NIMHD	National Institute on Minority Health and Health Disparities

PAGE	Population Architecture using Genomics and Epidemiology
PC	Principal component
PRODIGY	Progress in Diabetes Genetics in Youth
PS	Polygenic score
SIGMA	Slim Initiative for Genomic Medicine in the Americas
T2D-GENES	Type 2 Diabetes Genetics Exploration by Next-generation sequencing in multi-Ethnic Samples
TODAY	Treatment Options for Type 2 Diabetes in Adolescents and Youth
TOPMed	NHLBI Trans-Omics for Precision Medicine
TPM	Transcripts per million
UKBB	UK Biobank
WES	Whole-exome sequencing

## Introduction

Latino is a diverse ethnic group recently admixed from Native American, European and African ancestries, with a high prevalence of metabolic disorders including type 2 diabetes. Although genetic studies in the Latino population are limited, they have revealed unexpected pathways and potential therapeutic targets for type 2 diabetes [1–4]. This is the case for a Native American haplotype within the *SLC16A11* gene identified as the main genetic contributor to type 2 diabetes in the Latino population [1, 4], a rare risk variant within *HNF1A* unique to Latino population [2] and a loss-of-function (LoF) Latino-enriched variant within *IGF2* associated with a 22% decrease in the odds of type 2 diabetes in heterozygous carriers [3].

Unlike genetically homogenous populations, the complex linkage disequilibrium (LD) structure of admixed populations imposes challenges in implementing statistical methods that are crucial to maximise genetic discoveries [5]. This is especially relevant for genotype imputation, a method used to estimate the genotype probabilities at genetic variants that have not been experimentally genotyped [6]. A major factor limiting the accuracy of genotype imputation in Latino samples has been the poor representation of their haplotypes in the reference panels (i.e. 352 from the latest version of the 1000 Genomes [1000G] imputation model) [7]. The multi-ancestry National Heart, Lung, and Blood Institute (NHLBI) Trans-Omics for Precision Medicine (TOPMed) programme has released a reference panel for genotype imputation that includes the highest sequencing coverage (i.e. 30×) and the largest number of reference samples (i.e. 97,256) to date, of which ~15% are from Latino individuals. It has been shown to increase the number of well-imputed low-frequency variants

in the Hispanic Community Health Study/Study of Latinos (HCHS/SOL) [8, 9].

We hypothesised that by boosting the identification of variants in Latino samples with the recently released TOPMed reference panel, we would improve our knowledge of the genetic architecture of type 2 diabetes in the Latino population. The 1000G (1000G) panel was chosen as a comparison, since, besides TOPMed, it has the largest number of Latino samples. We performed a type 2 diabetes genome-wide association study (GWAS) meta-analysis, as well as association analyses on a collection of related phenotypes from TOPMed Latino imputed datasets to allow the interpretation of our novel variants that had low frequencies or were absent in other publicly available biobanks that mainly contained individuals of European ancestry. Finally, we leveraged the generated GWAS data to develop, in combination with GWAS data from other ancestries, a type 2 diabetes polygenic score (PS) for the Latino population.

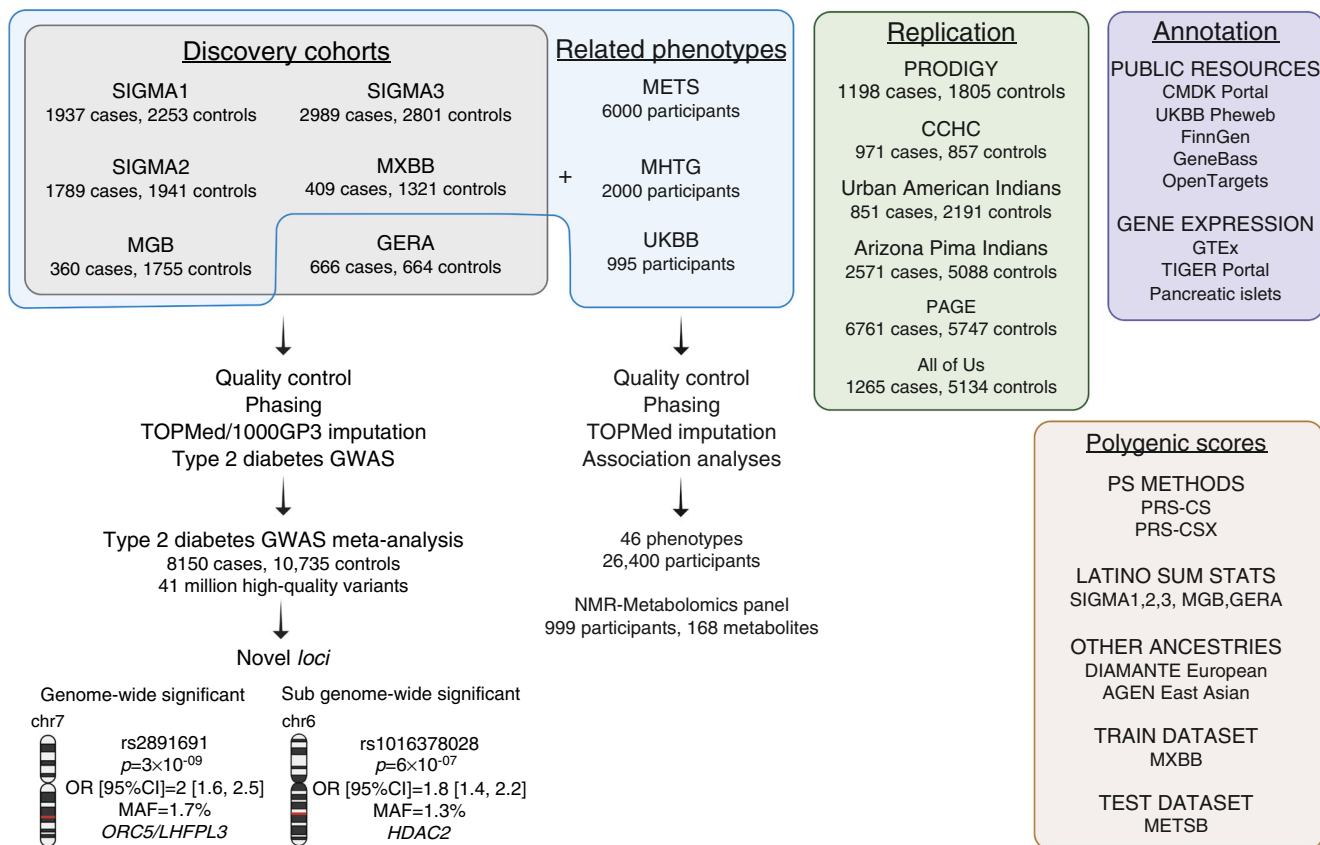
## Methods

Detailed descriptions of the methods are given in electronic supplementary material (ESM) [Methods](#).

**Discovery sample** We aggregated data from six Latino cohorts with a sample size of 18,885 individuals (8150 with type 2 diabetes [cases] and 10,735 without [controls]): the Slim Initiative for Genomic Medicine in the Americas (SIGMA) [1–3]; the Mexican Biobank (MXBB) [10]; the Mass General Brigham (MGB) Biobank [11]; and the Genetic Epidemiology Research on Aging (GERA) [12] (Fig. 1 and ESM Table 1). We selected Latino samples based on their genetically estimated ancestry using principal components (PCs) and Admixture v1.3.0 [13] (ESM Fig. 1). All human research was approved by the relevant Institutional Review Boards and conducted according to the Declaration of Helsinki. All participants provided written informed consent.

**Genotyping and imputation** Genotyping was done using several commercially available genome-wide arrays, and for a subset of the samples ( $N=9520$ ), we integrated whole-exome sequencing (WES) (ESM Table 1). We applied pre-imputation quality control to each dataset separately. Clean datasets were phased using SHAPEIT2 v2 [14]. For comparison purposes, we imputed the phased haplotypes using both 1000G Phase3 version 5 [15] and TOPMed reference panels freeze 8 [8].

**Imputation performance evaluation** We evaluated the performance of TOPMed and 1000G imputations by summarising



**Fig. 1** General overview of the study. Six cohorts of admixed Latino ancestry, representing a total of 8150 type 2 diabetes cases and 10,735 controls, were imputed with the TOPMed and 1000G Phase 3 panels (grey box). A type 2 diabetes GWAS meta-analysis of the imputed variants resulted in the identification of two novel loci, which were tested for replication in six additional Latino cohorts (green box). They were also interrogated for association with a collection of phenotypes in eight

Latino cohorts (blue box) and for functional evidence in multiple available resources (purple box). The generated Latino type 2 diabetes GWAS data were used, in combination with GWAS from other ancestries, to construct ancestry-specific and multi-ancestry type 2 diabetes PSs (brown box). CMDK, Common Metabolic Disease Knowledge; sum stats, summary statistics

the chromosome-wise  $r^2$  quality measure and the number of well-imputed ( $r^2 \geq 0.8$ ) variants at different allele frequency (AF) thresholds. We used available WES data from the SIGMA3 cohort and estimated the proportion of the sequenced variants in chromosome 22 that were well-imputed with TOPMed and 1000G panels at different WES AF thresholds. We used SnpEff v4.3 [16] to annotate the WES variants. We calculated the effective sample size (Neff) needed to reach 80% statistical power to detect genome-wide significant associations ( $\alpha=5 \times 10^{-8}$ ) at different effect sizes and AFs covered by the imputations (Fig. 2c).

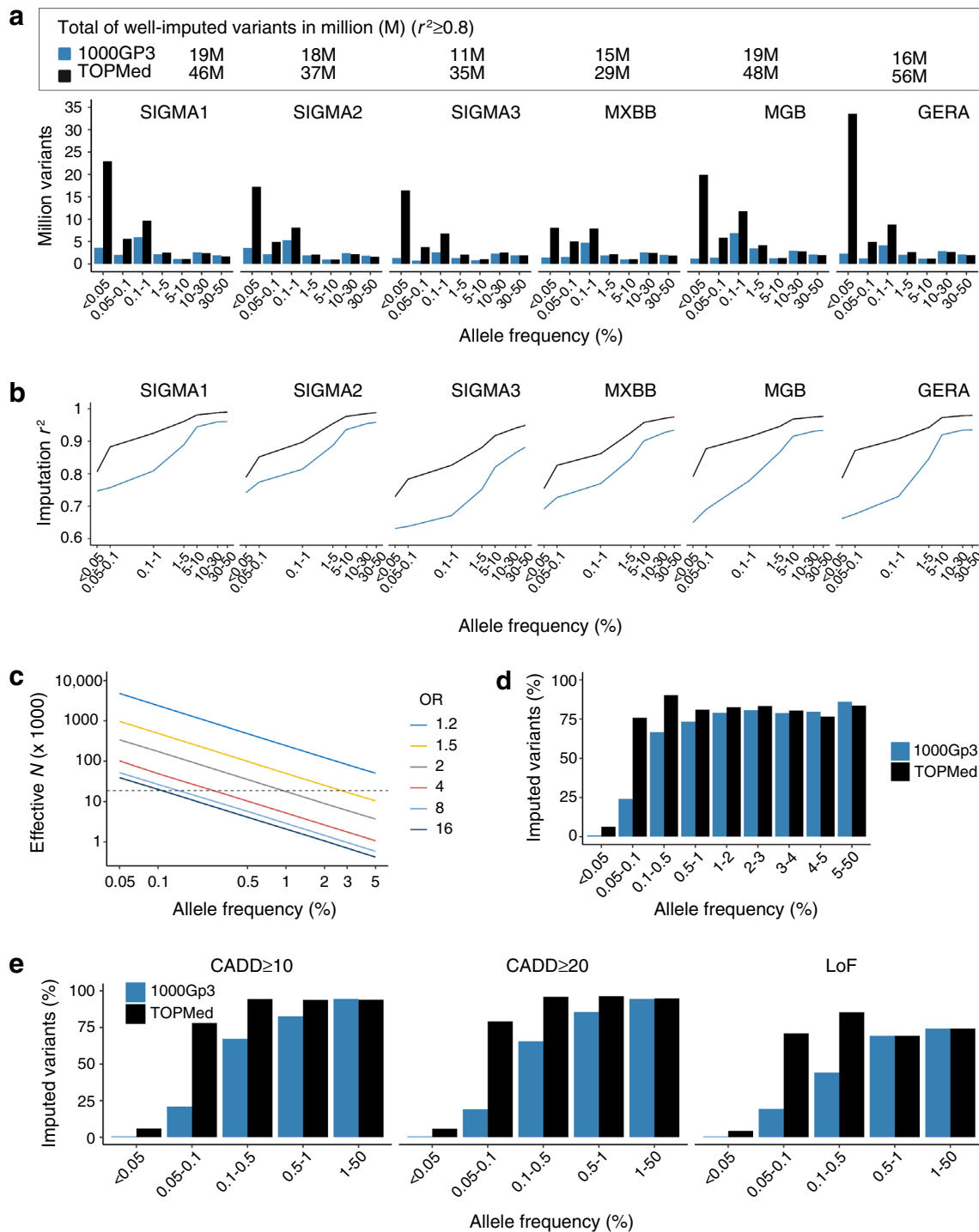
**Type 2 diabetes GWAS meta-analysis** Association analyses were performed in each cohort with SNPTTEST v2.5.4 [17]. Models were adjusted for sex, age, BMI and ten PCs to account for population structure. We ran additional models without adjusting for BMI. Only well-imputed variants ( $r^2 \geq 0.5$ ) were meta-analysed using the inverse of the

corresponding squared SEs in METALv2011-03-25 [18] We used a standard GWAS significance threshold of  $p < 5 \times 10^{-8}$ .

We performed LD-based clumping on the genome-wide significant variants to keep one representative variant per region of LD. If the lead SNP lay within a previously reported type 2 diabetes locus, we defined it as conditionally distinct if showing evidence of residual association ( $p < 5 \times 10^{-5}$ ) after conditioning on each of the reported variants.

Variants with sub-genome-wide significance ( $p < 1 \times 10^{-6}$ ) that were only imputed with the TOPMed panel, showed increased frequency in the Latino population and were >250 kb from other reported genome-wide significant variants from European or East Asian ancestry large consortia [19, 20] were considered for further investigation.

**Replication sample** Variants associated with type 2 diabetes at genome-wide and sub-genome-wide significance were tested for replication in six independent cohorts: the Cameron County Hispanic Cohort (CCHC) [21]; the Urban American



**Fig. 2** Performance of the TOPMed reference panel for the imputation of Latino samples. **(a)** Number of chromosome-wide well-imputed variants (imputation  $r^2 \geq 0.8$ ) by AF for each analysed cohort when using the 1000GP3 (blue) or the TOPMed (black) reference panels. **(b)** Average chromosome-wide imputation quality by AF for each analysed cohort when using the 1000GP3 (blue) or the TOPMed (black) reference panels. **(c)** Effective sample size required for reaching 80% statistical power to detect genome-wide significant signals at different effect sizes (OR). The

dotted lines show the discovery effective sample size of this study. **(d)** Percentage of the exome sequenced variants in chromosome 22 that could be imputed when using the 1000GP3 (blue) or the TOPMed (black) reference panels. **(e)** Percentage of the exome sequenced LoF and deleterious predicted variants based on CADD score in chromosome 22 that could be imputed when using the 1000GP3 (blue) or the TOPMed (black) reference panels



Indians and Arizona Pima Indians cohorts [22]; the Population Architecture using Genomics and Epidemiology (PAGE) study [23]; the All of Us Research Program [24]; and the Progress in Diabetes Genetics in Youth (PRODIGY), which comprises the Treatment Options for Type 2 Diabetes in Adolescents and Youth (TODAY) [25], the SEARCH for Diabetes in Youth studies [26], the Type 2 Diabetes Genetics Exploration by Next-generation sequencing in multi-Ethnic Samples (T2D-GENES) cohorts and the Mexican Metabolic Syndrome (METS) cohort [27] (ESM Table 2).

**Association with type 2 diabetes-related phenotypes** Given the lack of large-scale publicly available biobanks with Latino samples that may allow for better characterisation of our novel signals, we assembled a collection of cohorts to perform association analyses to several type 2 diabetes-related traits comprising 46 glycaemic, anthropometric and lipid traits. In addition to five of the Latino cohorts analysed in the type 2 diabetes meta-analysis (i.e. SIGMA1, SIGMA2, SIGMA3, MXBB and MGB Biobank), we included three extra cohorts, which we also imputed to the TOPMed panel: the METS and the Mexican Hypertriglyceridemia (MHTG) cohorts, as well as the genetically identified Latino samples from the UK Biobank (UKBB) [28]. We also analysed the Nightingale NMR-based panel of 168 metabolomic biomarkers from the UKBB. Association analyses were done with a maximum of 26,400 adult Latino individuals, depending on the trait, of whom 19,459 were diabetes-free.

**Credible sets** For each novel variant, we identified the set of variants with 99% probability of containing the causal variant. We used a Bayesian method [29], considering variants in LD with the lead variant ( $r^2 > 0.1$ ). We calculated LD using genetic data from 1996 Hispanic/Latino samples from TOPMed freeze 5b.

**Genomic annotation** We used the 99% credible sets to annotate their genomic effect using the VEP v100 [30] (GRCh38.p7) and SNP Nexus release Dec 2020 [31] applications. We used the Genotype–Tissue Expression project (GTEx) V8 [32] to assess the influence of the variants in gene-level expression, the TIGER Portal v7 [33] to evaluate the gene-level expression in pancreatic islets and the Islet Gene View (accessed 17 Dec 2022) [34] to assess the gene co-expression in human islets. We also assessed their association with a variety of phenotypes and diseases using the Common Metabolic Disease Knowledge Portal ([cmdgenkp.org](http://cmdgenkp.org), accessed 17 Dec 2022) and other resources.

**Expression of genes near novel variants** We assessed the expression levels of the genes  $\pm 500$  kb around the novel signals in human islets under different conditions pertaining

to type 1 and type 2 diabetes. Gene expression differences between groups were assessed using  $p$  values and adjusted  $p$  values (Benjamini Hochberg correction) determined by the Wald test using the DESeq2 pipeline [35]. Transcripts per million (TPM) was normalised by Salmon v1.4.0 [36].

**Polygenic scores** Polygenic scoring using single ancestry summary statistics and LD reference panels was calculated via Bayesian Regression and Continuous Shrinkage priors as implemented in PRS-CS release 4 Jun 2021 [37]. We used the UKBB LD reference panel and GWAS summary statistics from European [20], East Asian [19] and Latino populations. GWAS Latino summary statistics were calculated using a meta-analysis with five of the discovery cohorts (i.e. SIGMA1, SIGMA2, SIGMA3, MGB and GERA). Then, we used the estimated posterior SNP effect sizes for each ancestry to calculate and evaluate the performance of the polygenic scores (PSs) in a training cohort (i.e. MXBB). The best model was tested in a target cohort (i.e. the METS cohort).

Given that the ancestry-specific PSs were not highly correlated ( $r^2 < 0.3$ ), we also used PRS-CSx release 29 Jul 2021 [38], a method that improves multi-ancestry polygenic prediction by integrating GWAS summary statistics from multiple populations. We assessed the performance of the ancestry-specific vs the multi-ancestry PS.

## Results

**Overall strategy** Figure 1 summarises our overall strategy. We meta-analysed six type 2 diabetes GWAS of Latino ancestry, comprising 8150 cases and 10735 controls from hospital and population-based studies. All cohorts were imputed with TOPMed and 1000G panels and the imputation performance was evaluated. To replicate the novel loci, we analysed 13,617 type 2 diabetes cases and 20,822 controls from six independent cohorts of Latino ancestry. To gain further insight into the novel loci, we created a collection of type 2 diabetes-related phenotypes that included 26,400 Latino participants with 46 glycaemic and anthropometric traits, as well as 168 metabolomic traits. We used publicly available resources to interrogate our top signals, including functional annotation of the credible sets, and gene expression assessment of nearby genes in pancreatic islets from either type 1 or type 2 diabetes cases and controls or treated under conditions relevant for diabetes pathophysiology. We then used the generated Latino GWAS data, in combination with GWAS from other ancestries, to construct ancestry-specific and multi-ancestry type 2 diabetes PSs.

**TOPMed imputation performance** On average, imputation using the TOPMed panel resulted in 41 million (M) high-quality ( $r^2 \geq 0.8$ ) variants, being 24M rare (minor allele frequency [MAF] < 0.1%). This represents a 6.5-fold increased number of imputed rare variants compared with 1000G (Fig. 2a). The quality of imputation consistently improved when using TOPMed, particularly for low-frequency and rare variants (Fig. 2b).

We used WES data to confirm the improvement of TOPMed imputation to detect low-frequency and rare variants. The TOPMed panel allowed the identification of >80% of the WES variants with MAF  $\geq 0.1\%$  compared with 60% for the same MAF cut-off with the 1000G panel (Fig. 2d). It also improved the identification of likely pathogenic variants predicted as deleterious that usually occur at low frequency (Fig. 2e).

**Type 2 diabetes GWAS meta-analysis** To illustrate the gain in discovery when using TOPMed imputation, we tested the genetic variants for association with type 2 diabetes in six Latino cohorts. Our discovery sample comprised 18,885 Latino non-related individuals (8150 cases, 10,735 controls).

We identified 26 genome-wide significant variants ( $p < 5 \times 10^{-8}$ ) associated with type 2 diabetes at 13 loci. Twenty-five of these were previously reported type 2 diabetes-associated variants, including those consistently identified in multiple populations (e.g. variants at *KCNQ1* and *TCF7L2*) and others enriched in the Latino population (e.g. variants at *SLC16A11*) (Fig. 3a, ESM Fig. 2 and ESM Table 3).

We identified a novel locus between the *ORC5* and *LHFPL3* genes on chromosome 7. The intergenic lead variant, rs2891691, has low frequency in Latino people and is associated with a twofold increase in the odds of developing type 2 diabetes in the discovery sample (MAF 1.7%; OR 2.0 [95% CI 1.59, 2.52],  $p = 3.4 \times 10^{-9}$ ) (Fig. 3b,c). Although it was also imputed with the 1000G panel, TOPMed's higher imputation quality strengthened the association (1000G, mean  $\pm$  SD imputation  $r^2 = 0.948 \pm 0.057$ ,  $p = 2.3 \times 10^{-8}$ ; TOPMed, mean  $\pm$  SD imputation  $r^2 = 0.983 \pm 0.009$ ,  $p = 3.4 \times 10^{-9}$ ).

This variant is rare in Europeans (MAF 0.04%), yet prevalent among African (MAF 16%) and East Asian populations (MAF 7.6%). However, its association with type 2 diabetes does not replicate in either Africans ( $p = 0.149$ ) or East Asians ( $p = 0.095$ ). A fixed effects meta-analysis of the three ancestries showed no association of the variant with type 2 diabetes ( $p = 0.734$ ) but a significant heterogeneity in the allelic effects ( $p = 5 \times 10^{-8}$ ). To further investigate the source of such heterogeneity, we used MR-MEGA v1.0.5 software [39], which implements a multi-ancestry meta-regression approach to model allelic effects as a function of axes of the genetic variation. This meta-regression approach showed a significant association of rs2891691 with type 2 diabetes ( $p = 1.1 \times$

$10^{-7}$ ), as well as significant heterogeneity of the allelic effects between populations driven by ancestry ( $p = 2.9 \times 10^{-8}$ ). The residual heterogeneity accounting for other factors, such as phenotype definition or uncorrected population structure, was not significant ( $p = 0.944$ ) (ESM Fig. 3). These results show that the effects of rs2891691 on type 2 diabetes are specific to the Latino population and suggest that the lead variant we identified is in LD with the causal variant in Latino but not African or East Asian populations, a phenomenon also observed in a previous type 2 diabetes multi-ancestry meta-analysis [40]. The heterogeneity in the allelic effects across ancestries can also be explained by differences in environmental exposures.

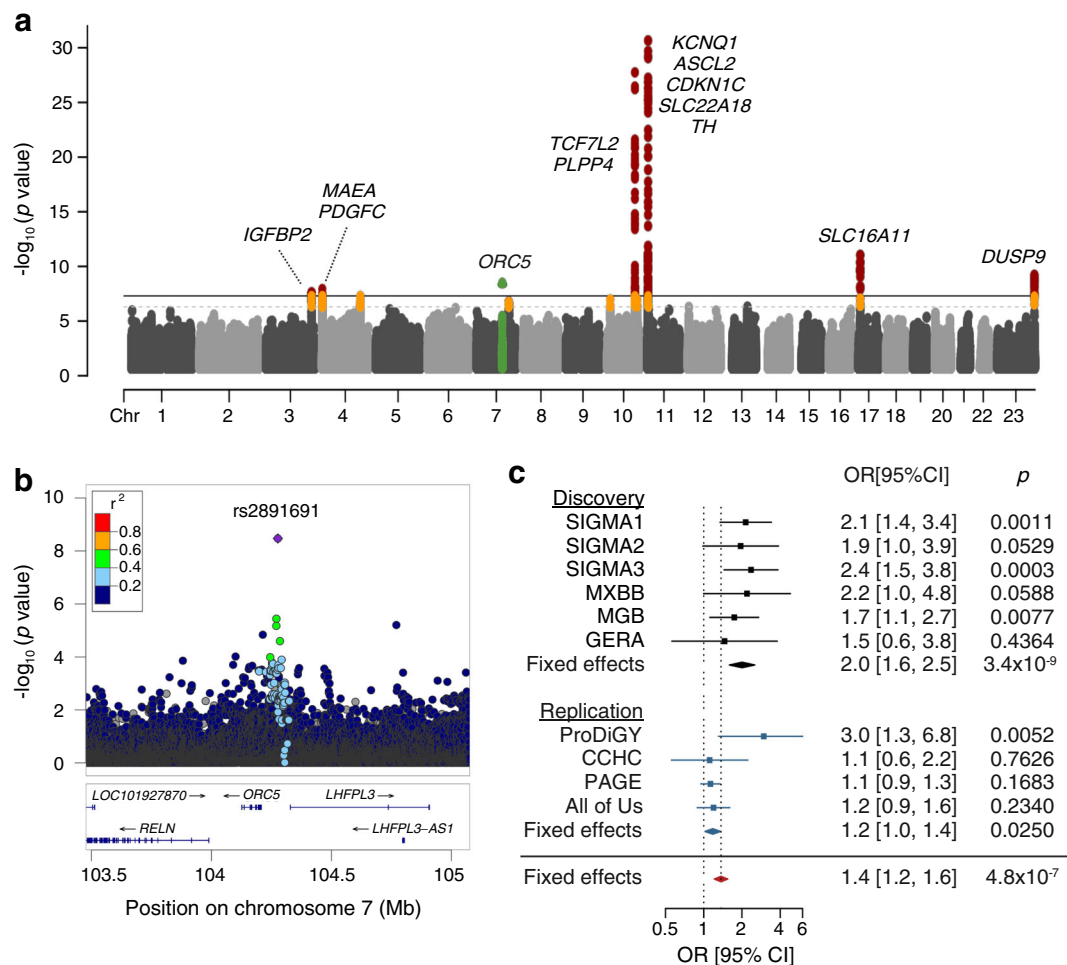
A sex-dimorphism in *RELN* gene expression has been documented, with higher *RELN* expression in women [41] and sex hormones likely mediating *RELN* expression. Because of the proximity of *RELN* to rs2891691, we evaluated the sex-specific association with type 2 diabetes and tested for heterogeneity between sex-specific allelic effects using GWAMA v2.2.2 [42]. rs2891691 showed a larger effect and was more associated with type 2 diabetes in women (Neff 10,228; OR 2.4 [95% CI 1.73, 3.22],  $p = 6.6 \times 10^{-8}$ ) compared with men (Neff 7206; OR 1.5 [95% CI 1.08, 2.19],  $p = 0.018$ ), yet the between-sex heterogeneity did not reach statistical significance ( $p = 0.076$ ) (ESM Table 4).

**Replication analysis** The replication analysis comprised 13,617 type 2 diabetes cases and 20,822 controls (ESM Table 2). The meta-analysis of the replication cohorts, where the variant was present, was nominally significant and showed a consistent direction of effect with the discovery sample (OR 1.18 [95% CI 1.02, 1.36],  $p = 0.025$ ) (Fig. 3c, ESM Table 5).

By querying our Latino collection of type 2 diabetes-related phenotypes, we found that the rs2891691 risk allele C was nominally associated with lower fasting glucose levels ( $p = 0.026$ ) (ESM Table 6). Such negative correlation might be induced by collider bias since specifically for glycaemic traits we only analysed diabetes-free individuals. Indeed, a positive association of rs2891691 risk allele with 2 h glucose adjusted for BMI has been previously reported in Latino ancestry participants ( $\beta = 3.4$  mg/dl [0.2 mmol/l],  $p = 0.006$ ) [43] and low potassium levels in East Asian ancestry participants ( $p = 8.5 \times 10^{-5}$ ) [44]. Accumulated epidemiological evidence points to a relationship between low potassium levels and decreased insulin secretion and risk of type 2 diabetes [45, 46].

The 99% credible set consisted only of the lead variant rs2891691 (ESM Table 7), yet we cannot discard other variants not called due to genotyping complexity nor those imputed to the TOPMed panel, such as a structural, variable tandem repeat or copy number variants.

To better characterise the role of the *ORC5/LHFPL3* locus, we assessed gene expression using the GTEx [32] and TIGER



**Fig. 3** Type 2 diabetes GWAS meta-analysis in the Latino population. (a) Manhattan plot of the meta-analysis association statistics, highlighting the loci with genome-wide significance (red) or sub-genome-wide significance (orange) for type 2 diabetes. (b) Regional association plot of the

novel *ORC5/LHFPL3* locus associated with type 2 diabetes risk. (c) Forest plot of the GWAS association statistics for the novel *ORC5/LHFPL3* locus in the discovery (black), the replication (blue) and overall (red) cohorts

[33] portals. *ORC5* is expressed ubiquitously, while *LHFPL3* is specifically expressed in the brain (ESM Fig. 5a, b). We then assessed the expression levels of genes  $\pm 500$  kb around the novel signal in human islets under different conditions relevant to diabetes pathophysiology. *ORC5* was downregulated after 2 h and 8 h exposure to IFN- $\alpha$ , and upregulated by exposure to brefeldin A (ESM Fig. 6a, c). Both IFN- $\alpha$  and brefeldin A are endoplasmic reticulum stress inducers that reduce the insulin content with a rise in the proinsulin/insulin ratio [47] and inhibit glucose-stimulated insulin secretion [48], respectively.

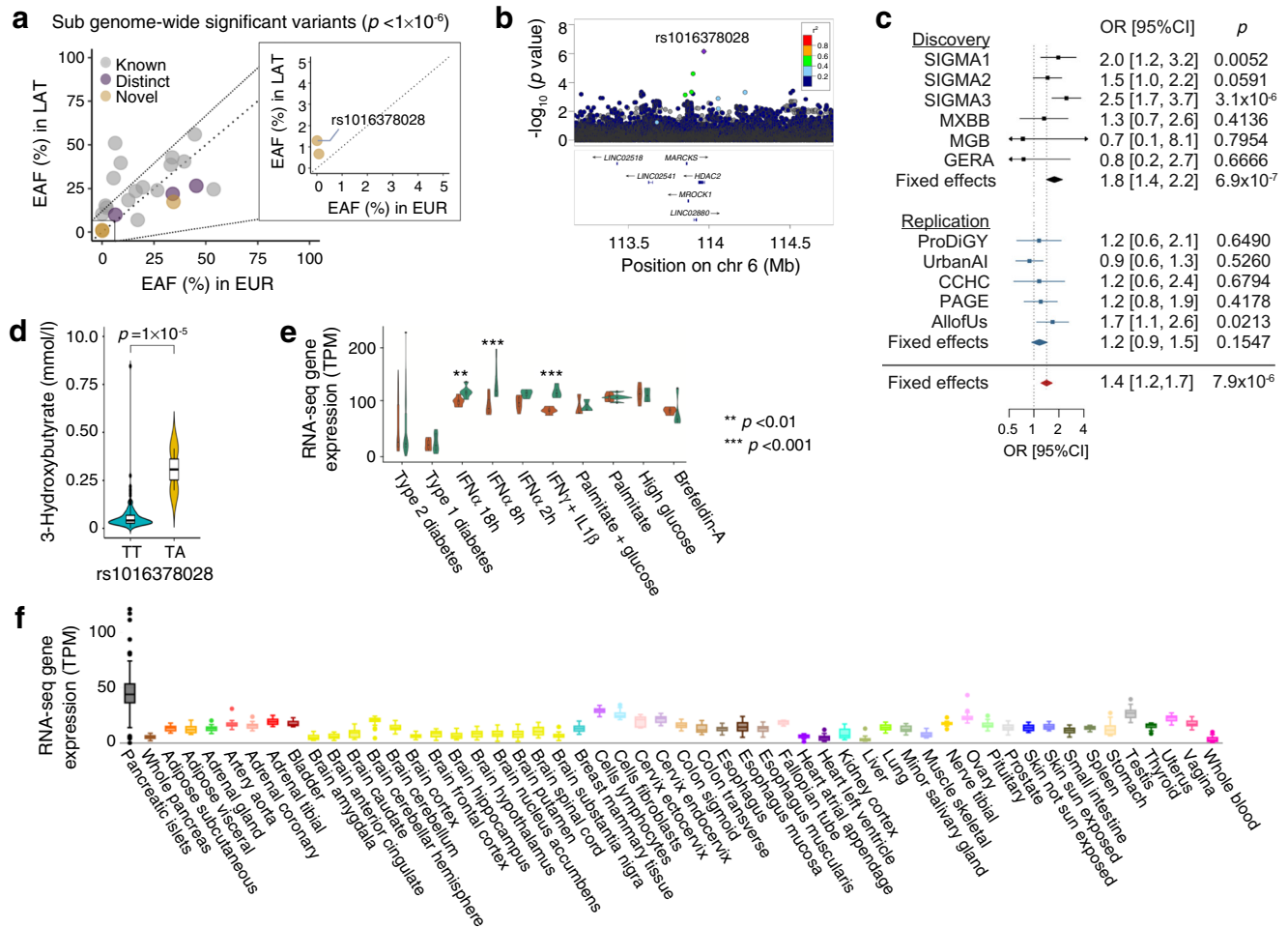
**Prioritising sub-genome-wide significant variants** We next searched for variants that were associated with type 2 diabetes at sub-genome-wide significance ( $p < 5 \times 10^{-6}$ ) but that deserved further study as they lay in previously unreported type 2 diabetes loci, were enriched or Latino-specific, and/or exclusively imputed with the TOPMed panel (Fig. 4a and

ESM Table 8). Three out of the 23 sub-genome-wide lead variants lay in or near the known type 2 diabetes loci *TACC2*, *FGFR2* and *CCND2*. We considered them as distinct variants as they retained locus-wide significance ( $p < 5 \times 10^{-5}$ ) after conditioning on the nearest known associated variant.

Three additional sub-genome-wide significant variants were located  $\pm 1$  Mb away from any reported type 2 diabetes locus (Fig. 4a and ESM Table 9). Of interest, rs1016378028 is a low-frequency variant (MAF 1.3%; OR 1.77 [95% CI 1.41, 2.21],  $p = 7.0 \times 10^{-7}$ ) that is Latino private (MAF < 0.01% in other populations) and is only imputed with the TOPMed panel. It is intronic of *HDAC2*, a gene under strong purifying selection (probability of being LoF intolerant [pLI]=1, gnomAD, [gnomAD-sg.org](http://gnomAD-sg.org), accessed 17 December 2022) and that is highly and mostly expressed in pancreatic islets (tiger.bsc.es, accessed 17 December 2022) (Fig. 4f) [33].

Although the replication results did not show statistical significance, the direction of the effect was consistent with the





**Fig. 4** Sub-genome-wide significant *HDAC2* novel type 2 diabetes loci. (a) Scatter plot of the effect allele frequencies (EAFs) from the sub-genome-wide significant variants in Latino (LAT) vs European (EUR) populations, highlighting those that are distinct from the known lead type 2 diabetes-associated variants (purple) and those that are in novel loci (yellow). (b) Regional association plot of the novel *HDAC2* locus associated with type 2 diabetes risk. (c) Forest plot of the association statistics in the discovery (black) and the replication (blue) cohorts. (d) Violin plots of serum 3-hydroxybutyrate levels in non-carriers (blue) and carriers (yellow) of the rs1016378028 variant. Whiskers range from upper and

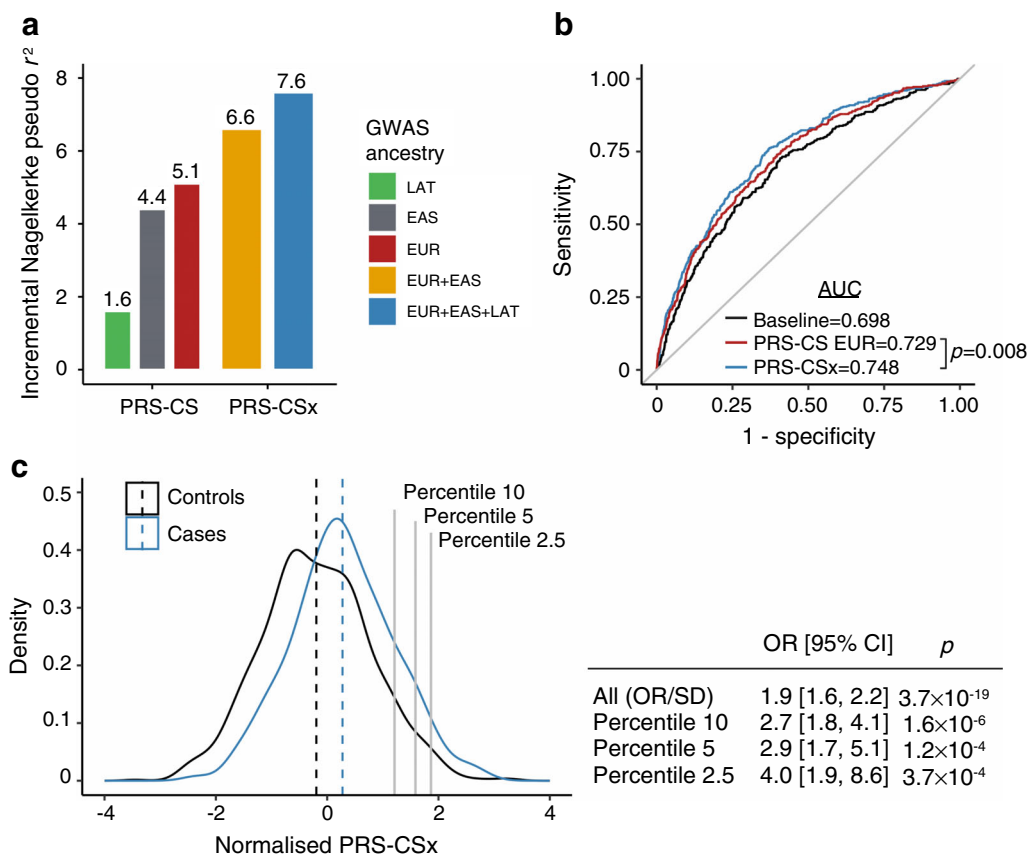
lower fences ( $1.5 \times \text{IQR}$ ); points represent outliers. (e) *HDAC2* gene expression in human islets from donors with type 1 and type 2 diabetes and control islets treated (brown) or not (green) with different cytokines or other stressor compounds.  $**p < 0.01$ ,  $***p < 0.001$  vs no treatment (adjusted  $p$  values, Benjamini–Hochberg correction). Whiskers range from upper and lower fences ( $1.5 \times \text{IQR}$ ); points represent outliers. (f) *HDAC2* gene expression in multiple tissues from GTEx and TIGER portals. Each box plot shows expression in a different tissue or cell line Whiskers range from upper and lower fences ( $1.5 \times \text{IQR}$ ); points represent outliers

discovery effect (OR 1.17 [95% CI 0.94, 1.45],  $p = 0.1547$ ) (Fig. 4b,c, ESM Table 5). The Diabetes Meta-Analysis of Trans-Ethnic association studies (DIAMANTE) European meta-analysis [20] reported a suggestive signal ~80 kb upstream of rs1016378028 (rs4945979,  $p = 4.8 \times 10^{-6}$ ). After conditioning for the rs4945979 variant, the statistical significance of our identified variant remained essentially the same (OR 1.75 [95% CI 1.4, 2.2],  $p = 4.5 \times 10^{-7}$ ).

The rs1016378028 risk allele was significantly associated with higher levels of acetone ( $p = 1.2 \times 10^{-7}$ ), 3-hydroxybutyrate ( $p = 1.01 \times 10^{-5}$ ) and acetoacetate ( $p = 3.3 \times 10^{-5}$ ) (Fig. 4d and ESM Table 10). It was also nominally associated with lower hip circumference ( $p = 0.02$ ) and higher WHR ( $p = 0.03$ ) (ESM Table 6).

*HDAC2* expression in human islets is downregulated after exposure to IFN- $\alpha$  (8 h  $\log_2$ -fold change =  $-0.38$ ,  $p = 6 \times 10^{-7}$ ; 18 h  $\log_2$ -fold change =  $-0.28$ ,  $p = 3 \times 10^{-4}$ ) or IFN- $\gamma$ +IL-1 $\beta$  ( $\log_2$ -fold change =  $-0.39$ ,  $p = 3 \times 10^{-7}$ ) (Fig. 4e). These cytokines mimic the proinflammatory milieu of type 1 diabetes, inhibit beta cell function [49, 50], induce beta cell stress and may trigger beta cell dedifferentiation in type 2 diabetes [51, 52].

**Development of PSs for the Latino population** We then developed a PS for type 2 diabetes in Latino people using our TOPMed imputed GWAS meta-analysis data. This PS explained 1.6% of the type 2 diabetes status variance



**Fig. 5** PS for the risk of type 2 diabetes in Latino population. **(a)** Variance explained by a PS using these Latino GWAS association statistics (LAT, green), the AGEN East Asian GWAS association statistics (EAS, grey), the DIAMANTE European GWAS association statistics (EUR, red), a combination of DIAMANTE European and AGEN East Asian GWAS association statistics (yellow) and a combination of DIAMANTE European, AGEN East Asian and these Latino GWAS association statistics (blue). METSB was used as the testing cohort. **(b)** Receiver operating characteristic curves for the type 2 diabetes risk prediction explained by a

model including sex, age and ten PCs of ancestry (black), a model including covariates and a PS constructed using DIAMANTE European (red) and a model including covariates and a PS constructed using a combination of DIAMANTE European, AGEN East Asian and these Latino GWAS association statistics (blue). **(c)** Distribution of a multi-ancestry PS using a combination of DIAMANTE European, AGEN East Asian and these Latino GWAS association statistics in type 2 diabetes cases (blue) and controls (black). The table shows the OR per SD attributed to the multi-ancestry PS, as well as the OR for high-risk individuals

(Fig. 5a), which is expected given the relatively small sample size of the Latino summary statistics compared with European and East Asian ancestries. The PS derived from the Diabetes Meta-Analysis of Trans-Ethnic association studies (DIAMANTE) European GWAS [20] and from Asian Genetic Epidemiology Network (AGEN) East Asian GWAS [19] explained 5.1% and 4.4% of the type 2 diabetes variance in the Latino population, respectively. The European and East Asian PSs showed a weak correlation ( $r^2 < 0.2$ ) with our Latino TOPMed-derived PS, suggesting that they could provide orthogonal information and improve the overall predictive performance. We developed a PS that incorporated GWAS data from the three ancestries using PRS-CSx [38], a method that allows for the integration of summary statistics and LD reference panels from different ancestries. The multi-ancestry PS including the three GWAS summary statistics explained 7.6% of the type 2 diabetes variance in the Latino target sample. Our Latino GWAS added 1% of the explained

variance compared with the PS using only European and East Asian GWAS, which explained 6.6% of the variance.

Each SD of the multi-ancestry PS was associated with an OR of 1.9 (95% CI 1.6, 2.2,  $p=3.7 \times 10^{-19}$ ) (Fig. 5c). People in the 2.5 percentile of the PS showed four times more risk of developing type 2 diabetes (OR 4.01 [95% CI 1.87, 8.62],  $p=3.7 \times 10^{-4}$ ) (Fig. 5c). The receiver operating characteristic AUC of the full model including the multi-ancestry PS was 0.748 (95% CI 0.72, 0.775) compared with 0.729 (95% CI 0.701, 0.758) of the PS including European GWAS only, representing a 2% improvement in the prediction accuracy ( $p=0.008$ ) (Fig. 5b).

## Discussion

The Latino population has been underrepresented in most genetic studies. Yet, recent studies of type 2 diabetes in

Latino populations have been fruitful, even with sample-size orders of magnitude smaller than those in studies of European or East Asian ancestries. The poor representation of Latino samples with genotype and phenotype data constrains nearly every step of a gene–disease association framework, including genotype imputation, a cost-effective technique to improve the resolution of a GWAS. This is more problematic for low-frequency and rare variation. Instead, next-generation sequencing technologies have typically been chosen but these are more expensive, precluding the study of large samples. This study was motivated by the recent release of the TOPMed imputation panel, which includes the largest number of Latino haplotypes compared with all available panels.

In this study, we aggregate genotype and WES data from six datasets to test the improvement in accuracy of the TOPMed imputation compared with 1000G. To illustrate how this panel can boost the discovery of complex disease variants we performed a type 2 diabetes GWAS meta-analysis using the imputed data. TOPMed imputation not only improved the statistical significance of our findings but allowed for the testing of up to 24 M rare variants, compared with 3 M properly imputed with the 1000G panel. The high quality of TOPMed imputation at low/rare frequencies is especially relevant for the study of disease-causing variation, because deleterious variants usually span such a spectrum. We show that by imputing with TOPMed, it is possible to test >90% of the variants with a  $MAF \geq 0.1\%$  predicted to be deleterious by the Combined Annotation Dependent Depletion (CADD) score; previously, it was only possible to detect these variants by relying on more expensive sequencing technologies. While ascertaining variants at frequencies <0.1% may still require whole-genome sequencing (WGS) or WES, we estimate that the power to identify associated variants may be limited unless we undertake sequencing efforts with sample sizes orders of magnitude larger than our study. For example, for  $MAF < 0.1\%$ , the effective sample size required to reach statistical power to detect associations with an effect of  $OR > 2.0$  is above 170,000 individuals (Fig. 2c). Since the cost of sequencing such a large sample size is a major constraint for the study of underrepresented populations, we propose that highly accurate imputation with dense reference panels may be a more cost-effective approach.

In this study, we identified a novel low-frequency variant associated with type 2 diabetes, rs2891691, which lies between the *ORC5* and *LHFPL3* genes and showed increased accuracy of imputation and association power when using the TOPMed panel. *ORC5* encodes the subunit 5 of the origin recognition complex implicated in the DNA replication origins, transcription silencing and heterochromatin formation [53] Lipoma HMGIC fusion partner-like 3 (LHFPL3) is a member of the tetraspanin superfamily, which functions as membrane protein organiser. The rs2891691 risk allele is

present in 1% of Latino people. Overall, in discovery and replication cohorts, carriers have 1.37-fold increased odds of developing type 2 diabetes, with a possibly higher risk in women.

We identified a second low-frequency variant, rs1016378028, associated with a 1.7-fold increased risk of type 2 diabetes, which is not imputed with the 1000G panel. This variant was prioritised from a subset of variants at a sub-genome-wide significant threshold that showed additional evidence of association. rs1016378028 is a Latino private variant (MAF: Latino, 1.3%; East Asian, 0.2%; other populations, <0.05%), and lies within *HDAC2*, a gene that is highly intolerant of protein-changing variation and is mostly expressed in pancreatic islets [33].

Histone deacetylase 2 (HDAC2) is a histone deacetylase involved in gene transcription repression. HDACs play a regulatory role in insulin signalling, beta cell function and pancreatic endocrine cell development. At low glucose levels, HDAC2 is recruited to the insulin promoter to downregulate its expression [54]. In human islets, *HDAC2* expression negatively correlates with insulin gene expression ( $r = -0.56$ , false discovery rate  $3.7 \times 10^{-16}$ ) and positively correlates with *IAPP* expression, which encodes for a satiety hormone ( $r = 0.38$ , false discovery rate  $1.8 \times 10^{-7}$ ) [34] HDAC2 also deacetylates IRS-1, uncoupling its downstream phosphorylation cascade. Both insulin expression and insulin signalling are partially restored after treatment with HDAC2 inhibitors [55, 56]. We show that cytokine treatment of pancreatic islets downregulated *HDAC2* expression.

Because there are no comprehensive genome-wide association data to guide the interpretation of variants enriched in Latino populations, we aggregated phenotypic glycaemic and cardiometabolic data from 26,400 Latino individuals to follow-up the identified variants. We found that rs1016378028 risk allele carriers have higher levels of ketone bodies, which are produced through the breakdown of fatty acids and serve as an alternative energy source to glucose. Uncoupled hepatic production of ketone bodies may be a pathological consequence of relative insulin deficiency in diabetes [57]. While the mechanism linking rs1016378028, diabetes and 3-hydroxybutyrate levels remains to be determined, our results suggest this variant as a potential genetic type 2 diabetes risk factor.

We leveraged our GWAS results and existing publicly available data to develop an improved PS for Latino ancestry. PSs developed in a particular ancestry group poorly transfer to other populations, exacerbating disparities between populations. We provide an improved PS for the Latino population, by using a combination of GWAS and LD data from East Asian, European and our Latino GWAS. This PS showed a similar performance to the previously reported in European ancestry [58] with individuals at the top 2.5 percentile showing a fourfold increased risk of type 2 diabetes. Evaluating this

PS in additional external datasets of Latino ancestry may prove useful in assessing its potential clinical utility.

Leveraging new resources to reanalyse Latino data, such as imputation with the TOPMed panel, proved to be successful in identifying additional type 2 diabetes-related loci. We acknowledge that the TOPMed panel allows the testing of an increased number of variants and additional evidence will be needed to confirm associations at the standard GWAS significance. Further efforts are needed to increase the power of discovery and to follow-up on novel findings in diverse populations. Until then, translation of identified genetic variation-to-function and application to the clinic in Latino populations will remain highly compromised compared with the resources available for European populations. In this study we gathered a high number of Latino samples with extensive biomarker and clinical characterisation; however, larger sample sizes are still needed to achieve sufficient statistical power to detect low-frequency variants. Efforts must be expanded to build shareable resources with a high representation of different ancestries, enabling ancestry-specific effects to be interpreted within the local ancestry context, which is instrumental to identify causal genes, to improve the biological mechanistic insight and to develop targeted therapies.

Overall, this study confirms the superior imputation performance of TOPMed, representing a cost-effective and unique opportunity to analyse low-frequency and rare genetic variants in Latino samples at scale. It also presents the largest type 2 diabetes GWAS meta-analysis performed in individuals of Latino ancestry imputed with the TOPMed reference panel. Despite the sample size being orders of magnitude smaller compared with studies performed in other populations, the novel discoveries presented here suggest that more novel genetic associations and new biology of type 2 diabetes will be revealed as the sample size of discovery samples, reference panels and large-scale biobanks with phenome-wide data increase in studies including non-European populations.

**Supplementary Information** The online version of this article (<https://doi.org/10.1007/s00125-023-05912-9>) contains peer-reviewed but unedited supplementary material.

**Acknowledgements** The PAGE consortium thanks the staff and participants of the PAGE studies for their important contributions. A list of PAGE senior investigators can be found at <http://www.pagestudy.org>. In addition, the All of Us Research Program would not be possible without the partnership of its participants. A full list of the Mexican Biobank contributors can be found in the [ESM](#).

**Data availability** Full summary statistics are available through the Common Metabolic Diseases Knowledge Portal (<https://t2d.hugeamp.org/downloads.html>) and through the GWAS catalog (<https://www.ebi.ac.uk/gwas/>, accession ID: GCST90255648). Polygenic scores (PS) weights for each ancestry are available via the PGS catalog (<https://www.pgscatalog.org>, publication ID: PGP000445, scores IDs: PGS003443, PGS003444 and PGS003445).

**Funding** SIGMA was partially supported by a joint USA–Mexico project funded by the Carlos Slim Health Institute. The UNAM/INCMNSZ Diabetes Study was supported by Consejo Nacional de Ciencia y Tecnología grants (138826, 128877, SALUD 2009-01-115250) and a grant from Dirección General de Asuntos del Personal Académico (UNAM, IT 214711). The Mexico City Diabetes Study was supported by the NHLBI (grant ROHL24799) and the Consejo Nacional de Ciencia y Tecnología (grants 2099, M9303, F671-M9407, 251M, 2005-CO1-14502 and SALUD 2010-2-15-1165). SIGMA was also supported by funds from the Fundación Carlos Slim (to JCF). GERA was supported by a grant (RC2 AG033067; principal investigators C. Schaefer and N. Risch) awarded to the Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH) and the UCSF Institute for Human Genetics. GERA was also supported by grants from the Robert Wood Johnson Foundation, the Wayne and Gladys Valley Foundation, the Ellison Medical Foundation, Kaiser Permanente Northern California, and the Kaiser Permanente National and Northern California Community Benefit Programs. The PAGE programme is funded by the National Human Genome Research Institute (NHGRI) with co-funding from the National Institute on Minority Health and Health Disparities (NIMHD), supported by U01HG007416 (Causal Variants Across the Life Course [CALiCo]), U01HG007417 (Icahn School of Medicine at Mount Sinai [ISMMS]), U01HG007397 (Multiethnic Cohort Study [MEC]), U01HG007376 (Women’s Health Initiative [WHI]), U01HG007419 (Coordinating Center), R01HG010297 and R01HL151152. The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

The data and materials included in this report result from collaboration between the following studies and organisations: The HCHS/SOL was carried out as a collaborative study supported by contracts from the NHLBI to the University of North Carolina (N01-HC65233), University of Miami (N01-HC65234), Albert Einstein College of Medicine (N01-HC65235), Northwestern University (N01-HC65236) and San Diego State University (N01-HC65237). The following Institutes/Centres/Offices contribute to the HCHS/SOL through a transfer of funds to the NHLBI: NIMHD; National Institute on Deafness and Other Communication Disorders; National Institute of Dental and Craniofacial Research; National Institute of Diabetes and Digestive and Kidney Diseases; National Institute of Neurological Disorders and Stroke; and NIH Institution-Office of Dietary Supplements. Samples and data of The Charles Bronfman Institute for Personalized Medicine (IPM) BioMe Biobank used in this study were provided by The Charles Bronfman Institute for Personalized Medicine at the Icahn School of Medicine at Mount Sinai (New York). Phenotype data collection was supported by The Andrea and Charles Bronfman Philanthropies. Funding support for the PAGE IPM BioMe Biobank study was provided through the NHGRI (U01 HG007417). The datasets used for the analyses described in this manuscript were obtained from dbGaP under accession phs000925. The Multiethnic Cohort study (MEC) characterisation of epidemiological architecture is funded through the NHGRI PAGE programme (U01 HG007397). The MEC study is funded through the National Cancer Institute (R37CA54281, R01CA63, P01CA33619, U01CA136792 and U01CA98758). The datasets used for the analyses described in this manuscript were obtained from dbGaP under accession phs000220. Funding support for the ‘Exonic variants and their relation to complex traits in minorities of the WHI’ study is provided through the NHGRI PAGE programme (U01HG007376). The WHI programme is funded by the NHLBI, NIH, US Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C. The authors thank the WHI investigators and staff for their dedication, and the study participants for making the programme possible. The datasets used for the analyses described in this manuscript were obtained from dbGaP under accession phs000227.

The Urban American Indians and Arizona Pima Indians cohort studies were supported by the intramural research programme of NIDDK. Parts



of this research were conducted using the UK Biobank Resource under Application Number 27892.

The All of Us Research Program is supported by the National Institutes of Health, Office of the Director: Regional Medical Centers: 1 OT2 OD026549; 1 OT2 OD026554; 1 OT2 OD026557; 1 OT2 OD026556; 1 OT2 OD026550; 1 OT2 OD 026552; 1 OT2 OD026553; 1 OT2 OD026548; 1 OT2 OD026551; 1 OT2 OD026555; IAA : AOD 16037; Federally Qualified Health Centers: HHSN 263201600085U; Data and Research Center: 5 U2C OD023196; Biobank: 1 U24 OD023121; The Participant Center: U24 OD023176; Participant Technology Systems Center: 1 U24 OD023163; Communications and Engagement: 3 OT2 OD023205; 3 OT2 OD023206; and Community Partners: 1 OT2 OD025277; 3 OT2 OD025315; 1 OT2 OD025337; 1 OT2 OD025276.

JMM is supported by ADA Innovative and Clinical Translational Award 1-19-ICTS-068, ADA grant no. 11-22-ICTSPM-16 and by NHGRI, grant FAIN no. U01HG011723. MC is supported by the Fonds National de la Recherche Scientifique (FNRS), the Walloon Region SPW-EER Win2Wal project BetaSource, and the FWO and FRS-FNRS under the Excellence of Science (EOS) programme, project Pandarome, Belgium. XY is supported by the Foundation ULB and the China Scholarship Council. DLE acknowledges the support of grants from the Welbio-FNRS (WELBIO-CR-2019C-04), Belgium. DLE, PM and MC acknowledge the support from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreements 115797 (INNODIA) and 945268 (INNODIA HARVEST), supported by the European Union's Horizon 2020 research and innovation programme. These joint undertakings receive support from the European Union's Horizon 2020 research and innovation programme and European Federation of Pharmaceutical Industries and Associations (EFPIA), JDRF, and the Leona M. and Harry B. Helmsley Charitable Trust. PM and LM acknowledge the support of European Union's Horizon 2020 research and innovation programme T2Systems under grant agreement no. 667191. HMH is supported by the NHLBI training grant T32 HL129982, ADA grant no. 1-19-PDF-045 and R01HL142825. JB acknowledges the support of grants R01DK127084, U01HG011723, R01HL142302 and R01GM133169. The Mexican Biobank (MXBB) project was supported by CONACYT (grant FONCICYT/50/2016) and the Newton Fund (grant MR/N028937/1) awarded to A.M.E. JHL is partially supported by a MGH ECOR Fund for Medical Discovery Clinical Research Fellowship Award. YL is supported by grants R56HL150186 and R01HL158884. AD is supported by NIDDK T32DK007028. JCF is supported by UM1 DK078616, K24 HL157960, UM1 DK126185, R01 HL151855, U01 HG011723 and UM1 DK105554. TT is supported by Fundación Gonzalo Río Arrente Project No. S.678. JBC is supported by NIDDK K99DK127196. AL was supported by grant 2020096 from the Doris Duke Charitable Foundation (<https://www.ddcf.org>).

**Authors' relationships and activities** As of April 2022, PD is an employee and stockholder at Regeneron Pharmaceuticals. DMD and PM are members of the editorial board of *Diabetologia* journal and thus, were excluded from the peer review process of this manuscript. All other authors declare that there are no relationships or activities that might bias, or be perceived to bias, their work.

**Contribution statement** AH-C and JMM conceived and designed the analyses and wrote the manuscript. AH-C, PS, RM and AJD led the main analyses. CAA-S and TT procured METS and MHTG data. AM-E, LG-G and TT procured MXBB data. WZ, LP, JBMc, SPF-H and JB procured CCHC data and analysis. LC, SS, JT and JF procured T2D-Genes data and analysis. RG-K, LL, RS, MK and BB procured TODAY data and analysis. DMD, JDi and SM procured SEARCH data and analysis. LSt, RJFL, BFD, CK, LMR, CH, QS and HMH procured PAGE data and analysis. RH procured Pima Indian data and analysis. RM, LSz, JHL and AMan procured All of Us data and analysis. XY, DLE, PM, LM

and MC procured the human islets functional characterisation. All the authors contributed to the interpretation of results, critically revised the article, and gave final approval for the version to be published. JCF, AL, and JMM supervised the study. JMM is responsible for the integrity of the work as a whole.

**Open Access** This article 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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References


- Williams Amy AL, Jacobs Suzanne SBR, Moreno-Macías H et al (2014) Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. *Nature* 506(7486):97–101. <https://doi.org/10.1038/nature12828>
- The SIGMA Type 2 Diabetes Consortium, Estrada K, Aukrust I et al (2014) Association of a low-frequency variant in HNF1A with type 2 diabetes in a Latino population. *JAMA* 311(22):2305–2314. <https://doi.org/10.1001/jama.2014.6511>
- Mercader JM, Liao RG, Bell AD et al (2017) A loss-of-function splice acceptor variant in IGF2 is protective for type 2 diabetes. *Diabetes* 66(11):2903–2914. <https://doi.org/10.2337/db17-0187>
- Rusu V, Rusu V, Hoch E et al (2017) Type 2 diabetes variants disrupt function of SLC16A11 through two distinct mechanisms. *Cell* 170(1):199–212.e20. <https://doi.org/10.1016/j.cell.2017.06.011>
- Mercader JM, Florez JC (2017) The genetic basis of type 2 diabetes in Hispanics and Latin Americans: challenges and opportunities. *Front Public Health* 5:239. <https://doi.org/10.3389/fpubh.2017.00329>
- Das S, Abecasis GR, Browning BL (2018) Genotype imputation from large reference panels. *Annu Rev Genomics Hum Genet* 19: 73–96. <https://doi.org/10.1146/annurev-genom-083117-021602>
- Auton A, Abecasis GR, Altshuler DM et al (2015) A global reference for human genetic variation. *Nature* 526(7571):68–74. <https://doi.org/10.1038/NATURE15393>
- Kowalski MH, Qian H, Hou Z et al (2019) Use of >100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed African and Hispanic/Latino populations. *PLoS Genet* 15(12):e1008500. <https://doi.org/10.1371/journal.pgen.1008500>
- Taliun D, Harris DN, Kessler MD et al (2021) Sequencing of 53, 831 diverse genomes from the NHLBI TOPMed Program. *Nature* 590(7845):290–299. <https://doi.org/10.1038/S41586-021-03205-Y>
- Sepúlveda J, Tapia-Conyer R, Velásquez O et al (2007) Diseño y metodología de la Encuesta Nacional de Salud 2000. *Salud Publica Mex* 49(Suppl 3):427–432 [article in Spanish]
- Karlson EW, Boutin NT, Hoffnagle AG, Allen NL (2016) Building the Partners Healthcare Biobank at Partners Personalized Medicine: informed consent, return of research results, recruitment lessons and

- operational considerations. *J Pers Med* 6(1):1–11. <https://doi.org/10.3390/jpm6010002>
12. Banda Y, Kvale MN, Hoffmann TJ et al (2015) Characterizing race/ethnicity and genetic ancestry for 100,000 subjects in the Genetic Epidemiology Research On Adult Health And Aging (GERA) cohort. *Genetics* 200(4):1285–1295. <https://doi.org/10.1534/genetics.115.178616>
  13. Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19(9):1655–1664. <https://doi.org/10.1101/gr.094052.109>
  14. Delaneau O, Marchini J, Zagury JF (2012) A linear complexity phasing method for thousands of genomes. *Nat Methods* 9(2):179–181. <https://doi.org/10.1038/nmeth.1785>
  15. Sudmant PH, Rausch T, Gardner EJ et al (2015) An integrated map of structural variation in 2,504 human genomes. *Nature* 526(7571):75–81. <https://doi.org/10.1038/nature15394>
  16. Cingolani P, Platts A, Wang LL et al (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 6(2):80–92. <https://doi.org/10.4161/FLY.19695>
  17. Marchini J, Howie B, Myers S, McVean G, Donnelly P (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 39(7):906–913. <https://doi.org/10.1038/ng2088>
  18. Willer CJ, Li Y, Abecasis GR (2010) METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26(17):2190–2191. <https://doi.org/10.1093/bioinformatics/btq340>
  19. Spracklen CN, Horikoshi M, Kim YJ et al (2020) Identification of type 2 diabetes loci in 433,540 East Asian individuals. *Nature* 582(7811):240–245. <https://doi.org/10.1038/s41586-020-2263-3>
  20. Mahajan A, Taliun D, Thumer M 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(11):1505–1513. <https://doi.org/10.1038/s41588-018-0241-6>
  21. Fisher-Hoch SP, Rentfro AR, Salinas JJ et al (2010) Socioeconomic status and prevalence of obesity and diabetes in a Mexican American community, Cameron County, Texas, 2004–2007. *Prev Chronic Dis* 7(3):A53. <https://doi.org/10.13016/vtrw-onkt>
  22. Nair AK, Sutherland JR, Traurig M et al (2018) Functional and association analysis of an Amerindian-derived population-specific p.(Thr280Met) variant in RBPJL, a component of the PTF1 complex. *Eur J Hum Genet* 26:238–246. <https://doi.org/10.1038/s41431-017-0062-6>
  23. Wojcik GL, Graff M, Nishimura KK et al (2019) Genetic analyses of diverse populations improves discovery for complex traits. *Nature* 570(7762):514–518. <https://doi.org/10.1038/S41586-019-1310-4>
  24. All of Us Research Program Investigators, Denny J, Rutter JL et al (2019) The “All of Us” research program. *N Engl J Med* 381(7):668–676. [https://doi.org/10.1056/NEJMSR1809937/SUPPL\\_FILE/NEJMSR1809937\\_APPENDIX.PDF](https://doi.org/10.1056/NEJMSR1809937/SUPPL_FILE/NEJMSR1809937_APPENDIX.PDF)
  25. Haymond M, Anderson B, Barrera P et al (2007) Treatment options for type 2 diabetes in adolescents and youth: a study of the comparative efficacy of metformin alone or in combination with rosiglitazone or lifestyle intervention in adolescents with type 2 diabetes. *Pediatr Diabetes* 8(2):74–87. <https://doi.org/10.1111/J.1399-5448.2007.00237.X>
  26. SEARCH Study Group (2004) SEARCH for Diabetes in Youth: a multicenter study of the prevalence, incidence and classification of diabetes mellitus in youth. *Control Clin Trials* 25(5):458–471. <https://doi.org/10.1016/J.CCT.2004.08.002>
  27. Arellano-Campos O, Gómez-Velasco DV, Bello-Chavolla OY et al (2019) Development and validation of a predictive model for incident type 2 diabetes in middle-aged Mexican adults: The metabolic syndrome cohort. *BMC Endocr Disord* 19(1):41. <https://doi.org/10.1186/s12902-019-0361-8>
  28. Ahola-Olli AV, Mustelin L, Kalimeri M et al (2019) Circulating metabolites and the risk of type 2 diabetes: a prospective study of 11,896 young adults from four Finnish cohorts. *Diabetologia* 62(12):2298–2309. <https://doi.org/10.1007/s00125-019-05001-w>
  29. The Wellcome Trust Case Control Consortium, Maller JB, McVean G et al (2012) Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat Genet* 44(12):1294. <https://doi.org/10.1038/NG.2435>
  30. McLaren W, Gil L, Hunt SE et al (2016) The ensembl variant effect predictor. *Genome Biol* 17(1):1–14. <https://doi.org/10.1186/S13059-016-0974-4>
  31. Oscanoa J, Sivapalan L, Gadaleta E, Dayem Ullah AZ, Lemoine NR, Chelala C (2020) SNPnexus: a web server for functional annotation of human genome sequence variation (2020 update). *Nucleic Acids Res* 48(W1):W185–W192. <https://doi.org/10.1093/NAR/GKAA420>
  32. GTEx Consortium (2013) The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 45(6):580–585. <https://doi.org/10.1038/NG.2653>
  33. Alonso L, Piron A, Morán I et al (2021) TIGER: The gene expression regulatory variation landscape of human pancreatic islets. *Cell Rep* 37(2):109807. <https://doi.org/10.1016/J.CELREP.2021.109807>
  34. Asplund O, Storm P, Chandra V et al (2022) Islet gene view—a tool to facilitate islet research. *Life Sci Alliance* 5(12):1–17. <https://doi.org/10.26508/lsa.202201376>
  35. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(12):1–21. <https://doi.org/10.1186/S13059-014-0550-8>
  36. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* 14(4):417–419. <https://doi.org/10.1038/nmeth.4197>
  37. Ge T, Chen CY, Ni Y, Feng YCA, Smoller JW (2019) Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun* 10(1):1776. <https://doi.org/10.1038/s41467-019-09718-5>
  38. Ruan Y, Lin Y-F, Feng Y-CA et al (2022) Improving polygenic prediction in ancestrally diverse populations. *Nat Genet* 54(5):573–580. <https://doi.org/10.1038/S41588-022-01054-7>
  39. Mägi R, Horikoshi M, Sofer T et al (2017) Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum Mol Genet* 26(18):3639–3650. <https://doi.org/10.1093/hmg/ddx280>
  40. Kanai M, Ullirsch JC, Karjalainen J et al (2021) Insights from complex trait fine-mapping across diverse populations. *medRxiv* 2021.09.03.21262975. <https://doi.org/10.1101/2021.09.03.21262975>
  41. Eastwood SL, Harrison PJ (2003) Interstitial white matter neurons express less reelin and are abnormally distributed in schizophrenia: towards an integration of molecular and morphologic aspects of the neurodevelopmental hypothesis. *Mol Psychiatry* 8(9):821–831. <https://doi.org/10.1038/SJ.MP.4001371>
  42. Magi R, Lindgren CM, Morris AP (2010) Meta-analysis of sex-specific genome-wide association studies. *Genet Epidemiol* 34(8):846. <https://doi.org/10.1002/GEPI.20540>
  43. Chen J, Spracklen CN, Marenne G et al (2021) The trans-ancestral genomic architecture of glycemic traits. *Nat Genet* 53(6):840–860. <https://doi.org/10.1038/s41588-021-00852-9>
  44. Kanai M, Akiyama M, Takahashi A et al (2018) Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat Genet* 50(3):390–400. <https://doi.org/10.1038/s41588-018-0047-6>

45. Ekmekcioglu C, Elmadfa I, Meyer AL, Moeslinger T (2016) The role of dietary potassium in hypertension and diabetes. *J Physiol Biochem* 72(1):93–106. <https://doi.org/10.1007/s13105-015-0449-1>
46. Heianza Y, Hara S, Arase Y et al (2011) Low serum potassium levels and risk of type 2 diabetes: the Toranomon Hospital Health Management Center Study 1 (TOPICS 1). *Diabetologia* 54(4):762–766. <https://doi.org/10.1007/s00125-010-2029-9>
47. Lombardi A, Tomer Y (2017) Interferon alpha impairs insulin production in human beta cells via endoplasmic reticulum stress. *J Autoimmun* 80:48–55. <https://doi.org/10.1016/J.JAUT.2017.02.002>
48. Bone R, Oyebamiji O, Talware S et al (2020) A computational approach for defining a signature of  $\beta$ -cell Golgi stress in diabetes. *Diabetes* 69(11):2364–2376. <https://doi.org/10.2337/DB20-0636>
49. Eizirik DL, Pasquali L, Cnop M (2020) Pancreatic  $\beta$ -cells in type 1 and type 2 diabetes mellitus: different pathways to failure. *Nat Rev Endocrinol* 16(7):349–362. <https://doi.org/10.1038/S41574-020-0355-7>
50. Marroqui L, dos Santos RS, Op de Beeck A et al (2017) Interferon- $\alpha$  mediates human beta cell HLA class I overexpression, endoplasmic reticulum stress and apoptosis, three hallmarks of early human type 1 diabetes. *Diabetologia* 60(4):656–667. <https://doi.org/10.1007/S00125-016-4201-3>
51. Oshima M, Knoch KP, Diedisheim M et al (2018) Virus-like infection induces human  $\beta$  cell dedifferentiation. *JCI Insight* 3(3):e97732. <https://doi.org/10.1172/JCI.INSIGHT.97732>
52. Sun J, Ni Q, Xie J et al (2019)  $\beta$ -Cell dedifferentiation in patients with T2D with adequate glucose control and nondiabetic chronic pancreatitis. *J Clin Endocrinol Metab* 104(1):83–94. <https://doi.org/10.1210/JC.2018-00968>
53. Shen Z (2013) The origin recognition complex in human diseases. *Biosci Rep* 33(3):475–483. <https://doi.org/10.1042/BSR20130036>
54. Mosley AL, Özcan S (2004) The pancreatic duodenal homeobox-1 protein (PDX-1) interacts with histone deacetylases HDAC-1 and HDAC-2 on low levels of glucose. *J Biol Chem* 279(52):54241–54247. <https://doi.org/10.1074/jbc.M410379200>
55. Christensen DP, Dahllöf M, Lundh M et al (2011) Histone deacetylase (HDAC) inhibition as a novel treatment for diabetes mellitus. *Mol Med* 17(5–6):378. <https://doi.org/10.2119/MOLMED.2011.00021>
56. Ye J (2013) Improving insulin sensitivity with HDAC inhibitor. *Diabetes* 62(3):685. <https://doi.org/10.2337/DB12-1354>
57. Longo VD, Mattson MP (2014) Fasting: molecular mechanisms and clinical applications. *Cell Metab* 19:181–192. <https://doi.org/10.1016/j.cmet.2013.12.008>
58. Khera AV, Chaffin M, Aragam KG et al (2018) Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* 50(9):1219–1224. <https://doi.org/10.1038/s41588-018-0183-z>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Affiliations

Alicia Huerta-Chagoya<sup>1,2,3,4</sup> · Philip Schroeder<sup>1,2,5</sup> · Ravi Mandla<sup>1,2,5</sup> · Aaron J. Deutsch<sup>1,2,5,6</sup> · Wanying Zhu<sup>7</sup> · Lauren Petty<sup>7</sup> · Xiaoyan Yi<sup>8</sup> · Joanne B. Cole<sup>1,2,9,10,11</sup> · Miriam S. Udler<sup>2</sup> · Peter Dombos<sup>1</sup> · Bianca Porneala<sup>12</sup> · Daniel DiCorpo<sup>13</sup> · Ching-Ti Liu<sup>13</sup> · Josephine H. Li<sup>1,2,5,6</sup> · Lukasz Szczerbiński<sup>1,2,5,14,15</sup> · Varinderpal Kaur<sup>5</sup> · Joohyun Kim<sup>16</sup> · Yingchang Lu<sup>16</sup> · Alicia Martin<sup>1,17</sup> · Decio L. Eizirik<sup>8,18</sup> · Piero Marchetti<sup>19</sup> · Lorella Marselli<sup>19</sup> · Ling Chen<sup>2</sup> · Shylaja Srinivasan<sup>20</sup> · Jennifer Todd<sup>21</sup> · Jason Flannick<sup>1,22</sup> · Rose Gubitosi-Klug<sup>23</sup> · Lynne Levitsky<sup>24</sup> · Rachana Shah<sup>25</sup> · Megan Kelsey<sup>26</sup> · Brian Burke<sup>27</sup> · Dana M. Dabelea<sup>28</sup> · Jasmin Divers<sup>29</sup> · Santica Marcovina<sup>30</sup> · Lauren Stalbow<sup>31</sup> · Ruth J. F. Loos<sup>31,32</sup> · Burcu F. Darst<sup>33</sup> · Charles Kooperberg<sup>33</sup> · Laura M. Raffield<sup>34</sup> · Christopher Haiman<sup>35,36</sup> · Quan Sun<sup>37</sup> · Joseph B. McCormick<sup>38</sup> · Susan P. Fisher-Hoch<sup>38</sup> · Maria L. Ordoñez<sup>39</sup> · James Meigs<sup>1,6,12</sup> · Leslie J. Baier<sup>40</sup> · Clicerio González-Villalpando<sup>41</sup> · Maria Elena González-Villalpando<sup>41</sup> · Lorena Orozco<sup>42</sup> · Lourdes García-García<sup>43</sup> · Andrés Moreno-Estrada<sup>44</sup> · Mexican Biobank · Carlos A. Aguilar-Salinas<sup>45</sup> · Teresa Tusie<sup>3,4</sup> · Josée Dupuis<sup>13</sup> · Maggie C. Y. Ng<sup>16</sup> · Alisa Manning<sup>46</sup> · Heather M. Highland<sup>47</sup> · Miriam Cnop<sup>8,48</sup> · Robert Hanson<sup>49</sup> · Jennifer Below<sup>7</sup> · Jose C. Florez<sup>1,2,5,6,9,50</sup> · Aaron Leong<sup>1,2,5,6,12,50</sup> · Josep M. Mercader<sup>1,2,5,9</sup> 

<sup>1</sup> Programs in Metabolism and Medical & Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA

<sup>2</sup> Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA

<sup>3</sup> Departamento de Medicina Genómica y Toxicología Ambiental, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Mexico City, Mexico

<sup>4</sup> Unidad de Biología Molecular y Medicina Genómica, Instituto Nacional de Ciencias Médicas y Nutrición, Mexico City, Mexico

<sup>5</sup> Diabetes Unit, Massachusetts General Hospital, Boston, MA, USA

<sup>6</sup> Department of Medicine, Massachusetts General Hospital, Boston, MA, USA

<sup>7</sup> Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA

<sup>8</sup> ULB Center for Diabetes Research, Medical Faculty, Université Libre de Bruxelles, Brussels, Belgium

<sup>9</sup> Department of Medicine, Harvard Medical School, Boston, MA, USA

<sup>10</sup> Division of Endocrinology, Boston Children's Hospital, Boston, MA, USA

- <sup>11</sup> Department of Biomedical Informatics, University of Colorado School of Medicine, Aurora, CO, USA
- <sup>12</sup> Division of General Internal Medicine, Massachusetts General Hospital, Boston, MA, USA
- <sup>13</sup> Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA
- <sup>14</sup> Department of Endocrinology, Diabetology and Internal Medicine, Medical University of Bialystok, Bialystok, Poland
- <sup>15</sup> Clinical Research Centre, Medical University of Bialystok, Bialystok, Poland
- <sup>16</sup> Vanderbilt Genetics Institute, Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
- <sup>17</sup> Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA
- <sup>18</sup> WELBIO, Université Libre de Bruxelles, Brussels, Belgium
- <sup>19</sup> Department of Clinical and Experimental Medicine, and AOUP Cisanello University Hospital, University of Pisa, Pisa, Italy
- <sup>20</sup> Department of Pediatrics, University of California San Francisco, San Francisco, CA, USA
- <sup>21</sup> Department of Pediatrics, University of Vermont, Burlington, VT, USA
- <sup>22</sup> Department of Pediatrics, Boston Children's Hospital, Boston, MA, USA
- <sup>23</sup> Pediatric Endocrinology, Diabetes, and Metabolism, Case Western Reserve University and Rainbow Babies and Children's Hospital, Cleveland, OH, USA
- <sup>24</sup> Department of Pediatrics, Division of Pediatric Endocrinology and Pediatric Diabetes Center, Massachusetts General Hospital, Boston, MA, USA
- <sup>25</sup> Pediatric Endocrinology and Diabetes, Children's Hospital of Philadelphia, Philadelphia, PA, USA
- <sup>26</sup> Pediatric Endocrinology, University of Colorado School of Medicine, Aurora, CO, USA
- <sup>27</sup> Biostatistics Center, The George Washington University, Rockville, MD, USA
- <sup>28</sup> Department of Epidemiology, University of Colorado School of Medicine, Aurora, CO, USA
- <sup>29</sup> NYU Langone Health, New York, NY, USA
- <sup>30</sup> Medpace Reference Laboratories, Cincinnati, OH, USA
- <sup>31</sup> The Charles Bronfman Institute of Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA
- <sup>32</sup> Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Science, University of Copenhagen, Copenhagen, Denmark
- <sup>33</sup> Division of Public Health Science, Fred Hutchinson Cancer Center, Seattle, WA, USA
- <sup>34</sup> Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
- <sup>35</sup> Department of Population and Public Health Sciences, University of Southern California, Los Angeles, CA, USA
- <sup>36</sup> Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA, USA
- <sup>37</sup> Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
- <sup>38</sup> School of Public Health, The University of Texas Health Science Center at Houston, Brownsville, TX, USA
- <sup>39</sup> Unidad de Biología Molecular y Medicina Genómica, Instituto Nacional de Ciencias Médicas y Nutrición, Mexico City, Mexico
- <sup>40</sup> Phoenix Epidemiology and Clinical Research Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Phoenix, AZ, USA
- <sup>41</sup> Centro de Estudios en Diabetes, Unidad de Investigación en Diabetes y Riesgo Cardiovascular, Centro de Investigación en Salud Poblacional, Instituto Nacional de Salud Pública, Mexico City, Mexico
- <sup>42</sup> Laboratorio Immunogenómica y Enfermedades Metabólicas, Instituto Nacional de Medicina Genómica, Mexico City, Mexico
- <sup>43</sup> Instituto Nacional de Salud Pública, Cuernavaca, Mexico
- <sup>44</sup> Laboratorio Nacional de Genómica para la Biodiversidad (LANGEBIO), Unidad de Genómica Avanzada (UGA), CINVESTAV, Irapuato, Mexico
- <sup>45</sup> Unidad de Investigación de Enfermedades Metabólicas y Dirección de Nutrición, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico
- <sup>46</sup> Clinical and Translational Epidemiology Unit, Massachusetts General Hospital, Boston, MA, USA
- <sup>47</sup> Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
- <sup>48</sup> Division of Endocrinology, Erasmus Hospital, Université Libre de Bruxelles, Brussels, Belgium
- <sup>49</sup> Diabetes Epidemiology and Clinical Research Section, National Institute of Diabetes and Digestive and Kidney Diseases, Phoenix, AZ, USA
- <sup>50</sup> Endocrine Division, Massachusetts General Hospital, Boston, MA, USA