

A loss-of-function splice acceptor variant in IGF2 is protective for type 2 diabetes

Journal:	*Diabetes*
Manuscript ID	DB17-0187.R1
Manuscript Type:	Original Article: Genetics/Genomes/Proteomics/Metabolomics
Date Submitted by the Author:	11-Jul-2017
Complete List of Authors:	Mercader, Josep M; Broad Institute, Program in Medical and Population Genetics Liao, Rachel G; Broad Institute, Program in Medical and Population Genetics Davis, Avery; Broad Institute, Program in Medical and Population Genetics Dymek, Zachary; Broad Institute, Program in Medical and Population Genetics Estrada, Karol; Broad Institute, Program in Medical and Population Genetics Tukiainen, Taru; Broad Institute, Program in Medical and Population Genetics Huerta-Chagoya, Alicia; UNAM/ Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán. Mexico City, México, Unidad de Biología Molecular y Medicina Genómica, I.d.I.B. Moreno-Macías, Hortensia; UNAM/ Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán. Mexico City, México, Unidad de Biología Molecular y Medicina Genómica, I.d.I.B. Jablonski, Kathleen A; George Washington University, The Biostatistics Center Hanson, Robert; DAES/NIDDK/NIH, Building 1 Walford, Geoffrey; Massachesetts General Hospital, Medicine Moran, Ignasi; Imperial College London, Department of Medicine Chen, Ling; Massachesetts General Hospital, Medicine Agarwala, Vineeta; Broad Institute, Program in Medical and Population Genetics Ordoñez-Sánchez, Maria Luisa; UNAM/ Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán. Mexico City, México, Unidad de Biología Molecular y Medicina Genómica, I.d.I.B. Rodríguez-Guillen, Rosario ; UNAM/ Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán. Mexico City, México, Unidad de Biología Molecular y Medicina Genómica, I.d.I.B. Rodríguez-Torres, Maribel; UNAM/ Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán. Mexico City, México, Unidad de Biología Molecular y Medicina Genómica, I.d.I.B. Segura-Kato, Yayoi; UNAM/ Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán. Mexico City, México, Unidad de Biología Molecular y Medicina Genómica, I.d.I.B. García-Ortiz, Humberto; Instituto Nacional de Medicina Genomica

Centeno-Cruz, Federico; Instituto Nacional de Medicina Genomica
Barajas-Olmos, Francisco; Instituto Nacional de Medicina Genomica
Caulkins, Lizz; Broad Institute, Program in Medical and Population Genetics
Puppala, Sobha; Texas Biomedical Research Institute
Fontanillas, Pierre; Broad Institute, Program in Medical and Population Genetics
Williams, Amy; Cornell University
Bonàs, Sílvia; Centro Nacional de Supercomputacion
Hartl, Chris; Broad Institute, Program in Medical and Population Genetics
Ripke, Stephan; Broad Institute, Program in Medical and Population Genetics
Tooley, Katherine; Broad Institute, Program in Medical and Population Genetics
Lane, Jacqueline; Broad Institute, Program in Medical and Population Genetics
Zerrweck, Carlos; Hospital General Tláhuac
Martínez, Angélica; Instituto Nacional de Medicina Genomica
Córdova, Emílio; Instituto Nacional de Medicina Genomica
Mendoza-Caamal, Elvia; Instituto Nacional de Medicina Genomica
Contreras-Cubas, Cecilia; Instituto Nacional de Medicina Genomica
González-Villalpando, Maria-Elena; Instituto Nacional de Salud Publica
Cruz-Bautista, Ivette; Instituto Nacional de Ciencias Medicas y Nutricion Salvador Zubiran
Muñoz-Hernández, Liliana ; Instituto Nacional de Ciencias Medicas y Nutricion Salvador Zubiran
Gómez-Velasco, Donaji; Instituto Nacional de Ciencias Medicas y Nutricion Salvador Zubiran
Alvidre, Ulises; Instituto Nacional de Ciencias Medicas y Nutricion Salvador Zubiran
Wilkens, Lynne R; University of Hawaii Cancer Center
Le Marchand, Loic; University of Hawaii Cancer Center
Arellano-Campos, Olimpia; University of Hawaii Cancer Center
Harden, Maegan; Broad Institute, Program in Medical and Population Genetics
Gabriel, Stacey; Broad Institute, Program in Medical and Population Genetics
Cortes, Marisa; Broad Institute, Program in Medical and Population Genetics
Revilla-Monsalve, Cristina; Instituto Mexicano del Seguro Social
Islas-Andrade, Sergio; Instituto Mexicano del Seguro Social
Soberon, Xavier; Instituto Nacional de Medicina Genomica
Curran, Joanne E; University of Texas Rio Grande Valley
Jenkinson, Christopher P; University of Texas Rio Grande Valley - Edinburg Campus
DeFronzo, Ralph; Univ. of Texas Health Science Center, Dept. of Med., Diabetes Division;
Lehman, Donna; UT Health Science Center, Medicine/Clinical Epidemiology
Hanis, Craig; Univ. of Texas Health Science, Human Genetics Center
Bell, Graeme I; University of Chicago
Boehnke, Michael; University of Michigan, School of Public Health
Blangero, John; University of Texas Rio Grande Valley
Saxena, Richa; Broad Institute, Program in Medical and Population Genetics
MacArthur, Daniel; Broad Institute, Program in Medical and Population Genetics
Ferrer, Jorge; Imperial College London, Department of Medicine
McCarroll, Steven; Broad Institute
Torrents, David; Centro Nacional de Supercomputacion
Knowler, William C; National Institute of Diabetes and Digestive and Kidney Diseases
Baier, Leslie; National Institutes of Health, NIDDK;

	Burtt, Noel; Broad Institute, Program in Medical and Population Genetics González, Clicerio; Instituto Nacional de Salud Publica Haiman, Christopher; University of Hawaii Cancer Center Aguilar-Salinas, Carlos; Instituto Nacional de la Nutricion, Endocrinology and Metabolism Tusié-Luna, María Teresa; Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán / Universidad Nacional Autónoma de México, Unidad de Biología Molecular y Medicina Genómica Flannick, Jason; Broad Institute, Program in Medical and Population Genetics Jacobs, Suzanne B. R.; Broad Institute, Program in Medical and Population Genetics Orozco, Lorena; Instituto Nacional de Medicina Genomica Altshuler, David; Massachesetts General Hospital, Medicine Florez, Jose; Massachusetts General Hospital / Broad Institute of Harvard and MIT, Center for Human Genetic Research and Diabetes Unit;

SCHOLARONE™
Manuscripts

A loss-of-function splice acceptor variant in *IGF2* is protective for type 2 diabetes

Running Title: *IGF2* loss-of-function type 2 diabetes protective variant.

The SIGMA T2D Genetics Consortium

Josep M Mercader[1,2,3], Rachel G. Liao[1]*, Avery Davis[4,5,6]*, Zachary Dymek[1]*, Karol Estrada[1,7,8], Taru Tukiainen[4,6,7], Alicia Huerta-Chagoya[9], Hortensia Moreno-Macías[9,10], Kathleen A. Jablonski[11], Robert L. Hanson[12], Geoffrey A. Walford[1,2,8], Ignasi Moran[13], Ling Chen[1,2], Vineeta Agarwala[6], María Luisa Ordoñez-Sánchez[9], Rosario Rodríguez-Guillen[9], Maribel Rodríguez-Torres[9], Yayoi Segura-Kato[9], Humberto García-Ortiz[14], Federico Centeno-Cruz[14], Francisco Barajas-Olmos[14], Lizz Caulkins[1], Sobha Puppala[15], Pierre Fontanillas[6], Amy Williams[16], Sílvia Bonàs-Guarch[3], Chris Hartl[6], Stephan Ripke[5,7,17], Diabetes Prevention Program Research Group[¢], Katherine Tooley[4,5,6], Jacqueline Lane[6,18,19], Carlos Zerrweck[20], Angélica Martínez-Hernández[14], Emilio J. Córdova[14], Elvia Mendoza-Caamal[14], Cecilia Contreras-Cubas[14], María E. González-Villalpando[21], Ivette Cruz-Bautista[22], Liliana Muñoz-Hernández[22], Donaji Gómez-Velasco[22], Ulises Alvirde[22], Brian E. Henderson[23], Lynne R. Wilkens[24], Loic Le Marchand[24], Olimpia Arellano-Campos[22], Laura Riba[22], Maegan Harden[25], Broad Genomics Platform[25], Stacey Gabriel[25], T2D-GENES Consortium[¢], Hanna E. Abboud[26], Maria L. Cortes[27], Cristina Revilla-Monsalve[28], Sergio Islas-Andrade[28], Xavier Soberon[14], Joanne E. Curran[29], Christopher P. Jenkinson[30], Ralph A. DeFronzo[31], Donna M. Lehman[32], Craig L. Hanis[33], Graeme I. Bell[34,35], Michael Boehnke[36], John Blangero[29], Ravindranath Duggirala[30], Richa Saxena[6,18,19], Daniel MacArthur[6,7,8], Jorge Ferrer[13,37,38], Steven A. McCarroll[4,5,6], David Torrents[3,39], William C. Knowler[12], Leslie J. Baier[12], Noel Burtt[1], Clicerio González-Villalpando[21], Christopher A. Haiman[24], Carlos A. Aguilar-Salinas[22], Teresa Tusié-Luna[9], Jason Flannick[1,2,40], Suzanne B.R. Jacobs[1,2], Lorena Orozco[14], David Altshuler[2,4,6,8,18,40,41], Jose C. Florez[1,2,8,#]

[¢]Members of the consortia are provided in Appendix S1.

*These authors contributed equally to this work.

#To whom correspondence should be addressed.

1. Programs in Metabolism and Medical & Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA.
2. Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA.
3. Barcelona Supercomputing Center (BSC). Joint BSC-CRG-IRB Research Program in Computational Biology, 08034 Barcelona, Spain.
4. Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA.
5. Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.
6. Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA.
7. Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.
8. Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA.

9. Unidad de Biología Molecular y Medicina Genómica, I.d.I.B., UNAM/ Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán. Mexico City, México. Instituto de Investigaciones Biomédicas, UNAM Unidad de Biología Molecular y Medicina Genómica, UNAM/INCMNSZ, Coyoacán, 04510 Mexico City, Mexico.
10. Universidad Autónoma Metropolitana, Tlalpan 14387, Mexico City, Mexico.
11. The Biostatistics Center, George Washington University, Rockville, MD, 20852, USA.
12. Phoenix Epidemiology and Clinical Research Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Phoenix, Arizona, 85004, USA.
13. Department of Medicine, Imperial College London, London W12 0NN, United Kingdom.
14. Instituto Nacional de Medicina Genómica, Tlalpan, 14610, Mexico City, Mexico.
15. Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX, USA.
16. Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, New York, USA.
17. Department of Psychiatry and Psychotherapy, Charité–Universitätsmedizin Berlin, Campus Mitte, 10117 Berlin, Germany.
18. Center for Genomic Medicne, Massachusetts General Hospital, Boston, Massachusetts, USA.
19. Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA.
20. Clínica de Integral de Cirugía para la Obesidad y Enfermedades Metabólicas, Hospital General Tláhuac, Secretaría de Salud del CDMX. México City.
21. Centro de Estudios en Diabetes, Unidad de Investigacion en Diabetes y Riesgo Cardiovascular, Centro de Investigacion en Salud Poblacional, Instituto Nacional de Salud Publica, Mexico City, Mexico.
22. Departamento de Endocrinología y Metabolismo. Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City.
23. Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California,USA.
24. Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii, USA.
25. The Genomics Platform, The Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA.
26. Department of Medicine,University of Texas Health Science Center at San Antonio, San Antonio, Texas, USA.
27. Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA.
28. Unidad de Investigación Médica en Enfermedades Metabólicas, CMN SXXI, Instituto Mexicano del Seguro Social, Mexico City, México.
29. South Texas Diabetes and Obesity Institute, School of Medicine, University of Texas Rio Grande Valley, Brownsville, TX, USA.
30. South Texas Diabetes and Obesity Institute, School of Medicine, University of Texas Rio Grande Valley, Edinburg, TX, USA.
31. Division of Diabetes, Department of Medicine, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA.
32. Departments of Medicine and Cellular & Structural Biology, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA.
33. Human Genetics Center, University of Texas Health Science Center at Houston, Houston, Texas 77030, USA.
34. Department of Medicine, University of Chicago, Chicago, Illinois 60637, USA.
35. Department of Human Genetics, University of Chicago, Chicago, Illinois 60637, USA.
36. Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan 48109, USA.

37. Genomic Programming of Beta-cells Laboratory, Institut d'Investigacions August Pi i Sunyer (IDIBAPS), 08036 Barcelona, Spain.
38. CIBER de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), 08036 Barcelona, Spain.
39. Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.
40. Department of Molecular Biology, Harvard Medical School, Boston, Massachusetts, USA.
41. Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

Corresponding author:

Jose C. Florez, M. D. Ph. D.
Chief, Diabetes Unit
Massachusetts General Hospital
Associate Professor of Medicine
Harvard Medical School
Institute Member
Broad Institute
Diabetes Unit, Department of Medicine
Center for Genomic Medicine
Richard B. Simches Research Center
Massachusetts General Hospital
185 Cambridge Street, CPZN 5.250
Boston, MA 02114
Office: 617-643-3308
Fax: 617-726-5735
Email: jcflorez@mgh.harvard.edu

Type 2 diabetes (T2D) affects more than 415 million people worldwide and its costs to the health care system continue to rise. To identify common or rare genetic variation with potential therapeutic implications for T2D, we analyzed and replicated genome-wide protein coding variation in a total of 8,227 individuals with T2D and 12,966 individuals without T2D of Latino descent. We identified a novel genetic variant in the *IGF2* gene associated with ~20% reduced risk for T2D. This variant, which has an allele frequency of 17% in the Mexican population but is rare in Europe, prevents splicing between *IGF2* exons 1 and 2. We show *in vitro* and in human liver and adipose tissue that the variant is associated with a specific, allele-dosage dependent reduction in expression of *IGF2* isoform 2. In individuals who do not carry the protective allele, expression of *IGF2* isoform 2 in adipose is positively correlated with both incidence of T2D and increased plasma glycated hemoglobin in individuals without T2D, providing support that the protective effects are mediated by reductions in *IGF2* isoform 2. Broad phenotypic examination of carriers of the protective variant revealed no association with other disease states or impaired reproductive health. These findings suggest that reducing *IGF2* isoform 2 expression in relevant tissues has potential as a new therapeutic strategy for T2D, also beyond the Latin-American population, with no major adverse effects on health or reproduction.

For Peer Review Only

Introduction

Type 2 diabetes (T2D) affects 415 million people worldwide and is predicted to be the 7th leading cause of death by 2030 (1). T2D is also the leading cause of preventable blindness (2) and end-stage renal disease (3) and is a major risk factor for heart attack and stroke (4).

An individual's risk of developing T2D is influenced by a combination of lifestyle, environmental, and genetic factors. Uncovering the genetic contributors to diabetes holds promise for clinical impact by revealing new therapeutic targets aimed at the molecular and cellular mechanisms that lead to disease. Genome-wide association studies (GWAS) performed during the past decade have uncovered more than 100 regions associated with T2D (5-12). While these studies have provided a better understanding of T2D genetics, the majority of identified variants fall outside protein-coding regions, leaving the molecular mechanism by which these variants confer altered disease risk obscure. Consequently, T2D GWAS have identified few *loci* with clear therapeutic potential.

The identification of loss-of-function variants associated with reduced risk of disease is of particular interest, as their protective genetic effect can be potentially recapitulated by pharmacological inhibition. Furthermore, if carriers of protective, loss-of-function variants are otherwise healthy, this suggests that specific pharmacological perturbation of the effector protein could confer benefit without significant adverse health effects (13).

Genetic explorations in traditionally understudied populations have succeeded in identifying novel T2D variants in Mexican populations (6; 14), as well as in East-Asians (15), Greenlanders (16), and African Americans (8). In Mexico, T2D is one of the leading causes of death and has a prevalence twice that of non-Hispanic whites in the US and among the highest worldwide (17; 18). While different environmental and lifestyle risk factors in Mexico partially explain the increased prevalence of T2D, unique genetic influences also contribute (6; 19). Here, we explored protein-coding variants present at higher frequency in people of Latino descent to shed further light on

For Peer Review Only

genetic risk factors for T2D in Mexico. We identified a novel T2D association with a protective, splice-acceptor variant that disrupts expression of *IGF2* isoform 2, providing a clear hypothesis for future mechanism of action and therapeutic inquiries.

Research Design and Methods

Study participants

This study was performed as part of the Slim Initiative in Genomic Medicine for the Americas (SIGMA) Type2 Diabetes Consortium, whose goal is to improve the understanding of the genetic basis of type 2 diabetes in Mexican and Latin American populations. The discovery dataset consisted of four studies from Mexico or Mexicans living in the US comprising a total of 4,210 cases and 4,786 controls, which resulted in a final sample size of 4,052 cases and 4,606 controls after quality control of the genotyping data (Table 1, details of these studies are provided in the Supplementary Note). All participants from the discovery and replication datasets provided informed consent for conducting this study. Their respective local ethics committees approved all contributing studies.

Genotyping and quality control

The genotyping of the discovery sample was done using the Exome Illumina array at the genomics platform at the Broad Institute (Cambridge, MA). The Genomics Platform at the Broad Institute received, quality controlled and tracked DNA samples for Exome array processing. The exome array was designed in order to cover rare and low-frequency coding variants identified through whole-exome sequencing studies of 12,031 individuals from different populations including 362 individuals of Hispanic ancestry. Details on the genotyping of the different discovery and replication cohorts are provided in the Supplementary Note.

3,732 of the samples genotyped by the exome array also underwent whole-exome sequencing (19), and were used to create a population-specific reference panel in order to fine-map the association at the *IGF2* locus (Supplementary Figure 1, 2, Supplementary Note).

In vitro splicing assay

IGF2 minigenes including the first three exons and two introns of the *IGF2* gene (chr11:2150342-2156088, Hg19), and containing either the G or A allele of rs149483638, were synthesized by Genewiz and subcloned into the mammalian expression vector pcDNA3.1. A stop codon was introduced at the end of exon 3 to stop translation of the expressed protein. Human Embryonic Kidney 293 cells (HEK293 cells) were transfected with either minigene using TransIT transfection reagent (Mirus Bio). RNA was extracted from the cells 24 hours post-transfection using the RNeasy Extraction Kit (Qiagen) and 1µg of RNA was reverse-transcribed into cDNA using a High Capacity cDNA Reverse Transcription Kit (Applied Biosystems).

We used two probes to detect *IGF2* expression by droplet digital PCR (ddPCR): one probe that targets exon 3 and recognizes all *IGF2* isoforms (Bio-Rad, custom probe 10031276), and one probe that targets the exon 1-2 junction and recognizes only isoforms with exon1-2 splicing (Life technologies Cat# Hs04188275). A probe targeting *ACTB* (Bio-Rad Cat# 10031255) was used as an endogenous control for both *IGF2* probes. Reaction mixtures consisted of 1 µL of cDNA (diluted 200X from the RT-PCR reaction), 1x of Supermix (Version 1) for Probes (Bio-Rad), 1x of each probe (*IGF2*-specific and *ACTB*-specific), and water to a final volume of 20 µL. Each reaction was partitioned into droplets using a QX200 automatic droplet generator (Bio-Rad). The droplets then underwent PCR as follows: 95°C for 10 minutes, 40 cycles of 94°C for 30 seconds and 60°C for 1 minute, followed by 98°C for 10 minutes. The QX200 droplet reader (Bio-Rad) was then used to measure the fluorescence of each of the two fluorophores corresponding to the *ACTB* and *IGF2* probes. After subtracting the background *IGF2* signal detected in untransfected cells (which was

minimal), *IGF2* was normalized to *ACTB* within each sample. The level of exon 1-2 splicing is presented relative to the total *IGF2* for that sample, as determined by the exon 3 probe.

Visceral adipose and liver tissue collection

Visceral adipose and liver samples were collected from subjects undergoing bariatric surgery for severe obesity (body mass index [BMI] greater than 40 kg/m^2, or greater than 35 kg/m^2 with comorbid entities) or elective surgery in nonobese patients. Patients were selected for bariatric surgery after 6 months of rigorous lifestyle intervention. All individuals were Mexican Mestizos older than 18 years, carefully selected from the Integral Clinic of Surgery for Obesity and Metabolic Diseases or General Surgery Department at the Tláhuac Hospital in Mexico City. Tissue samples were obtained at the beginning of the surgery with harmonic scalpel in all cases as follows: visceral fat was obtained from the greater omentum at the middle of the greater curvature of the stomach. Liver biopsy was obtained at the distal end of the left hepatic lobe, just above the spleen. VAT and liver samples were frozen immediately after removal. The protocol for collecting VAT and liver samples was approved by the respective local research and ethics committees and all patients signed an informed consent. Genomic DNA was purified from whole blood samples and of the rs149483638 variant was performed as the described to the DMS2 cohort.

RNA was isolated at the Broad Institute genomics platform (Cambridge, MA, Online Supplementary Note).

RNA-seq analysis of adult and ESC-derived cell lines

RNA-seq datasets for ESC-derived human pancreatic progenitor cells (20), ESC-derived neuronal progenitor, trophoblast, mesendoderm and mesenchymal cells, as well as adult liver (21) and adult pancreatic islets (22) were aligned using STAR (23) against the hg19 reference genome, allowing for up to 10 mismatches and disallowing multimapping. Exon expression level was calculated in RPKM as described in Mortazavi et al (24).

The expression of IGF2 exon 2 across adult human tissues was queried using RNA sequencing data from the GTEx consortium (25) spanning 54 tissue types and 550 individuals (dbGaP Accession phs000424.v5.p1). The sample collection, sequencing and data processing have been described previously (25). For these analyses, the exon-level quantifications were generated using RNA-SeQC (26) with GENCODE version 18 reference annotations.

IGF2 isoform expression in vivo by droplet digital PCR (ddPCR)

For the tissue samples, we employed reverse-transcriptase droplet-digital PCR (RT-ddPCR, Bio-Rad) to measure the expression of *IGF2* using probes that targeted all *IGF2* isoforms (Life Technologies assay Hs01005963) and the specific isoform disrupted by the splice site variant (Life Technologies assay Hs04188276). Each assay was run separately, with an assay targeting *G2E3* used as an endogenous control, which was selected for stability across different samples and for showing levels of expression similar to IGF2 isoform 2 (forward primer: GTCCACACACCCTTTGAAAGTT; reverse primer: CAGGTTTATGACACAGGATGCTA; probe: CACCAAGGGTTTTCAGACCCTGC, HEX-labeled). In adipose tissue we used 30 ng of RNA to quantify exon 2 of *IGF2* and 5 ng to quantify total *IGF2* expression. In liver, we used 20 ng of RNA to quantify exon 2 of *IGF2* and 15 ng to quantify total *IGF2* expression. We used 1x of IGF2 assay, 1x of G2E3 assay primer probe mix (20x mixture containing 18 μM of forward and reverse primers each and 5 μM of fluorescent probe), 1x of 2x One-Step RT-ddPCR Supermix (Bio-Rad), 1mM manganese acetate (Bio-Rad), and water to a final volume of 20 μL. Each reaction was partitioned into thousands of nanoliter-sized droplets using a QX200 manual or automatic droplet generator (Bio-Rad). The droplets underwent PCR as follows: 60°C for 30 minutes, 95°C for 5 minutes, 50 cycles of 94°C for 30 seconds and 60°C for 1 minute, followed by 98°C for 10 minutes. Following PCR, the fluorescence from each of the two fluorophores corresponding to *IGF2* and *G2E3* was read by a QX200 droplet reader (Bio-Rad), yielding precise, digital counts of the number of droplets containing the RNA targeted by each assay. Data were processed using QuantaSoft

For Peer Review Only

software (Bio-Rad), which estimates the absolute concentration of input RNA templates by Poisson-correcting the fraction of droplets that are positive for each amplicon. We used the ratio of IGF2 concentration to control G2E3 concentration as the normalized IGF2 expression value for downstream analyses.

Plasma IGF2 measurements

Total, circulating IGF2 levels were measured in plasma from 120 individuals, 40 per genotype at rs149483638, which were matched by ancestry, BMI, age, sex, and T2D status. IGF2 measurements were performed by the VUMC Hormone Assay and Analytical Services Core, using a Millipore Human IGFI,II Magnetic Bead Panel (Catalog # HIGFMAG-52K). The assay was read on a Luminex MAGPIX instrument. The association results were compared using linear regression adjusting for the first two principal components, BMI, age, sex, and T2D status.

Statistics

We used efficient mixed-model association (EMMAX) in order to test the genetic variants for association with T2D adjusted by age, BMI and sex, while controlling for sample structure (27). Odds ratios (ORs) were estimated using logistic regression models on T2D status adjusting for age, BMI, and ancestry as specified in the Supplementary Note. The experiment wide statistical significance threshold was set to $p < 5 \times 10^{-8}$ to adjust for the number of variants evaluated.

For functional analyses, statistical analyses were performed using linear and logistic regression, as well as non-parametric tests and $p < 0.05$ was considered significant for these functional studies.

Integration of data and imputation

For the credible set analysis we first built two datasets. One dataset was comprised of 4,478 samples that had been genotyped by exome chip and OMNI 2.5. The other dataset comprised another subset of 3,732 samples genotyped by exome chip, OMNI2.5 and whole-exome sequencing, which we

integrated to build a population-specific reference panel for protein coding variation. We kept all the variants with MAF higher than 0.001 for both datasets. We phased both datasets with SHAPEIT2 (28) (version 2.5,). We then imputed the 1000 G (phase 3, release June 2014) into both datasets separately. We also imputed the whole-exome variants with the population specific reference panel described above into the samples that did not undergo whole-exome sequencing. We used *impute 2 information score* > 0.8 as a post-imputation quality control. We then performed the association analysis separately in each cohort using SNPTEST and adjusting for BMI, age, sex and the first two principal components to adjust for population stratification. We then meta-analyzed both results using Metal (29).

Results

We performed association analysis between T2D and each of the 158,892 non-monomorphic variants genotyped in the Illumina exome array that passed stringent quality control in 4,210 T2D cases and 4,786 controls from four different cohorts in Mexico or Mexicans living in the US (Table 1, See Online Methods). The top genome-wide significant ($p<5\times10^{-8}$) signals replicated previously reported variants, including those at *TCF7L2* and *KCNQ1* (30; 31), with consistent effect sizes and directions of effect (Figure 1a, Supplementary Table 1), and confirmed the association of variants in *SLC16A11*, originally identified in a genome-wide study of the same subjects included in the present analysis (6).

To identify variants enriched in the Mexican population, we next focused our analysis on variants of low or rare frequency in Europeans (minor allele frequency [MAF]<0.05), but common (MAF>0.05) in Mexicans (Figure 1b). Of novel findings in this analysis, a SNP predicted to disrupt a canonical splice acceptor site in *IGF2* achieved the highest statistical significance (rs149483638, MAF=0.17; OR=0.80, $p=1.6\times10^{-7}$). Heterozygous carriers of this variant have a 22% decreased risk of T2D, and risk in homozygous carriers is reduced by 40%. We did not find associations between rs194483638 and other glycemic or metabolic traits (Supplementary Table 2). This variant is rare in

individuals of European ancestry (MAF=0.0002) and at low-frequency in individuals of East Asian (MAF=0.01) or African ancestry (MAF=0.001) (32) (http://exac.broadinstitute.org/). This variant showed a stronger association with T2D when adjusting for population stratification using principal components, as the protective T allele was more frequent in individuals with higher Native American ancestry, which is also a risk factor for T2D. Thus, we identified a protective genetic factor for T2D, present in 17% of a Latino population.

We performed several analyses that suggest rs149483638 is the most likely causal variant for the protective signal. First, we confirmed that other rare variants do not explain the association through a phenomenon called "synthetic association" (33) (Supplementary Figure 1, 2, 3, Supplementary Text). Second, we established that known T2D variants at the nearby *KCNQ1* locus (6; 31; 34) do not explain the association signal, as the two independently-associated variants at the *KCNQ1* locus are in weak linkage disequilibrium (LD) with rs149483638 in our dataset, (r^2 with rs139647931=0.026, r^2 with rs2237897=0.028) and the T2D association with rs149483638 remains significant after conditioning for these two variants (OR=0.81, $p=6.9\times10^{-6}$). Last, we carried out an analysis to identify the most likely causal variant(s). To do so, we first integrated whole-exome sequencing data, available for a subset of 3,732 samples, with exome chip and genotyping array data from OMNI 2.5 and performed imputation with 1000G phase 3 reference panel in all the samples (Supplementary Figure 1, 2, Supplementary Text). We then used a Bayesian approach to prioritize and rank variants according to likelihood of being causal (Supplementary Text). This analysis identified the splice acceptor variant (rs149483638) as having the highest probability of being causal for the T2D-protective association (Supplementary Figure 2, Supplementary Table 3).

We then sought to replicate the rs149483638 association in four independent data sets: T2D cases and controls of Hispanic origin from the T2D-GENES Consortium (19) (MAF=0.12, OR=0.89, p=0.3), individuals of full-heritage American Indian ancestry from the Pima cohort (35) (MAF=0.14, OR=0.68, $p=0.1\times10^{-5}$), self-identified indigenous individuals from different ethnic

groups in Mexico (DMS2 cohort) (19) (MAF=0.36, OR=0.71, p=0.001) (see Online Methods), and a subsample of Hispanic individuals from the GERA cohort (36) (MAF=0.06, OR=0.82, p=0.11). A meta-analysis of the discovery and these replication studies produced a genome-wide significant association (OR=0.78, $p=5.6\times10^{-14,}$ Figure 1c). We also tested the association of rs149483638 with diabetes incidence in the subset of 616 Hispanic or American Indian prediabetic individuals that were followed for an average of three years in the Diabetes Prevention Program (DPP) (37). The direction of effect was consistent with the results in other datasets, but was not statistically significant (HR=0.76, p=0.24, Supplementary Table 4), possibly because of lower power in this dataset due to its smaller sample size and/or the inclusion of prediabetic individuals who are at high risk for T2D at baseline. As an additional replication, and to further confirm that the findings are not due to potential population stratification, we analyzed this variant in the San Antonio Families Study, using a family-based association approach (38-40) (N=2,980); results are consistent with those obtained through the population-based analyses (z=-2.3, p=0.02, Supplementary Table 5). The overall meta-analyses including these two last datasets further strengthened the observed association between rs149483638 and T2D (overall $p=4.8\times10^{-14}$, Supplementary Table 4).

Having confirmed that the rs149483638 is driving the association for T2D protection, we performed experiments to understand the mechanism through which this beneficial metabolic action occurs. Using *in silico* analyses, we found that the protective A allele of rs149483638 variant (allele defined in the reverse strand, in which *IGF2* is expressed) is predicted to disrupt a canonical splice-site acceptor controlling inclusion of exon 2 in *IGF2* isoform 2 (P01344-3, Uniprot). Compared to isoform 1, *IGF2* isoform 2 has 56 additional N-terminal amino acids, encoded by exon 2. Therefore, the A allele is predicted to specifically disrupt expression of isoform 2 (Figure 2a). *IGF2* isoform 2 is lowly expressed in most adult tissues (25), showing the highest expression in pancreatic islets, liver and adipose tissue, where it represents approximately 1-2 % of total *IGF2* transcripts (Supplementary Figure 5, Supplementary Figure 6).

To determine if rs149483638 affects splicing as predicted, we measured exon 1-2 splicing in human cells transfected with *IGF2* minigenes consisting of the first three exons and two introns of *IGF2* (chr11:2150342-2156088) and containing either the G or A allele of rs149483638. In contrast to the high level of exon 1-2 splicing detected from the G allele, no exon 1-2 splicing was detected in samples expressing the *IGF2* minigene containing the A allele (Figure 2b, c), indicating a specific effect of the rs14983836 variant on *IGF2* isoform 2 splicing.

To assess whether the alternative allele at rs149483638 alters transcript expression *in vivo*, we collected 34 liver and 133 adipose tissue samples from Mexican rs149483638 variant carriers and non-carriers and analyzed expression of *IGF2* isoform 2 by measuring levels of the exon 1-2 splice junction using droplet digital PCR. The dosage of the A allele was negatively correlated with expression of *IGF2* isoform 2 in both liver (rho=-0.75 spearman $p=3.2\times10^{-7}$) and adipose tissue (rho=-0.22 spearman $p=0.01$) (Figure 3). In contrast, no significant correlation was detected for *IGF2* isoform 1 expression, as measured by exon 3 (common to both isoforms but representative of isoform 1, which constitutes ~98% of *IGF2* in these tissues (Supplementary Figure 7a and 7b). Similarly, we observed no association between rs149483638 genotype and circulating levels of total IGF2, which is expected to reflect the majority isoform, isoform 1 (Supplementary Figure 8). Together, *in vitro* and *in vivo* studies indicate that the T2D-protective A allele cause a reduction of the expression of *IGF2* isoform 2 via disruption of exon 1-2 splicing.

Collectively, our results suggested that decreased expression of *IGF2* isoform 2 is associated with decreased risk of T2D. We formally tested the association between expression of *IGF2* isoform 2 and T2D status and glycemic traits relevant to T2D in homozygous non-carriers (GG) and observed reduced expression of the isoform 2 in visceral adipose in non-diabetic individuals, compared to T2D ($p=0.003$, Figure 4a). This finding provides a link between the genetic association, gene expression, and T2D risk, suggesting that a "dose-response" curve may exist between *IGF2* isoform 2 expression and T2D risk.

For Peer Review Only

Furthermore, expression of *IGF2* isoform 2 in visceral adipose tissue is positively correlated with plasma glycated haemoglobin (HbA1c) in individuals without T2D, or untreated subjects with T2D, in homozygous non-carriers (GG) (p=0.004, Figure 4b). We did not detect significant associations between *IGF2* isoform 2 expression and glycemic traits or T2D status in the liver, possibly due to smaller sample size and, therefore, reduced statistical power for this tissue. We also did not find associations between the expression of isoform 1 and HbA1c or T2D in either adipose tissue or liver, suggesting the protective effect is specific to *IGF2* isoform 2. Overall, these results suggest that pharmacological inhibition of *IGF2* isoform 2 levels or activity could recapitulate the protective effect of the rs149483638 variant.

To assess potential negative effects of isoform 2 perturbation, we screened available datasets containing information on humans homozygous for the A allele of rs149483638. In the Exome Aggregation Consortium database (32) (ExAC, http://exac.broadinstitute.org/), we observed that there were 240 AA homozygotes (isoform 2 knockouts) within the Latin American population, all of whom were free of severe clinically recognized pediatric diseases. Furthermore, within the discovery and replication cohorts, we identified 293 AA homozygous individuals for whom clinical history of other diseases and fertility records were available and compared them to up to 6,407 GG homozygous individuals. We found that A allele homozygotes show reduced risk for T2D (OR=0.63, p=0.004) but do not exhibit increased prevalence of other diseases, and have indistinguishable reproductive performance based on number of children and percentage of individuals with children (Supplementary Table 6, Supplementary Figure 9). We also performed a phenome-wide association analysis in the Genetic Epidemiology Research on Aging (GERA) cohort, which revealed that rs149483638 is not associated with increased risk for any of the 18 available relevant medical conditions (Supplementary Table 7, Figure 5). Together, these data suggest that reduced activity or levels of *IGF2* isoform 2 does not a have a negative impact on general health or fertility.

For Peer Review Only

Discussion

Human genetics has proved successful in uncovering genetic risk factors for both Mendelian and complex diseases. In addition to identifying individuals at increased risk for disease, knowledge of genetic variants that influence disease risk can be translated for clinical impact by identifying new potential therapeutic targets and improving the success rate of drug development. In particular, loss-of-function variants that are protective for disease are attractive drug development targets, as their reduced function can be mimicked by pharmacological inhibition (41). Through whole-genome screening of protein-coding variants in Latin Americans, we have identified a protective T2D variant that disrupts a protein-coding exon of *IGF2*, leading to lower *IGF2* isoform 2 expression. Even in the absence of the protective variant, lower expression of *IGF2* isoform 2 is observed in non-T2D subjects compared with T2D subjects, and correlates with lower HbA1c levels in non-diabetic individuals. Importantly, we found no genetic evidence that loss of *IGF2* isoform 2 has a major negative impact on human health or reproduction. These findings suggest that reducing *IGF2* isoform 2 levels could provide therapeutic benefit for patients with T2D without adverse side-effects.

While a role for IGF2 in T2D and related glycemic traits has been previously suggested, our findings validate *IGF2* as a gene relevant to T2D pathophysiology in human populations. Recent human genetic studies in Latino and African American populations identified T2D risk associations at the *INS-IGF2* locus (6; 8); however, the effector gene responsible for the modified T2D risk was not identified. Here, we identified a protective signal for which the most probable causal variant is functionally validated as having an impact on *IGF2* isoform 2 expression.

It is unclear why the allele frequency of rs149483638 shows so much variability across populations. Even within different Latin American populations, the MAF ranges from ~5% in Puerto Ricans to ~23% in Peruvians. In Mexicans, the frequency increases with the percentage of Native American

For Peer Review Only

ancestry. Future analyses should clarify whether this protective variant underwent positive selection or its variable frequency is a result of genetic drift (54).

IGF2 has previously been implicated in T2D due to well-established metabolic functions of isoform 1. IGF2 is a peptide hormone with 47% amino acid sequence identity to insulin that regulates growth and metabolism through binding with insulin receptor, insulin-like growth factor 1 receptor (IGF1R) and insulin-like growth factor 2 receptor (IGFR2) (42). IGF2 regulates fetal development and differentiation, and has an important role in embryonic growth (43). In the adult, *IGF2* is expressed in several tissues, with highest levels in liver, where it is synthesized and released into the periphery. In the pancreas, IGF2 promotes □-cell proliferation and survival (44). Dysregulation of *IGF2* expression has been reported in several metabolic diseases, including growth disorders (45), obesity (46), and diabetes (47).

In mice, which only express isoform 1, *Igf2* inactivation promotes brown pre-adipocyte differentiation, protecting from insulin resistance (48). On the other hand, *Igf2* isoform 1 overexpression in murine pancreatic □ cells causes □-cell dysfunction that ultimately leads to hyperglycemia (49; 50). In humans, IGF2 has also been indirectly implicated in T2D due to the association of variants in IGF2 mRNA binding protein 2 (encoded by *IGF2BP2*) with T2D (51; 52). The protein encoded by *IGF2BP2*, IMP2, regulates *IGF2* mRNA levels, and mice deficient for *Igf2bp2* are resistant to diet-induced obesity and show higher glucose tolerance and insulin sensitivity (52; 53). While these studies support a potential role for IGF2 in T2D pathogenesis, they are based on the function of IGF2 isoform 1, while the function(s) of IGF2 isoform 2 -the isoform implicated in our study- remain unknown. Future studies are needed to elucidate how IGF2 isoform 2 differs from isoform 1 in its regulation and its effects on human cellular metabolism and physiology.

Though the molecular and cellular links between reduced expression of *IGF2* isoform 2, glucose regulation, and T2D have not yet been established, the observations reported here suggest inhibition

of *IGF2* isoform 2 might be a potential strategy for prevention or treatment of T2D. Previously identified loss-of-function mutations that cause beneficial metabolic phenotypes have spurred the development of recently approved drugs. One such example is that of sodium-glucose transporter 2 (SGLT2) inhibitors, a new class of oral anti-diabetic agents that lower blood glucose and, for at least one agent in the class, also reduce cardiovascular events among individuals at high risk for such events (55). These agents mimic physiology observed in familial renal glucosuria, a condition caused by loss-of-function mutations in *SLC5A2*, which encodes SGLT2 (56). Therefore, perturbation of *IGF2* isoform 2, which protects against T2D without observable adverse phenotypes in humans, has potential for development as a novel metabolic therapy. Our findings that reduced expression of *IGF2* isoform 2 is correlated with lower prevalence of T2D in individuals who do not carry the protective allele further suggests that such a treatment could provide therapeutic benefit beyond Latin American populations.

Alternative splicing is a major source of proteome diversity, and miss-splicing of genes is a cause of several Mendelian diseases (57). Here, we demonstrate that disruption of a canonical splice-acceptor site is also associated with altered risk of a complex disease. This mechanism of variant action suggests a specific pharmacological strategy for T2D, namely, inducing exon skipping of *IGF2* exon 2 to prevent *IGF2* isoform 2 expression. Indeed, induction of exon-skipping through use of modified antisense oligonucleotides has been successfully applied as a molecular therapy for a form of Duchenne muscular dystrophy caused by a stop-coding mutation in exon 51 of *DMD* gene, and two drugs based on this idea are currently in advanced clinical trials (58). Overall, our identification of a T2D-protective splice variant in IGF2 suggests that modulating *IGF2* isoform splicing, possibly in accessible hepatic tissue, could be a possible strategy for preventing or treating T2D, and opens a new line of investigation to characterize the mechanism through which disruption of *IGF2* isoform 2 protects against T2D.

Conclusions

IGF2 loss-of-function T2D protective variant

Genetic and functional evidence suggest a specific *IGF2* isoform as functionally relevant for the T2D physiology. Loss-of-function of this isoform is associated with reduced risk of T2D and shows no evidence of increased risk for other diseases, highlighting this isoform as a potential therapeutic target for T2D. Our results open a new line of investigation to characterize the mechanism through which disruption of *IGF2* isoform 2 protects against T2D.

For Peer Review Only

Author contributions

J.M.M, K.E., A.H.C, H.M.M, L.R., H.E.A, C.P.J., R.A.D., D.M.L., C.L.H, R.E.H., G.B., M.B., J.B., R.D., D.M., C.G.V., C.A.H., C.A.A.S, T.T.L., J.F., S.B.R.J., L.O., D.A., J.C.F. conceived, planned and oversaw the study. J.M.M, K.O., J.F., A.H.C, H.M.M, designed, performed, and analyzed most experiments. L.C., S.R., V.A., P.F., A.W., C.H., performed additional statistical analyses. B.G.P., M.H., S.G., L.C, C.H. J.F., P.F., J.M.M., performed the genotyping and quality control of the data. J.M.M, S.B.G., D.T., G.A.W., S.B.R.J., designed and performed the phenome-wide analysis. R.G.L., Z.D., J.M.M, S.B.R.J, performed, analyzed, and oversaw the in vitro splicing experiments, A.D, K.T., J.M.M, S.A.M. performed and oversaw the expression assays in human samples. T.K., I.M., J.F. analyzed the RNA-seq data. R.L.H., M.L.O.S, R.R.G., M.R.T., Y.Y.K., H.G.O., F.C.C., F.,B.O., S.P., C.G.V., C.A.H., C.A.A.S, T.T.L., L.O. K.A.J., R.S., J.L., C.Z., A.M.H., E.J.C., E.M.C., C.C.C, M.E.G.V., I.C.B., L.M.H., D.G.V., U.A., L.R.W., L.L.M., O.A.C., L.R., S.I.A., X.S., J.E.C., W.K., provided patient samples and genetic data. L.C., N.P.B, M.L.C, provided administrative, technical, or material support. The T2D-GENES Consortium and DPP Research Group provided genetic data. J.M.M, K.E., H.M.M., G.W.A., R. L. H., J.F., S.B.R.J., D. A., J.C.F. wrote and edited the manuscript.

Acknowledgements

Dr. Jose C. Florez is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

J.M.M. declares no conflicts of interest; R.G.L. declares no conflicts of interest; A.D. declares no conflicts of interest; Z.D. declares no conflicts of interest; K.E. declares no conflicts of interest; T.T. declares no conflicts of interest; A.H.C. declares no conflicts of interest; H.M.M. declares no conflicts of interest; K.A.J. declares no conflicts of interest; R.L.H. declares no conflicts of interest; G.A.W. declares no conflicts of interest; I.M. declares no conflicts of interest; L.C. declares no

For Peer Review Only

conflicts of interest; V.A. declares no conflicts of interest; M.L.O.S. declares no conflicts of interest; R.R.G. declares no conflicts of interest; M.R.T. declares no conflicts of interest; Y.S.K. declares no conflicts of interest; H.G.O. declares no conflicts of interest; F.C.C. declares no conflicts of interest; F.B.O. declares no conflicts of interest; L.C. declares no conflicts of interest; S.P. declares no conflicts of interest; P.F. declares no conflicts of interest; A.W. declares no conflicts of interest; S.B.G. declares no conflicts of interest; C.H. declares no conflicts of interest; S.R. declares no conflicts of interest; K.T. declares no conflicts of interest; J.L. declares no conflicts of interest; C.Z. declares no conflicts of interest; A.M.H. declares no conflicts of interest; E.J.C. declares no conflicts of interest; E.M.C. declares no conflicts of interest; C.C.C. declares no conflicts of interest; M.E.G.V. declares no conflicts of interest; I.C.B. declares no conflicts of interest; L.M.H. declares no conflicts of interest; D.G.V. declares no conflicts of interest; U.A. declares no conflicts of interest; L.R.W. declares no conflicts of interest; L.L.M. declares no conflicts of interest; O.A.C. declares no conflicts of interest; M.H. declares no conflicts of interest; S.G. declares no conflicts of interest; M.L.C. declares no conflicts of interest; C.R.M. declares no conflicts of interest; S.I.A. declares no conflicts of interest; X.S. declares no conflicts of interest; J.E.C. declares no conflicts of interest; C.P.J. declares no conflicts of interest; R.A.D. declares no conflicts of interest; D.M.L. declares no conflicts of interest; C.L.H. declares no conflicts of interest; G.I.B. declares no conflicts of interest; M.B. declares no conflicts of interest; J.B. declares no conflicts of interest; R.D. declares no conflicts of interest; R.S. declares no conflicts of interest; D.M. declares no conflicts of interest; J.Fe. declares no conflicts of interest; S.A.M. declares no conflicts of interest; D.T. declares no conflicts of interest; W.C.K. declares no conflicts of interest; L.J.B. declares no conflicts of interest; N.B. declares no conflicts of interest; C.G.V. declares no conflicts of interest; C.A.H. declares no conflicts of interest; C.A.A.S. declares no conflicts of interest; T.T.L. declares no conflicts of interest; J.Fl. declares no conflicts of interest; S.B.R.J. declares no conflicts of interest; L.O. declares no conflicts of interest; D.A. declares no conflicts of interest. J.C.F received consulting honoraria from Merck and from Boehringer-Ingelheim.

This work was conducted as part of the Slim Initiative for Genomic Medicine, a joint U.S.-Mexico project funded by the Carlos Slim Foundation. The UNAM/INCMNSZ diabetes study was supported by Consejo Nacional de Ciencia y Tecnología grants 138826, 128877, CONACyT-SALUD 2009-01-115250, and a grant from Dirección General de Asuntos del Personal Académico, UNAM, IT 214711. The Diabetes in Mexico Study was supported by Consejo Nacional de Ciencia y Tecnología grant number 86867 and by Carlos Slim Foundation. The Mexico City Diabetes Study was supported by National Institutes of Health grant number R01HL24799 from the National Heart, Lung and Blood Institute and by the Consejo Nacional de Ciencia y Tenologia grants numbers: 2092, M9303, F677-M9407, 251M and 2005-C01-14502, SALUD 2010-2-151165. The Multiethnic Cohort was supported by National Institutes of Health grants CA54281 and CA063464. A.L.W. is supported by National Institutes of Health Ruth L. Kirschstein National Research Service Award number F32HG005944. The DMS2 cohort and the visceral adipose tissue and liver samples collection were supported by Consejo Nacional de Ciencia y Tecnología grants number SALUD-233970 and 223019 respectively. The Pima longitudinal study is supported by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases. The Diabetes Prevention Program Research Group is supported by R01 DK072041 and by the Intramural Research Program of NIDDK and by the Indian Health Service. The VUMC Hormone Assay and Analytical Services Core is supported by NIH grants DK059637 and DK020593.

Josep M. Mercader was supported by Sara Borrell Fellowship from the Instituto Carlos III, grant SEV-2011-00067 of Severo Ochoa Program, and EMBO short term fellowship, EFSD/Lilly research fellowship and Beatriu de Pinós fellowship from the Agency for Management of University and Research Grants (AGAUR).

The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health. Additional funds were provided by the NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. Donors were enrolled at Biospecimen Source Sites

For Peer Review Only

funded by NCI\SAIC-Frederick, Inc. (SAIC-F) subcontracts to the National Disease Research Interchange (10XS170), Roswell Park Cancer Institute (10XS171), and Science Care, Inc. (X10S172). The Laboratory, Data Analysis, and Coordinating Center (LDACC) was funded through a contract (HHSN268201000029C) to The Broad Institute, Inc. Biorepository operations were funded through an SAIC-F subcontract to Van Andel Institute (10ST1035). Additional data repository and project management were provided by SAIC-F (HHSN261200800001E). The Brain Bank was supported by a supplements to University of Miami grants DA006227 & DA033684 and to contract N01MH000028. Statistical Methods development grants were made to the University of Geneva (MH090941 & MH101814), the University of Chicago (MH090951, MH090937, MH101820, MH101825), the University of North Carolina-Chapel Hill (MH090936 & MH101819), Harvard University (MH090948), Stanford University (MH101782), Washington University St Louis (MH101810), and the University of Pennsylvania (MH101822). Researchers of the DMS2 study thank Olaf Iván Corro Labra and José Luis de Jesus García Ruíz from the "Comisión Nacional para el Desarrollo de los Pueblos Indígenas" for their support in sample collection, for which they were not compensated. We also acknowledge Saúl Cano-Colín for his technical assistance in the genotyping of rs149483638 variant. We also acknowledge Vicky Kaur for her technical assistance in collecting the plasma samples for measuring IGF2 circulating levels. We also thank Joan Bacardí for his assistance in the preparation figures.

This paper is dedicated to the memories of our colleagues Laura Riba, Hanna Abboud and Brian Henderson.

References

1. World Health Organization: Global status report on noncommunicable diseases 2010. World Health Organization 2011;
2. Kohner EM, Barry PJ: Prevention of blindness in diabetic retinopathy. Diabetologia 1984;26:173-179
3. Mogensen CE: Preventing end-stage renal disease. Diabet Med 1998;15 Suppl 4:S51-56
4. Laakso M: Hyperglycemia and cardiovascular disease in type 2 diabetes. Diabetes 1999;48:937-942
5. DIAbetes Genetics Replication Meta-analysis Consortium, Asian Genetic Epidemiology Network Type 2 Diabetes Consortium, South Asian Type 2 Diabetes Consortium, Mexican American Type 2 Diabetes Consortium, Type 2 Diabetes Genetic Exploration by Nex-generation sequencing in muylti-Ethnic Samples Consortium, Mahajan A, Go MJ, Zhang W, Below JE, Gaulton KJ, Ferreira T, Horikoshi M, Johnson AD, Ng MC, Prokopenko I, Saleheen D, Wang X, Zeggini E, Abecasis GR, Adair LS, Almgren P, Atalay M, Aung T, Baldassarre D, Balkau B, Bao Y, Barnett AH, Barroso I, Basit A, Been LF, Beilby J, Bell GI, Benediktsson R, Bergman RN, Boehm BO, Boerwinkle E, Bonnycastle LL, Burtt N, Cai Q, Campbell H, Carey J, Cauchi S, Caulfield M, Chan JC, Chang LC, Chang TJ, Chang YC, Charpentier G, Chen CH, Chen H, Chen YT, Chia KS, Chidambaram M, Chines PS, Cho NH, Cho YM, Chuang LM, Collins FS, Cornelis MC, Couper DJ, Crenshaw AT, van Dam RM, Danesh J, Das D, de Faire U, Dedoussis G, Deloukas P, Dimas AS, Dina C, Doney AS, Donnelly PJ, Dorkhan M, van Duijn C, Dupuis J, Edkins S, Elliott P, Emilsson V, Erbel R, Eriksson JG, Escobedo J, Esko T, Eury E, Florez JC, Fontanillas P, Forouhi NG, Forsen T, Fox C, Fraser RM, Frayling TM, Froguel P, Frossard P, Gao Y, Gertow K, Gieger C, Gigante B, Grallert H, Grant GB, Grrop LC, Groves CJ, Grundberg E, Guiducci C, Hamsten A, Han BG, Hara K, Hassanali N, Hattersley AT, Hayward C, Hedman AK, Herder C, Hofman A, Holmen OL, Hovingh K, Hreidarsson AB, Hu C, Hu FB, Hui J, Humphries SE, Hunt SE, Hunter DJ, Hveem K, Hydrie ZI, Ikegami H, Illig T, Ingelsson E, Islam M, Isomaa B, Jackson AU, Jafar T, James A, Jia W, Jockel KH, Jonsson A, Jowett JB, Kadowaki T, Kang HM, Kanoni S, Kao WH, Kathiresan S, Kato N, Katulanda P, Keinanen-Kiukaanniemi KM, Kelly AM, Khan H, Khaw KT, Khor CC, Kim HL, Kim S, Kim YJ, Kinnunen L, Klopp N, Kong A, Korpi-Hyovalti E, Kowlessur S, Kraft P, Kravic J, Kristensen MM, Krithika S, Kumar A, Kumate J, Kuusisto J, Kwak SH, Laakso M, Lagou V, Lakka TA, Langenberg C, Langford C, Lawrence R, Leander K, Lee JM, Lee NR, Li M, Li X, Li Y, Liang J, Liju S, Lim WY, Lind L, Lindgren CM, Lindholm E, Liu CT, Liu JJ, Lobbens S, Long J, Loos RJ, Lu W, Luan J, Lyssenko V, Ma RC, Maeda S, Magi R, Mannisto S, Matthews DR, Meigs JB, Melander O, Metspalu A, Meyer J, Mirza G, Mihailov E, Moebus S, Mohan V, Mohlke KL, Morris AD, Muhleisen TW, Muller-Nurasyid M, Musk B, Nakamura J, Nakashima E, Navarro P, Ng PK, Nica AC, Nilsson PM, Njolstad I, Nothen MM, Ohnaka K, Ong TH, Owen KR, Palmer CN, Pankow JS, Park KS, Parkin M, Pechlivanis S, Pedersen NL, Peltonen L, Perry JR, Peters A, Pinidiyapathirage JM, Platou CG, Potter S, Price JF, Qi L, Radha V, Rallidis L, Rasheed A, Rathman W, Rauramaa R, Raychaudhuri S, Rayner NW, Rees SD, Rehnberg E, Ripatti S, Robertson N, Roden M, Rossin EJ, Rudan I, Rybin D, Saaristo TE, Salomaa V, Saltevo J, Samuel M, Sanghera DK, Saramies J, Scott J, Scott LJ, Scott RA, Segre AV, Sehmi J, Sennblad B, Shah N, Shah S, Shera AS, Shu XO, Shuldiner AR, Sigurdsson G, Sijbrands E, Silveira A, Sim X, Sivapalaratnam S, Small KS, So WY, Stancakova A, Stefansson K, Steinbach G, Steinthorsdottir V, Stirrups K, Strawbridge RJ, Stringham HM, Sun Q, Suo C, Syvanen AC, Takayanagi R, Takeuchi F, Tay WT, Teslovich TM, Thorand B, Thorleifsson G, Thorsteinsdottir U, Tikkanen E, Trakalo J, Tremoli E, Trip MD, Tsai FJ, Tuomi T, Tuomilehto J, Uitterlinden AG, Valladares-Salgado A, Vedantam S, Veglia F, Voight BF, Wang C, Wareham NJ, Wennauer R, Wickremasinghe AR, Wilsgaard T, Wilson JF, Wiltshire S, Winckler W, Wong TY, Wood AR, Wu JY, Wu Y, Yamamoto K, Yamauchi T, Yang M, Yengo L, Yokota M, Young R, Zabaneh D, Zhang F, Zhang R, Zheng W, Zimmet PZ, Altshuler D, Bowden DW, Cho YS, Cox NJ, Cruz M, Hanis CL, Kooner

J, Lee JY, Seielstad M, Teo YY, Boehnke M, Parra EJ, Chambers JC, Tai ES, McCarthy MI, Morris AP: Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat Genet 2014;46:234-244
6. SIGMA Type 2 Diabetes Consortium, Williams AL, Jacobs SB, Moreno-Macias H, Huerta-Chagoya A, Churchhouse C, Marquez-Luna C, Garcia-Ortiz H, Gomez-Vazquez MJ, Burtt NP, Aguilar-Salinas CA, Gonzalez-Villalpando C, Florez JC, Orozco L, Haiman CA, Tusie-Luna T, Altshuler D: Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. Nature 2014;506:97-101
7. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segre AV, Steinthorsdottir V, Strawbridge RJ, Khan H, Grallert H, Mahajan A, Prokopenko I, Kang HM, Dina C, Esko T, Fraser RM, Kanoni S, Kumar A, Lagou V, Langenberg C, Luan J, Lindgren CM, Muller-Nurasyid M, Pechlivanis S, Rayner NW, Scott LJ, Wiltshire S, Yengo L, Kinnunen L, Rossin EJ, Raychaudhuri S, Johnson AD, Dimas AS, Loos RJ, Vedantam S, Chen H, Florez JC, Fox C, Liu CT, Rybin D, Couper DJ, Kao WH, Li M, Cornelis MC, Kraft P, Sun Q, van Dam RM, Stringham HM, Chines PS, Fischer K, Fontanillas P, Holmen OL, Hunt SE, Jackson AU, Kong A, Lawrence R, Meyer J, Perry JR, Platou CG, Potter S, Rehnberg E, Robertson N, Sivapalaratnam S, Stancakova A, Stirrups K, Thorleifsson G, Tikkanen E, Wood AR, Almgren P, Atalay M, Benediktsson R, Bonnycastle LL, Burtt N, Carey J, Charpentier G, Crenshaw AT, Doney AS, Dorkhan M, Edkins S, Emilsson V, Eury E, Forsen T, Gertow K, Gigante B, Grant GB, Groves CJ, Guiducci C, Herder C, Hreidarsson AB, Hui J, James A, Jonsson A, Rathmann W, Klopp N, Kravic J, Krjutskov K, Langford C, Leander K, Lindholm E, Lobbens S, Mannisto S, Mirza G, Muhleisen TW, Musk B, Parkin M, Rallidis L, Saramies J, Sennblad B, Shah S, Sigurethsson G, Silveira A, Steinbach G, Thorand B, Trakalo J, Veglia F, Wennauer R, Winckler W, Zabaneh D, Campbell H, van Duijn C, Uitterlinden AG, Hofman A, Sijbrands E, Abecasis GR, Owen KR, Zeggini E, Trip MD, Forouhi NG, Syvanen AC, Eriksson JG, Peltonen L, Nothen MM, Balkau B, Palmer CN, Lyssenko V, Tuomi T, Isomaa B, Hunter DJ, Qi L, Wellcome Trust Case Control C, Meta-Analyses of G, Insulin-related traits Consortium I, Genetic Investigation of ATC, Asian Genetic Epidemiology Network-Type 2 Diabetes C, South Asian Type 2 Diabetes C, Shuldiner AR, Roden M, Barroso I, Wilsgaard T, Beilby J, Hovingh K, Price JF, Wilson JF, Rauramaa R, Lakka TA, Lind L, Dedoussis G, Njolstad I, Pedersen NL, Khaw KT, Wareham NJ, Keinanen-Kiukaanniemi SM, Saaristo TE, Korpi-Hyovalti E, Saltevo J, Laakso M, Kuusisto J, Metspalu A, Collins FS, Mohlke KL, Bergman RN, Tuomilehto J, Boehm BO, Gieger C, Hveem K, Cauchi S, Froguel P, Baldassarre D, Tremoli E, Humphries SE, Saleheen D, Danesh J, Ingelsson E, Ripatti S, Salomaa V, Erbel R, Jockel KH, Moebus S, Peters A, Illig T, de Faire U, Hamsten A, Morris AD, Donnelly PJ, Frayling TM, Hattersley AT, Boerwinkle E, Melander O, Kathiresan S, Nilsson PM, Deloukas P, Thorsteinsdottir U, Groop LC, Stefansson K, Hu F, Pankow JS, Dupuis J, Meigs JB, Altshuler D, Boehnke M, McCarthy MI, Replication DIG, Meta-analysis C: Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet 2012;44:981-990
8. Ng MC, Shriner D, Chen BH, Li J, Chen WM, Guo X, Liu J, Bielinski SJ, Yanek LR, Nalls MA, Comeau ME, Rasmussen-Torvik LJ, Jensen RA, Evans DS, Sun YV, An P, Patel SR, Lu Y, Long J, Armstrong LL, Wagenknecht L, Yang L, Snively BM, Palmer ND, Mudgal P, Langefeld CD, Keene KL, Freedman BI, Mychaleckyj JC, Nayak U, Raffel LJ, Goodarzi MO, Chen YD, Taylor HA, Jr., Correa A, Sims M, Couper D, Pankow JS, Boerwinkle E, Adeyemo A, Doumatey A, Chen G, Mathias RA, Vaidya D, Singleton AB, Zonderman AB, Igo RP, Jr., Sedor JR, Consortium F, Kabagambe EK, Siscovick DS, McKnight B, Rice K, Liu Y, Hsueh WC, Zhao W, Bielak LF, Kraja A, Province MA, Bottinger EP, Gottesman O, Cai Q, Zheng W, Blot WJ, Lowe WL, Pacheco JA, Crawford DC, e MC, Consortium D, Grundberg E, Mu TC, Rich SS, Hayes MG, Shu XO, Loos RJ, Borecki IB, Peyser PA, Cummings SR, Psaty BM, Fornage M, Iyengar SK, Evans MK, Becker DM, Kao WH, Wilson JG, Rotter JI, Sale MM, Liu S, Rotimi CN, Bowden DW, Consortium ME-aotDiAA: Meta-analysis of genome-wide association studies in African Americans provides insights into the genetic architecture of type 2 diabetes. PLoS Genet 2014;10:e1004517

9. Kim YJ, Go MJ, Hu C, Hong CB, Kim YK, Lee JY, Hwang JY, Oh JH, Kim DJ, Kim NH, Kim S, Hong EJ, Kim JH, Min H, Kim Y, Zhang R, Jia W, Okada Y, Takahashi A, Kubo M, Tanaka T, Kamatani N, Matsuda K, consortium M, Park T, Oh B, Kimm K, Kang D, Shin C, Cho NH, Kim HL, Han BG, Lee JY, Cho YS: Large-scale genome-wide association studies in East Asians identify new genetic loci influencing metabolic traits. Nat Genet 2011;43:990-995
10. The Wellcome Trust Case Control Consortium: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661-678
11. Flannick J, Florez JC: Type 2 diabetes: genetic data sharing to advance complex disease research. Nat Rev Genet 2016;17:535-549
12. Fuchsberger C, Flannick J, Teslovich TM, Mahajan A, Agarwala V, Gaulton KJ, Ma C, Fontanillas P, Moutsianas L, McCarthy DJ, Rivas MA, Perry JR, Sim X, Blackwell TW, Robertson NR, Rayner NW, Cingolani P, Locke AE, Fernandez Tajes J, Highland HM, Dupuis J, Chines PS, Lindgren CM, Hartl C, Jackson AU, Chen H, Huyghe JR, van de Bunt M, Pearson RD, Kumar A, Muller-Nurasyid M, Grarup N, Stringham HM, Gamazon ER, Lee J, Chen Y, Scott RA, Below JE, Chen P, Huang J, Go MJ, Stitzel ML, Pasko D, Parker SC, Varga TV, Green T, Beer NL, Day-Williams AG, Ferreira T, Fingerlin T, Horikoshi M, Hu C, Huh I, Ikram MK, Kim BJ, Kim Y, Kim YJ, Kwon MS, Lee J, Lee S, Lin KH, Maxwell TJ, Nagai Y, Wang X, Welch RP, Yoon J, Zhang W, Barzilai N, Voight BF, Han BG, Jenkinson CP, Kuulasmaa T, Kuusisto J, Manning A, Ng MC, Palmer ND, Balkau B, Stancakova A, Abboud HE, Boeing H, Giedraitis V, Prabhakaran D, Gottesman O, Scott J, Carey J, Kwan P, Grant G, Smith JD, Neale BM, Purcell S, Butterworth AS, Howson JM, Lee HM, Lu Y, Kwak SH, Zhao W, Danesh J, Lam VK, Park KS, Saleheen D, So WY, Tam CH, Afzal U, Aguilar D, Arya R, Aung T, Chan E, Navarro C, Cheng CY, Palli D, Correa A, Curran JE, Rybin D, Farook VS, Fowler SP, Freedman BI, Griswold M, Hale DE, Hicks PJ, Khor CC, Kumar S, Lehne B, Thuillier D, Lim WY, Liu J, van der Schouw YT, Loh M, Musani SK, Puppala S, Scott WR, Yengo L, Tan ST, Taylor HA, Jr., Thameem F, Wilson G, Sr., Wong TY, Njolstad PR, Levy JC, Mangino M, Bonnycastle LL, Schwarzmayr T, Fadista J, Surdulescu GL, Herder C, Groves CJ, Wieland T, Bork-Jensen J, Brandslund I, Christensen C, Koistinen HA, Doney AS, Kinnunen L, Esko T, Farmer AJ, Hakaste L, Hodgkiss D, Kravic J, Lyssenko V, Hollensted M, Jorgensen ME, Jorgensen T, Ladenvall C, Justesen JM, Karajamaki A, Kriebel J, Rathmann W, Lannfelt L, Lauritzen T, Narisu N, Linneberg A, Melander O, Milani L, Neville M, Orho-Melander M, Qi L, Qi Q, Roden M, Rolandsson O, Swift A, Rosengren AH, Stirrups K, Wood AR, Mihailov E, Blancher C, Carneiro MO, Maguire J, Poplin R, Shakir K, Fennell T, DePristo M, Hrabe de Angelis M, Deloukas P, Gjesing AP, Jun G, Nilsson P, Murphy J, Onofrio R, Thorand B, Hansen T, Meisinger C, Hu FB, Isomaa B, Karpe F, Liang L, Peters A, Huth C, O'Rahilly SP, Palmer CN, Pedersen O, Rauramaa R, Tuomilehto J, Salomaa V, Watanabe RM, Syvanen AC, Bergman RN, Bharadwaj D, Bottinger EP, Cho YS, Chandak GR, Chan JC, Chia KS, Daly MJ, Ebrahim SB, Langenberg C, Elliott P, Jablonski KA, Lehman DM, Jia W, Ma RC, Pollin TI, Sandhu M, Tandon N, Froguel P, Barroso I, Teo YY, Zeggini E, Loos RJ, Small KS, Ried JS, DeFronzo RA, Grallert H, Glaser B, Metspalu A, Wareham NJ, Walker M, Banks E, Gieger C, Ingelsson E, Im HK, Illig T, Franks PW, Buck G, Trakalo J, Buck D, Prokopenko I, Magi R, Lind L, Farjoun Y, Owen KR, Gloyn AL, Strauch K, Tuomi T, Kooner JS, Lee JY, Park T, Donnelly P, Morris AD, Hattersley AT, Bowden DW, Collins FS, Atzmon G, Chambers JC, Spector TD, Laakso M, Strom TM, Bell GI, Blangero J, Duggirala R, Tai ES, McVean G, Hanis CL, Wilson JG, Seielstad M, Frayling TM, Meigs JB, Cox NJ, Sladek R, Lander ES, Gabriel S, Burtt NP, Mohlke KL, Meitinger T, Groop L, Abecasis G, Florez JC, Scott LJ, Morris AP, Kang HM, Boehnke M, Altshuler D, McCarthy MI: The genetic architecture of type 2 diabetes. Nature 2016;536:41-47
13. Plenge RM, Scolnick EM, Altshuler D: Validating therapeutic targets through human genetics. Nat Rev Drug Discov 2013;12:581-594
14. Estrada K, Aukrust I, Bjorkhaug L, Burtt NP, Mercader JM, Garcia-Ortiz H, Huerta-Chagoya A, Moreno-Macias H, Walford G, Flannick J, Williams AL, Gomez-Vazquez MJ, Fernandez-Lopez JC, Martinez-Hernandez A, Centeno-Cruz F, Mendoza-Caamal E, Revilla-Monsalve C, Islas-

Andrade S, Cordova EJ, Soberon X, Gonzalez-Villalpando ME, Henderson E, Wilkens LR, Le Marchand L, Arellano-Campos O, Ordonez-Sanchez ML, Rodriguez-Torres M, Rodriguez-Guillen R, Riba L, Najmi LA, Jacobs SB, Fennell T, Gabriel S, Fontanillas P, Hanis CL, Lehman DM, Jenkinson CP, Abboud HE, Bell GI, Cortes ML, Boehnke M, Gonzalez-Villalpando C, Orozco L, Haiman CA, Tusie-Luna T, Aguilar-Salinas CA, Altshuler D, Njolstad PR, Florez JC, MacArthur DG: Association of a low-frequency variant in HNF1A with type 2 diabetes in a Latino population. JAMA 2014;311:2305-2314

15. Hara K, Fujita H, Johnson TA, Yamauchi T, Yasuda K, Horikoshi M, Peng C, Hu C, Ma RC, Imamura M, Iwata M, Tsunoda T, Morizono T, Shojima N, So WY, Leung TF, Kwan P, Zhang R, Wang J, Yu W, Maegawa H, Hirose H, consortium D, Kaku K, Ito C, Watada H, Tanaka Y, Tobe K, Kashiwagi A, Kawamori R, Jia W, Chan JC, Teo YY, Shyong TE, Kamatani N, Kubo M, Maeda S, Kadowaki T: Genome-wide association study identifies three novel loci for type 2 diabetes. Human molecular genetics 2014;23:239-246

16. Moltke I, Grarup N, Jorgensen ME, Bjerregaard P, Treebak JT, Fumagalli M, Korneliussen TS, Andersen MA, Nielsen TS, Krarup NT, Gjesing AP, Zierath JR, Linneberg A, Wu X, Sun G, Jin X, Al-Aama J, Wang J, Borch-Johnsen K, Pedersen O, Nielsen R, Albrechtsen A, Hansen T: A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. Nature 2014;512:190-193

17. Prevention CfDCa: National diabetes fact sheet: National estimates and general information on diabetes and prediabetes in the United States. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention 2011;

18. Villalpando S, de la Cruz V, Rojas R, Shamah-Levy T, Avila MA, Gaona B, Rebollar R, Hernandez L: Prevalence and distribution of type 2 diabetes mellitus in Mexican adult population: a probabilistic survey. Salud publica de Mexico 2010;52 Suppl 1:S19-26

19. Sigma Type 2 Diabetes Consortium, Estrada K, Aukrust I, Bjorkhaug L, Burtt NP, Mercader JM, Garcia-Ortiz H, Huerta-Chagoya A, Moreno-Macias H, Walford G, Flannick J, Williams AL, Gomez-Vazquez MJ, Fernandez-Lopez JC, Martinez-Hernandez A, Jimenez-Morales S, Centeno-Cruz F, Mendoza-Caamal E, Revilla-Monsalve C, Islas-Andrade S, Cordova EJ, Soberon X, Gonzalez-Villalpando ME, Henderson E, Wilkens LR, Le Marchand L, Arellano-Campos O, Ordonez-Sanchez ML, Rodriguez-Torres M, Rodriguez-Guillen R, Riba L, Najmi LA, Jacobs SB, Fennell T, Gabriel S, Fontanillas P, Hanis CL, Lehman DM, Jenkinson CP, Abboud HE, Bell GI, Cortes ML, Boehnke M, Gonzalez-Villalpando C, Orozco L, Haiman CA, Tusie-Luna T, Aguilar-Salinas CA, Altshuler D, Njolstad PR, Florez JC, MacArthur DG: Association of a low-frequency variant in HNF1A with type 2 diabetes in a Latino population. JAMA 2014;311:2305-2314

20. Cebola I, Rodriguez-Segui SA, Cho CH, Bessa J, Rovira M, Luengo M, Chhatriwala M, Berry A, Ponsa-Cobas J, Maestro MA, Jennings RE, Pasquali L, Moran I, Castro N, Hanley NA, Gomez-Skarmeta JL, Vallier L, Ferrer J: TEAD and YAP regulate the enhancer network of human embryonic pancreatic progenitors. Nature cell biology 2015;17:615-626

21. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR, Farnham PJ, Hirst M, Lander ES, Mikkelsen TS, Thomson JA: The NIH Roadmap Epigenomics Mapping Consortium. Nature biotechnology 2010;28:1045-1048

22. Moran I, Akerman I, van de Bunt M, Xie R, Benazra M, Nammo T, Arnes L, Nakic N, Garcia-Hurtado J, Rodriguez-Segui S, Pasquali L, Sauty-Colace C, Beucher A, Scharfmann R, van Arensbergen J, Johnson PR, Berry A, Lee C, Harkins T, Gmyr V, Pattou F, Kerr-Conte J, Piemonti L, Berney T, Hanley N, Gloyn AL, Sussel L, Langman L, Brayman KL, Sander M, McCarthy MI, Ravassard P, Ferrer J: Human beta cell transcriptome analysis uncovers lncRNAs that are tissue-specific, dynamically regulated, and abnormally expressed in type 2 diabetes. Cell Metab 2012;16:435-448

23. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR: STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15-21

24. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature methods 2008;5:621-628
25. The GTEx Consortium: Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 2015;348:648-660
26. DeLuca DS, Levin JZ, Sivachenko A, Fennell T, Nazaire MD, Williams C, Reich M, Winckler W, Getz G: RNA-SeQC: RNA-seq metrics for quality control and process optimization. Bioinformatics 2012;28:1530-1532
27. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, Sabatti C, Eskin E: Variance component model to account for sample structure in genome-wide association studies. Nat Genet 2010;42:348-354
28. Delaneau O, Marchini J, Zagury JF: A linear complexity phasing method for thousands of genomes. Nature methods 2012;9:179-181
29. Willer CJ, Li Y, Abecasis GR: METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26:2190-2191
30. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M: A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 2007;316:1341-1345
31. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, Zeggini E, Huth C, Aulchenko YS, Thorleifsson G, McCulloch LJ, Ferreira T, Grallert H, Amin N, Wu G, Willer CJ, Raychaudhuri S, McCarroll SA, Langenberg C, Hofmann OM, Dupuis J, Qi L, Segre AV, van Hoek M, Navarro P, Ardlie K, Balkau B, Benediktsson R, Bennett AJ, Blagieva R, Boerwinkle E, Bonnycastle LL, Bengtsson Bostrom K, Bravenboer B, Bumpstead S, Burtt NP, Charpentier G, Chines PS, Cornelis M, Couper DJ, Crawford G, Doney AS, Elliott KS, Elliott AL, Erdos MR, Fox CS, Franklin CS, Ganser M, Gieger C, Grarup N, Green T, Griffin S, Groves CJ, Guiducci C, Hadjadj S, Hassanali N, Herder C, Isomaa B, Jackson AU, Johnson PR, Jorgensen T, Kao WH, Klopp N, Kong A, Kraft P, Kuusisto J, Lauritzen T, Li M, Lieverse A, Lindgren CM, Lyssenko V, Marre M, Meitinger T, Midthjell K, Morken MA, Narisu N, Nilsson P, Owen KR, Payne F, Perry JR, Petersen AK, Platou C, Proenca C, Prokopenko I, Rathmann W, Rayner NW, Robertson NR, Rocheleau G, Roden M, Sampson MJ, Saxena R, Shields BM, Shrader P, Sigurdsson G, Sparso T, Strassburger K, Stringham HM, Sun Q, Swift AJ, Thorand B, Tichet J, Tuomi T, van Dam RM, van Haeften TW, van Herpt T, van Vliet-Ostaptchouk JV, Walters GB, Weedon MN, Wijmenga C, Witteman J, Bergman RN, Cauchi S, Collins FS, Gloyn AL, Gyllensten U, Hansen T, Hide WA, Hitman GA, Hofman A, Hunter DJ, Hveem K, Laakso M, Mohlke KL, Morris AD, Palmer CN, Pramstaller PP, Rudan I, Sijbrands E, Stein LD, Tuomilehto J, Uitterlinden A, Walker M, Wareham NJ, Watanabe RM, Abecasis GR, Boehm BO, Campbell H, Daly MJ, Hattersley AT, Hu FB, Meigs JB, Pankow JS, Pedersen O, Wichmann HE, Barroso I, Florez JC, Frayling TM, Groop L, Sladek R, Thorsteinsdottir U, Wilson JF, Illig T, Froguel P, van Duijn CM, Stefansson K, Altshuler D, Boehnke M, McCarthy MI, investigators M, Consortium G: Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet 2010;42:579-589
32. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won HH, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M,

McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG, Exome Aggregation C: Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;536:285-291
33. Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB: Rare variants create synthetic genome-wide associations. PLoS Biol 8:e1000294
34. Yasuda K, Miyake K, Horikawa Y, Hara K, Osawa H, Furuta H, Hirota Y, Mori H, Jonsson A, Sato Y, Yamagata K, Hinokio Y, Wang HY, Tanahashi T, Nakamura N, Oka Y, Iwasaki N, Iwamoto Y, Yamada Y, Seino Y, Maegawa H, Kashiwagi A, Takeda J, Maeda E, Shin HD, Cho YM, Park KS, Lee HK, Ng MC, Ma RC, So WY, Chan JC, Lyssenko V, Tuomi T, Nilsson P, Groop L, Kamatani N, Sekine A, Nakamura Y, Yamamoto K, Yoshida T, Tokunaga K, Itakura M, Makino H, Nanjo K, Kadowaki T, Kasuga M: Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. Nat Genet 2008;40:1092-1097
35. Hanson RL, Rong R, Kobes S, Muller YL, Weil EJ, Curtis JM, Nelson RG, Baier LJ: Role of Established Type 2 Diabetes-Susceptibility Genetic Variants in a High Prevalence American Indian Population. Diabetes 2015;64:2646-2657
36. Kvale MN, Hesselson S, Hoffmann TJ, Cao Y, Chan D, Connell S, Croen LA, Dispensa BP, Eshragh J, Finn A, Gollub J, Iribarren C, Jorgenson E, Kushi LH, Lao R, Lu Y, Ludwig D, Mathauda GK, McGuire WB, Mei G, Miles S, Mittman M, Patil M, Quesenberry CP, Jr., Ranatunga D, Rowell S, Sadler M, Sakoda LC, Shapero M, Shen L, Shenoy T, Smethurst D, Somkin CP, Van Den Eeden SK, Walter L, Wan E, Webster T, Whitmer RA, Wong S, Zau C, Zhan Y, Schaefer C, Kwok PY, Risch N: Genotyping Informatics and Quality Control for 100,000 Subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort. Genetics 2015;200:1051-1060
37. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, Nathan DM: Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. The New England journal of medicine 2002;346:393-403
38. Mitchell BD, Kammerer CM, Blangero J, Mahaney MC, Rainwater DL, Dyke B, Hixson JE, Henkel RD, Sharp RM, Comuzzie AG, VandeBerg JL, Stern MP, MacCluer JW: Genetic and environmental contributions to cardiovascular risk factors in Mexican Americans. The San Antonio Family Heart Study. Circulation 1996;94:2159-2170
39. Hunt KJ, Lehman DM, Arya R, Fowler S, Leach RJ, Goring HHH, Almasy L, Blangero J, Dyer TD, Duggirala R, Stern MP: Genome-wide linkage analyses of type 2 diabetes in Mexican Americans - The San Antonio Family Diabetes/Gallbladder Study. Diabetes 2005;54:2655-2662
40. Coletta DK, Schneider J, Hu SL, Dyer TD, Puppala S, Farook VS, Arya R, Lehman DM, Blangero J, DeFronzo RA, Duggirala R, Jenkinson CP: Genome-wide linkage scan for genes influencing plasma triglyceride levels in the Veterans Administration Genetic Epidemiology Study. Diabetes 2009;58:279-284
41. Harper AR, Nayee S, Topol EJ: Protective alleles and modifier variants in human health and disease. Nat Rev Genet 2015;16:689-701
42. Livingstone C, Borai A: Insulin-like growth factor-II: its role in metabolic and endocrine disease. Clin Endocrinol (Oxf) 2014;80:773-781
43. Morali OG, Jouneau A, McLaughlin KJ, Thiery JP, Larue L: IGF-II promotes mesoderm formation. Dev Biol 2000;227:133-145
44. Hill DJ, Strutt B, Arany E, Zaina S, Coukell S, Graham CF: Increased and persistent circulating insulin-like growth factor II in neonatal transgenic mice suppresses developmental apoptosis in the pancreatic islets. Endocrinology 2000;141:1151-1157
45. Sparago A, Cerrato F, Vernucci M, Ferrero GB, Silengo MC, Riccio A: Microdeletions in the human H19 DMR result in loss of IGF2 imprinting and Beckwith-Wiedemann syndrome. Nat Genet 2004;36:958-960

46. Frystyk J, Skjaerbaek C, Vestbo E, Fisker S, Orskov H: Circulating levels of free insulin-like growth factors in obese subjects: the impact of type 2 diabetes. Diabetes Metab Res Rev 1999;15:314-322
47. Estil les E, Tellez N, Soler J, Montanya E: High sensitivity of beta-cell replication to the inhibitory effects of interleukin-1beta: modulation by adenoviral overexpression of IGF2 in rat islets. J Endocrinol 2009;203:55-63
48. Poher AL, Altirriba J, Veyrat-Durebex C, Rohner-Jeanrenaud F: Brown adipose tissue activity as a target for the treatment of obesity/insulin resistance. Frontiers in physiology 2015;6:4
49. Devedjian JC, George M, Casellas A, Pujol A, Visa J, Pelegrin M, Gros L, Bosch F: Transgenic mice overexpressing insulin-like growth factor-II in beta cells develop type 2 diabetes. J Clin Invest 2000;105:731-740
50. Casellas A, Mallol C, Salavert A, Jimenez V, Garcia M, Agudo J, Obach M, Haurigot V, Vila L, Molas M, Lage R, Morro M, Casana E, Ruberte J, Bosch F: Insulin-like Growth Factor 2 Overexpression Induces Beta-cell Dysfunction and Increases Beta-cell Susceptibility to Damage. J Biol Chem 2015;
51. Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JR, Rayner NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ, Harries LW, Marchini JL, Owen KR, Knight B, Cardon LR, Walker M, Hitman GA, Morris AD, Doney AS, McCarthy MI, Hattersley AT: Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science 2007;316:1336-1341
52. Diabetes Genetics Initiative of Broad Institute of H, Mit LU, Novartis Institutes of BioMedical R, Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, Hughes TE, Groop L, Altshuler D, Almgren P, Florez JC, Meyer J, Ardlie K, Bengtsson Bostrom K, Isomaa B, Lettre G, Lindblad U, Lyon HN, Melander O, Newton-Cheh C, Nilsson P, Orho-Melander M, Rastam L, Speliotes EK, Taskinen MR, Tuomi T, Guiducci C, Berglund A, Carlson J, Gianniny L, Hackett R, Hall L, Holmkvist J, Laurila E, Sjogren M, Sterner M, Surti A, Svensson M, Svensson M, Tewhey R, Blumenstiel B, Parkin M, Defelice M, Barry R, Brodeur W, Camarata J, Chia N, Fava M, Gibbons J, Handsaker B, Healy C, Nguyen K, Gates C, Sougnez C, Gage D, Nizzari M, Gabriel SB, Chirn GW, Ma Q, Parikh H, Richardson D, Ricke D, Purcell S: Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 2007;316:1331-1336
53. Dai N, Zhao L, Wrighting D, Kramer D, Majithia A, Wang Y, Cracan V, Borges-Rivera D, Mootha VK, Nahrendorf M, Thorburn DR, Minichiello L, Altshuler D, Avruch J: IGF2BP2/IMP2-Deficient mice resist obesity through enhanced translation of Ucp1 mRNA and Other mRNAs encoding mitochondrial proteins. Cell Metab 2015;21:609-621
54. Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooper R, Kwiatkowski D, Ward R, Lander ES: Detecting recent positive selection in the human genome from haplotype structure. Nature 2002;419:832-837
55. Zinman B, Wanner C, Lachin JM, Fitchett D, Bluhmki E, Hantel S, Mattheus M, Devins T, Johansen OE, Woerle HJ, Broedl UC, Inzucchi SE, Investigators E-RO: Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes. The New England journal of medicine 2015;373:2117-2128
56. Santer R, Kinner M, Lassen CL, Schneppenheim R, Eggert P, Bald M, Brodehl J, Daschner M, Ehrich JH, Kemper M, Li Volti S, Neuhaus T, Skovby F, Swift PG, Schaub J, Klaerke D: Molecular analysis of the SGLT2 gene in patients with renal glucosuria. Journal of the American Society of Nephrology : JASN 2003;14:2873-2882
57. Scotti MM, Swanson MS: RNA mis-splicing in disease. Nat Rev Genet 2016;17:19-32
58. Kole R, Krieg AM: Exon skipping therapy for Duchenne muscular dystrophy. Adv Drug Deliv Rev 2015;87:104-107

IGF2 loss-of-function T2D protective variant

For Peer Review Only

Table 1. Study cohorts comprising the SIGMA type 2 diabetes exome chip project data set

Study	Sample location	Study design		n	Percent male	Age (years)	Age-of-onset (years)	BMI (kg/m^2)	Fasting plasma glucose (mmol/l)
UNAM/INCMNSZ Diabetes Study (UIDS)	Mexico City, Mexico	Prospective cohort	Controls	1164	41.3	55.4 (9.4)	-	28.2 (3.9)	4.8 (0.5)
			T2D cases	835	40.1	56.3 (12.4)	44.2 (11.4)	28.6 (4.6)	9.8 (4.5)
Diabetes in Mexico Study (DMS)	Mexico City, Mexico	Prospective cohort	Controls	486	25.3	52.6 (7.8)	-	28.1 (4.5)	5 (0.4)
			T2D cases	715	32.3	55.9 (11)	47.7 (10.4)	29 (5.6)	8.8 (3.9)
Mexico City Diabetes Study (MCDS)	Mexico City, Mexico	Prospective cohort	Controls	671	38.3	62.2 (7.7)	-	29.4 (4.6)	5 (0.6)
			T2D cases	315	40.3	63.9 (7.5)	54.7 (9.7)	30 (5.3)	8.8 (4)
Multiethnic Cohort (MEC)	Los Angeles, California, USA	Case-control	Controls	2285	48.5	59.2 (7)	-	26.6 (3.9)	-
			T2D cases	2187	47.6	59.1 (6.9)	-	29.9 (5.3)	-

The table shows sample location, study design, numbers of cases and controls, percent male participants, age, age-of-onset in cases, body mass index, and fasting plasma glucose in controls. Data are mean (SD).

Figure legends

Figure 1. Discovery and replication of the rs149483638 T2D protective variant. QQ plot for all common variants (**a**) and for Mexican population-specific variants (**b**). The plot shows the two most significant variants that have low frequency in Europeans, but higher frequency in the Mexican population. (**c**) Forest plot for the meta-analysis of rs149483638 variant in *IGF2*. We replicated the rs149483638 association in four independent data sets: 1,007 T2D cases and 917 controls of Hispanic origin from the T2D-GENES project (minor allele frequency [MAF]=0.12, odds ratio [OR]=0.98, p=0.3), 1,519 T2D cases and 1,680 controls of full-heritage American Indian ancestry from the Pima cohort (MAF=0.14, OR=0.68, $p=10^{-6}$), 427 cases and 751 controls of self-identified indigenous individuals from different ethnic groups in Mexico (DMS2 cohort) (MAF=0.36, OR=0.71, p=0.001) and 1,064 cases and 4,832 controls from the subset of cases from Latino ancestry (MAF=0.06, OR=0.82, p=0.11).

Figure 2. rs149483638 prevents splicing *in vitro*. (**a**) This variant is located at a canonical splice acceptor site, and is predicted to cause skipping of exon 2 of *IGF2* isoform 2. (**b**) 293T cells were transfected with *IGF2* minigenes containing the first three exons and two introns of the *IGF2* gene, and either allele of the rs149483638 C>T variant (G>A in the reverse strand) and cDNA was analyzed by droplet digital (ddPCR). This analysis revealed no expression of the *IGF2* exon 1-2 junction in cells transfected with the minigene containing the T2D-protective rs149483638 A allele. This was in contrast to the high levels of exon 1-2 splicing detected in cells transfected with the G allele. (**c**) One-dimensional plots of the ddPCR droplets plotted in (b). No *IGF2* transcript was detected in untransfected samples. *ACTB* was used as an internal control.

For Peer Review Only

Figure 3. rs149483638 prevents splicing between exon 2 in liver and in adipose tissue. The dosage of the T2D protective A allele is correlated with lower expression of *IGF2* isoform 2 (as measured by expression levels of the exon 1-2 junction) in liver (n(GG) = 21, n(GA) = 9, n(AA) = 4) (**a**) and in adipose tissue (n(GG) = 83, n(GA) = 43, n(AA) = 5) (**b**).

Figure 4. Expression of *IGF2* isoform 2 with T2D and related. (a) Boxplots representing the expression of *IGF2* isoform 2 across T2D cases and controls in individuals homozygous for the G common allele. The linear model p-value represents the association between IGF2 isoform 2 expression, adjusted by age, body mass index, and sex. (**b**) *IGF2* isoform 2 positively correlates with higher plasma glycated hemoglobin (HbA1c) in non-diabetic participants. The grey area limited by the dashed red lines represent the 95% confidence interval of the slope of the linear regression. *Patients with HbA1c above 6.5% were non-T2D subjects according to the diagnostic criteria of Mexico at the time of extraction, as HbA1c was not considered a criterion in Mexico at the time of extraction. Therefore, none of the subjects were receiving any lowering glucose treatment. **For clarity, since the genotype is strongly associated with isoform 2 expression, only individuals carrying the GG genotype are plotted in (**a**) and (**b**).

Figure 5. Phenome-wide analysis of rs149483638 variant. The protective variant was tested for association across 18 different disease traits previously categorized in the subsample of GERA cohort of Latino ancestry (5,896 individuals). While the rs149483638 variant was associated with reduced risk of T2D, there was no significant association seen for other 18 conditions. IBS: Irritable bowel syndrome; Mac. Degen.: Macular degeneration; Psychiatric: any psychiatric condition; PVD: Peripheral vascular disease; Stress: acute reaction to stress. Association analyses were done by logistic regression analyses, considering additive model, and correcting for age, BMI, sex, and the first two principal components to correct for population stratification.

A loss-of-function splice acceptor variant in *IGF2* is protective for type 2 diabetes

Running Title: *IGF2* loss-of-function type 2 diabetes protective variant.

The SIGMA T2D Genetics Consortium

Josep M Mercader[1,2,3], Rachel G. Liao[1]*, Avery Davis[4,5,6]*, Zachary Dymek[1]*, Karol Estrada[1,7,8], Taru Tukiainen[4,6,7], Alicia Huerta-Chagoya[9], Hortensia Moreno-Macías[9,10], Kathleen A. Jablonski[11], Robert L. Hanson[12], Geoffrey A. Walford[1,2,8], Ignasi Moran[13], Ling Chen[1,2], Vineeta Agarwala[6], María Luisa Ordoñez-Sánchez[9], Rosario Rodríguez-Guillen[9], Maribel Rodríguez-Torres[9], Yayoi Segura-Kato[9], Humberto García-Ortiz[14], Federico Centeno-Cruz[14], Francisco Barajas-Olmos[14], Lizz Caulkins[1], Sobha Puppala[15], Pierre Fontanillas[6], Amy Williams[16], Sílvia Bonàs-Guarch[3], Chris Hartl[6], Stephan Ripke[5,7,17], Diabetes Prevention Program Research Group[¢], Katherine Tooley[4,5,6], Jacqueline Lane[6,18,19], Carlos Zerrweck[20], Angélica Martínez-Hernández[14], Emilio J. Córdova[14], Elvia Mendoza-Caamal[14], Cecilia Contreras-Cubas[14], María E. González-Villalpando[21], Ivette Cruz-Bautista[22], Liliana Muñoz-Hernández[22], Donaji Gómez-Velasco[22], Ulises Alvirde[22], Brian E. Henderson[23], Lynne R. Wilkens[24], Loic Le Marchand[24], Olimpia Arellano-Campos[22], Laura Riba[22], Maegan Harden[25], Broad Genomics Platform[25], Stacey Gabriel[25], T2D-GENES Consortium[¢], Hanna E. Abboud[26], Maria L. Cortes[27], Cristina Revilla-Monsalve[28], Sergio Islas-Andrade[28], Xavier Soberon[14], Joanne E. Curran[29], Christopher P. Jenkinson[30], Ralph A. DeFronzo[31], Donna M. Lehman[32], Craig L. Hanis[33], Graeme I. Bell[34,35], Michael Boehnke[36], John Blangero[29], Ravindranath Duggirala[30], Richa Saxena[6,18,19], Daniel MacArthur[6,7,8], Jorge Ferrer[13,37,38], Steven A. McCarroll[4,5,6], David Torrents[3,39], William C. Knowler[12], Leslie J. Baier[12], Noel Burtt[1], Clicerio González-Villalpando[21], Christopher A. Haiman[24], Carlos A. Aguilar-Salinas[22], Teresa Tusié-Luna[9], Jason Flannick[1,2,40], Suzanne B.R. Jacobs[1,2], Lorena Orozco[14], David Altshuler[2,4,6,8,18,40,41], Jose C. Florez[1,2,8,#]

[¢]Members of the consortia are provided in Appendix S1.

*These authors contributed equally to this work.

#To whom correspondence should be addressed.

1. Programs in Metabolism and Medical & Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA.
2. Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA.
3. Barcelona Supercomputing Center (BSC). Joint BSC-CRG-IRB Research Program in Computational Biology, 08034 Barcelona, Spain.
4. Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA.
5. Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.
6. Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA.
7. Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.
8. Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA.

For Peer Review Only

9. Unidad de Biología Molecular y Medicina Genómica, I.d.I.B., UNAM/ Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán. Mexico City, México. Instituto de Investigaciones Biomédicas, UNAM Unidad de Biología Molecular y Medicina Genómica, UNAM/INCMNSZ, Coyoacán, 04510 Mexico City, Mexico.
10. Universidad Autónoma Metropolitana, Tlalpan 14387, Mexico City, Mexico.
11. The Biostatistics Center, George Washington University, Rockville, MD, 20852, USA.
12. Phoenix Epidemiology and Clinical Research Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Phoenix, Arizona, 85004, USA.
13. Department of Medicine, Imperial College London, London W12 0NN, United Kingdom.
14. Instituto Nacional de Medicina Genómica, Tlalpan, 14610, Mexico City, Mexico.
15. Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX, USA.
16. Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, New York, USA.
17. Department of Psychiatry and Psychotherapy, Charité–Universitätsmedizin Berlin, Campus Mitte, 10117 Berlin, Germany.
18. Center for Genomic Medicne, Massachusetts General Hospital, Boston, Massachusetts, USA.
19. Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA.
20. Clínica de Integral de Cirugía para la Obesidad y Enfermedades Metabólicas, Hospital General Tláhuac, Secretaría de Salud del CDMX. México City.
21. Centro de Estudios en Diabetes, Unidad de Investigacion en Diabetes y Riesgo Cardiovascular, Centro de Investigacion en Salud Poblacional, Instituto Nacional de Salud Publica, Mexico City, Mexico.
22. Departamento de Endocrinología y Metabolismo. Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City.
23. Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California,USA.
24. Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii, USA.
25. The Genomics Platform, The Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA.
26. Department of Medicine,University of Texas Health Science Center at San Antonio, San Antonio, Texas, USA.
27. Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA.
28. Unidad de Investigación Médica en Enfermedades Metabólicas, CMN SXXI, Instituto Mexicano del Seguro Social, Mexico City, México.
29. South Texas Diabetes and Obesity Institute, School of Medicine, University of Texas Rio Grande Valley, Brownsville, TX, USA.
30. South Texas Diabetes and Obesity Institute, School of Medicine, University of Texas Rio Grande Valley, Edinburg, TX, USA.
31. Division of Diabetes, Department of Medicine, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA.
32. Departments of Medicine and Cellular & Structural Biology, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA.
33. Human Genetics Center, University of Texas Health Science Center at Houston, Houston, Texas 77030, USA.
34. Department of Medicine, University of Chicago, Chicago, Illinois 60637, USA.
35. Department of Human Genetics, University of Chicago, Chicago, Illinois 60637, USA.
36. Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan 48109, USA.

37. Genomic Programming of Beta-cells Laboratory, Institut d'Investigacions August Pi i Sunyer (IDIBAPS), 08036 Barcelona, Spain.
38. CIBER de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), 08036 Barcelona, Spain.
39. Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.
40. Department of Molecular Biology, Harvard Medical School, Boston, Massachusetts, USA.
41. Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

Corresponding author:

Jose C. Florez, M. D. Ph. D.
Chief, Diabetes Unit
Massachusetts General Hospital
Associate Professor of Medicine
Harvard Medical School
Institute Member
Broad Institute
Diabetes Unit, Department of Medicine
Center for Genomic Medicine
Richard B. Simches Research Center
Massachusetts General Hospital
185 Cambridge Street, CPZN 5.250
Boston, MA 02114
Office: 617-643-3308
Fax: 617-726-5735
Email: jcflorez@mgh.harvard.edu

Type 2 diabetes (T2D) affects more than 415 million people worldwide and its costs to the health care system continue to rise. To identify common or rare genetic variation with potential therapeutic implications for T2D, we analyzed and replicated genome-wide protein coding variation in a total of 8,227 individuals with T2D and 12,966 individuals without T2D of Latino descent. We identified a novel genetic variant in the *IGF2* gene associated with ~20% reduced risk for T2D. This variant, which has an allele frequency of 17% in the Mexican population but is rare in Europe, prevents splicing between *IGF2* exons 1 and 2. We show *in vitro* and in human liver and adipose tissue that the variant is associated with a specific, allele-dosage dependent reduction in expression of *IGF2* isoform 2. In individuals who do not carry the protective allele, expression of *IGF2* isoform 2 in adipose is positively correlated with both incidence of T2D and increased plasma glycated hemoglobin in individuals without T2D, providing support that the protective effects are mediated by reductions in *IGF2* isoform 2. Broad phenotypic examination of carriers of the protective variant revealed no association with other disease states or impaired reproductive health. These findings suggest that reducing *IGF2* isoform 2 expression in relevant tissues has potential as a new therapeutic strategy for T2D, also beyond the Latin-American population, with no major adverse effects on health or reproduction.

For Peer Review Only

Introduction

Type 2 diabetes (T2D) affects 415 million people worldwide and is predicted to be the 7th leading cause of death by 2030 (1). T2D is also the leading cause of preventable blindness (2) and end-stage renal disease (3) and is a major risk factor for heart attack and stroke (4).

An individual's risk of developing T2D is influenced by a combination of lifestyle, environmental, and genetic factors. Uncovering the genetic contributors to diabetes holds promise for clinical impact by revealing new therapeutic targets aimed at the molecular and cellular mechanisms that lead to disease. Genome-wide association studies (GWAS) performed during the past decade have uncovered more than 100 regions associated with T2D (5-12). While these studies have provided a better understanding of T2D genetics, the majority of identified variants fall outside protein-coding regions, leaving the molecular mechanism by which these variants confer altered disease risk obscure. Consequently, T2D GWAS have identified few *loci* with clear therapeutic potential.

The identification of loss-of-function variants associated with reduced risk of disease is of particular interest, as their protective genetic effect can be potentially recapitulated by pharmacological inhibition. Furthermore, if carriers of protective, loss-of-function variants are otherwise healthy, this suggests that specific pharmacological perturbation of the effector protein could confer benefit without significant adverse health effects (13).

Genetic explorations in traditionally understudied populations have succeeded in identifying novel T2D variants in Mexican populations (6; 14), as well as in East-Asians (15), Greenlanders (16), and African Americans (8). In Mexico, T2D is one of the leading causes of death and has a prevalence twice that of non-Hispanic whites in the US and among the highest worldwide (17; 18). While different environmental and lifestyle risk factors in Mexico partially explain the increased prevalence of T2D, unique genetic influences also contribute (6; 19). Here, we explored protein-coding variants present at higher frequency in people of Latino descent to shed further light on

For Peer Review Only

genetic risk factors for T2D in Mexico. We identified a novel T2D association with a protective, splice-acceptor variant that disrupts expression of *IGF2* isoform 2, providing a clear hypothesis for future mechanism of action and therapeutic inquiries.

Research Design and Methods

Study participants

This study was performed as part of the Slim Initiative in Genomic Medicine for the Americas (SIGMA) Type2 Diabetes Consortium, whose goal is to improve the understanding of the genetic basis of type 2 diabetes in Mexican and Latin American populations. The discovery dataset consisted of four studies from Mexico or Mexicans living in the US comprising a total of 4,210 cases and 4,786 controls, which resulted in a final sample size of 4,052 cases and 4,606 controls after quality control of the genotyping data (Table 1, details of these studies are provided in the Supplementary Note). All participants from the discovery and replication datasets provided informed consent for conducting this study. Their respective local ethics committees approved all contributing studies.

Genotyping and quality control

The genotyping of the discovery sample was done using the Exome Illumina array at the genomics platform at the Broad Institute (Cambridge, MA). The Genomics Platform at the Broad Institute received, quality controlled and tracked DNA samples for Exome array processing. The exome array was designed in order to cover rare and low-frequency coding variants identified through whole-exome sequencing studies of 12,031 individuals from different populations including 362 individuals of Hispanic ancestry. Details on the genotyping of the different discovery and replication cohorts are provided in the Supplementary Note.

3,732 of the samples genotyped by the exome array also underwent whole-exome sequencing (19), and were used to create a population-specific reference panel in order to fine-map the association at the *IGF2* locus (Supplementary Figure 1, 2, Supplementary Note).

In vitro splicing assay

IGF2 minigenes including the first three exons and two introns of the *IGF2* gene (chr11:2150342-2156088, Hg19), and containing either the G or A allele of rs149483638, were synthesized by Genewiz and subcloned into the mammalian expression vector pcDNA3.1. A stop codon was introduced at the end of exon 3 to stop translation of the expressed protein. Human Embryonic Kidney 293 cells (HEK293 cells) were transfected with either minigene using TransIT transfection reagent (Mirus Bio). RNA was extracted from the cells 24 hours post-transfection using the RNeasy Extraction Kit (Qiagen) and 1µg of RNA was reverse-transcribed into cDNA using a High Capacity cDNA Reverse Transcription Kit (Applied Biosystems).

We used two probes to detect *IGF2* expression by droplet digital PCR (ddPCR): one probe that targets exon 3 and recognizes all *IGF2* isoforms (Bio-Rad, custom probe 10031276), and one probe that targets the exon 1-2 junction and recognizes only isoforms with exon1-2 splicing (Life technologies Cat# Hs04188275). A probe targeting *ACTB* (Bio-Rad Cat# 10031255) was used as an endogenous control for both *IGF2* probes. Reaction mixtures consisted of 1 µL of cDNA (diluted 200X from the RT-PCR reaction), 1x of Supermix (Version 1) for Probes (Bio-Rad), 1x of each probe (*IGF2*-specific and *ACTB*-specific), and water to a final volume of 20 µL. Each reaction was partitioned into droplets using a QX200 automatic droplet generator (Bio-Rad). The droplets then underwent PCR as follows: 95°C for 10 minutes, 40 cycles of 94°C for 30 seconds and 60°C for 1 minute, followed by 98°C for 10 minutes. The QX200 droplet reader (Bio-Rad) was then used to measure the fluorescence of each of the two fluorophores corresponding to the *ACTB* and *IGF2* probes. After subtracting the background *IGF2* signal detected in untransfected cells (which was

minimal), *IGF2* was normalized to *ACTB* within each sample. The level of exon 1-2 splicing is presented relative to the total *IGF2* for that sample, as determined by the exon 3 probe.

Visceral adipose and liver tissue collection

Visceral adipose and liver samples were collected from subjects undergoing bariatric surgery for severe obesity (body mass index [BMI] greater than 40 kg/m^2, or greater than 35 kg/m^2 with comorbid entities) or elective surgery in nonobese patients. Patients were selected for bariatric surgery after 6 months of rigorous lifestyle intervention. All individuals were Mexican Mestizos older than 18 years, carefully selected from the Integral Clinic of Surgery for Obesity and Metabolic Diseases or General Surgery Department at the Tláhuac Hospital in Mexico City. Tissue samples were obtained at the beginning of the surgery with harmonic scalpel in all cases as follows: visceral fat was obtained from the greater omentum at the middle of the greater curvature of the stomach. Liver biopsy was obtained at the distal end of the left hepatic lobe, just above the spleen. VAT and liver samples were frozen immediately after removal. The protocol for collecting VAT and liver samples was approved by the respective local research and ethics committees and all patients signed an informed consent. Genomic DNA was purified from whole blood samples and of the rs149483638 variant was performed as the described to the DMS2 cohort.

RNA was isolated at the Broad Institute genomics platform (Cambridge, MA, Online Supplementary Note).

RNA-seq analysis of adult and ESC-derived cell lines

RNA-seq datasets for ESC-derived human pancreatic progenitor cells (20), ESC-derived neuronal progenitor, trophoblast, mesendoderm and mesenchymal cells, as well as adult liver (21) and adult pancreatic islets (22) were aligned using STAR (23) against the hg19 reference genome, allowing for up to 10 mismatches and disallowing multimapping. Exon expression level was calculated in RPKM as described in Mortazavi et al (24).

The expression of IGF2 exon 2 across adult human tissues was queried using RNA sequencing data from the GTEx consortium (25) spanning 54 tissue types and 550 individuals (dbGaP Accession phs000424.v5.p1). The sample collection, sequencing and data processing have been described previously (25). For these analyses, the exon-level quantifications were generated using RNA-SeQC (26) with GENCODE version 18 reference annotations.

IGF2 isoform expression in vivo by droplet digital PCR (ddPCR)

For the tissue samples, we employed reverse-transcriptase droplet-digital PCR (RT-ddPCR, Bio-Rad) to measure the expression of *IGF2* using probes that targeted all *IGF2* isoforms (Life Technologies assay Hs01005963) and the specific isoform disrupted by the splice site variant (Life Technologies assay Hs04188276). Each assay was run separately, with an assay targeting *G2E3* used as an endogenous control, which was selected for stability across different samples and for showing levels of expression similar to IGF2 isoform 2 (forward primer: GTCCACACACCCTTTGAAAGTT; reverse primer: CAGGTTTATGACACAGGATGCTA; probe: CACCAAGGGTTTTCAGACCCTGC, HEX-labeled). In adipose tissue we used 30 ng of RNA to quantify exon 2 of *IGF2* and 5 ng to quantify total *IGF2* expression. In liver, we used 20 ng of RNA to quantify exon 2 of *IGF2* and 15 ng to quantify total *IGF2* expression. We used 1x of IGF2 assay, 1x of G2E3 assay primer probe mix (20x mixture containing 18 μM of forward and reverse primers each and 5 μM of fluorescent probe), 1x of 2x One-Step RT-ddPCR Supermix (Bio-Rad), 1mM manganese acetate (Bio-Rad), and water to a final volume of 20 μL. Each reaction was partitioned into thousands of nanoliter-sized droplets using a QX200 manual or automatic droplet generator (Bio-Rad). The droplets underwent PCR as follows: 60°C for 30 minutes, 95°C for 5 minutes, 50 cycles of 94°C for 30 seconds and 60°C for 1 minute, followed by 98°C for 10 minutes. Following PCR, the fluorescence from each of the two fluorophores corresponding to *IGF2* and *G2E3* was read by a QX200 droplet reader (Bio-Rad), yielding precise, digital counts of the number of droplets containing the RNA targeted by each assay. Data were processed using QuantaSoft

For Peer Review Only

software (Bio-Rad), which estimates the absolute concentration of input RNA templates by Poisson-correcting the fraction of droplets that are positive for each amplicon. We used the ratio of IGF2 concentration to control G2E3 concentration as the normalized IGF2 expression value for downstream analyses.

Plasma IGF2 measurements

Total, circulating IGF2 levels were measured in plasma from 120 individuals, 40 per genotype at rs149483638, which were matched by ancestry, BMI, age, sex, and T2D status. IGF2 measurements were performed by the VUMC Hormone Assay and Analytical Services Core, using a Millipore Human IGFI,II Magnetic Bead Panel (Catalog # HIGFMAG-52K). The assay was read on a Luminex MAGPIX instrument. The association results were compared using linear regression adjusting for the first two principal components, BMI, age, sex, and T2D status.

Statistics

We used efficient mixed-model association (EMMAX) in order to test the genetic variants for association with T2D adjusted by age, BMI and sex, while controlling for sample structure (27). Odds ratios (ORs) were estimated using logistic regression models on T2D status adjusting for age, BMI, and ancestry as specified in the Supplementary Note. The experiment wide statistical significance threshold was set to $p < 5 \times 10^{-8}$ to adjust for the number of variants evaluated.

For functional analyses, statistical analyses were performed using linear and logistic regression, as well as non-parametric tests and $p < 0.05$ was considered significant for these functional studies.

Integration of data and imputation

For the credible set analysis we first built two datasets. One dataset was comprised of 4,478 samples that had been genotyped by exome chip and OMNI 2.5. The other dataset comprised another subset of 3,732 samples genotyped by exome chip, OMNI2.5 and whole-exome sequencing, which we

For Peer Review Only

integrated to build a population-specific reference panel for protein coding variation. We kept all the variants with MAF higher than 0.001 for both datasets. We phased both datasets with SHAPEIT2 (28) (version 2.5,). We then imputed the 1000 G (phase 3, release June 2014) into both datasets separately. We also imputed the whole-exome variants with the population specific reference panel described above into the samples that did not undergo whole-exome sequencing. We used *impute 2 information score* > 0.8 as a post-imputation quality control. We then performed the association analysis separately in each cohort using SNPTEST and adjusting for BMI, age, sex and the first two principal components to adjust for population stratification. We then meta-analyzed both results using Metal (29).

Results

We performed association analysis between T2D and each of the 158,892 non-monomorphic variants genotyped in the Illumina exome array that passed stringent quality control in 4,210 T2D cases and 4,786 controls from four different cohorts in Mexico or Mexicans living in the US (Table 1, See Online Methods). The top genome-wide significant ($p<5\times10^{-8}$) signals replicated previously reported variants, including those at *TCF7L2* and *KCNQ1* (30; 31), with consistent effect sizes and directions of effect (Figure 1a, Supplementary Table 1), and confirmed the association of variants in *SLC16A11*, originally identified in a genome-wide study of the same subjects included in the present analysis (6).

To identify variants enriched in the Mexican population, we next focused our analysis on variants of low or rare frequency in Europeans (minor allele frequency [MAF]<0.05), but common (MAF>0.05) in Mexicans (Figure 1b). Of novel findings in this analysis, a SNP predicted to disrupt a canonical splice acceptor site in *IGF2* achieved the highest statistical significance (rs149483638, MAF=0.17; OR=0.80, $p=1.6\times10^{-7}$). Heterozygous carriers of this variant have a 22% decreased risk of T2D, and risk in homozygous carriers is reduced by 40%. We did not find associations between rs194483638 and other glycemic or metabolic traits (Supplementary Table 2). This variant is rare in

individuals of European ancestry (MAF=0.0002) and at low-frequency in individuals of East Asian (MAF=0.01) or African ancestry (MAF=0.001) (32) (http://exac.broadinstitute.org/). This variant showed a stronger association with T2D when adjusting for population stratification using principal components, as the protective T allele was more frequent in individuals with higher Native American ancestry, which is also a risk factor for T2D. Thus, we identified a protective genetic factor for T2D, present in 17% of a Latino population.

We performed several analyses that suggest rs149483638 is the most likely causal variant for the protective signal. First, we confirmed that other rare variants do not explain the association through a phenomenon called "synthetic association" (33) (Supplementary Figure 1, 2, 3, Supplementary Text). Second, we established that known T2D variants at the nearby *KCNQ1* locus (6; 31; 34) do not explain the association signal, as the two independently-associated variants at the *KCNQ1* locus are in weak linkage disequilibrium (LD) with rs149483638 in our dataset, (r^2 with rs139647931=0.026, r^2 with rs2237897=0.028) and the T2D association with rs149483638 remains significant after conditioning for these two variants (OR=0.81, $p=6.9\times10^{-6}$). Last, we carried out an analysis to identify the most likely causal variant(s). To do so, we first integrated whole-exome sequencing data, available for a subset of 3,732 samples, with exome chip and genotyping array data from OMNI 2.5 and performed imputation with 1000G phase 3 reference panel in all the samples (Supplementary Figure 1, 2, Supplementary Text). We then used a Bayesian approach to prioritize and rank variants according to likelihood of being causal (Supplementary Text). This analysis identified the splice acceptor variant (rs149483638) as having the highest probability of being causal for the T2D-protective association (Supplementary Figure 2, Supplementary Table 3).

We then sought to replicate the rs149483638 association in four independent data sets: T2D cases and controls of Hispanic origin from the T2D-GENES Consortium (19) (MAF=0.12, OR=0.89, $p=0.3$), individuals of full-heritage American Indian ancestry from the Pima cohort (35) (MAF=0.14, OR=0.68, $p=0.1\times10^{-5}$), self-identified indigenous individuals from different ethnic

groups in Mexico (DMS2 cohort) (19) (MAF=0.36, OR=0.71, p=0.001) (see Online Methods), and a subsample of Hispanic individuals from the GERA cohort (36) (MAF=0.06, OR=0.82, p=0.11). A meta-analysis of the discovery and these replication studies produced a genome-wide significant association (OR=0.78, $p=5.6\times10^{-14,}$ Figure 1c). We also tested the association of rs149483638 with diabetes incidence in the subset of 616 Hispanic or American Indian prediabetic individuals that were followed for an average of three years in the Diabetes Prevention Program (DPP) (37). The direction of effect was consistent with the results in other datasets, but was not statistically significant (HR=0.76, p=0.24, Supplementary Table 4), possibly because of lower power in this dataset due to its smaller sample size and/or the inclusion of prediabetic individuals who are at high risk for T2D at baseline. As an additional replication, and to further confirm that the findings are not due to potential population stratification, we analyzed this variant in the San Antonio Families Study, using a family-based association approach (38-40) (N=2,980); results are consistent with those obtained through the population-based analyses (z=-2.3, p=0.02, Supplementary Table 5). The overall meta-analyses including these two last datasets further strengthened the observed association between rs149483638 and T2D (overall $p=4.8\times10^{-14}$, Supplementary Table 4).

Having confirmed that the rs149483638 is driving the association for T2D protection, we performed experiments to understand the mechanism through which this beneficial metabolic action occurs. Using *in silico* analyses, we found that the protective A allele of rs149483638 variant (allele defined in the reverse strand, in which *IGF2* is expressed) is predicted to disrupt a canonical splice-site acceptor controlling inclusion of exon 2 in *IGF2* isoform 2 (P01344-3, Uniprot). Compared to isoform 1, *IGF2* isoform 2 has 56 additional N-terminal amino acids, encoded by exon 2. Therefore, the A allele is predicted to specifically disrupt expression of isoform 2 (Figure 2a). *IGF2* isoform 2 is lowly expressed in most adult tissues (25), showing the highest expression in pancreatic islets, liver and adipose tissue, where it represents approximately 1-2 % of total *IGF2* transcripts (Supplementary Figure 5, Supplementary Figure 6).

To determine if rs149483638 affects splicing as predicted, we measured exon 1-2 splicing in human cells transfected with *IGF2* minigenes consisting of the first three exons and two introns of *IGF2* (chr11:2150342-2156088) and containing either the G or A allele of rs149483638. In contrast to the high level of exon 1-2 splicing detected from the G allele, no exon 1-2 splicing was detected in samples expressing the *IGF2* minigene containing the A allele (Figure 2b, c), indicating a specific effect of the rs14983836 variant on *IGF2* isoform 2 splicing.

To assess whether the alternative allele at rs149483638 alters transcript expression *in vivo*, we collected 34 liver and 133 adipose tissue samples from Mexican rs149483638 variant carriers and non-carriers and analyzed expression of *IGF2* isoform 2 by measuring levels of the exon 1-2 splice junction using droplet digital PCR. The dosage of the A allele was negatively correlated with expression of *IGF2* isoform 2 in both liver (rho=-0.75 spearman $p=3.2\times10^{-7}$) and adipose tissue (rho=-0.22 spearman $p=0.01$) (Figure 3). In contrast, no significant correlation was detected for *IGF2* isoform 1 expression, as measured by exon 3 (common to both isoforms but representative of isoform 1, which constitutes ~98% of *IGF2* in these tissues (Supplementary Figure 7a and 7b). Similarly, we observed no association between rs149483638 genotype and circulating levels of total IGF2, which is expected to reflect the majority isoform, isoform 1 (Supplementary Figure 8). Together, *in vitro* and *in vivo* studies indicate that the T2D-protective A allele cause a reduction of the expression of *IGF2* isoform 2 via disruption of exon 1-2 splicing.

Collectively, our results suggested that decreased expression of *IGF2* isoform 2 is associated with decreased risk of T2D. We formally tested the association between expression of *IGF2* isoform 2 and T2D status and glycemic traits relevant to T2D in homozygous non-carriers (GG) and observed reduced expression of the isoform 2 in visceral adipose in non-diabetic individuals, compared to T2D ($p=0.003$, Figure 4a). This finding provides a link between the genetic association, gene expression, and T2D risk, suggesting that a "dose-response" curve may exist between *IGF2* isoform 2 expression and T2D risk.

For Peer Review Only

Furthermore, expression of *IGF2* isoform 2 in visceral adipose tissue is positively correlated with plasma glycated haemoglobin (HbA1c) in individuals without T2D, or untreated subjects with T2D, in homozygous non-carriers (GG) (p=0.004, Figure 4b). We did not detect significant associations between *IGF2* isoform 2 expression and glycemic traits or T2D status in the liver, possibly due to smaller sample size and, therefore, reduced statistical power for this tissue. We also did not find associations between the expression of isoform 1 and HbA1c or T2D in either adipose tissue or liver, suggesting the protective effect is specific to *IGF2* isoform 2. Overall, these results suggest that pharmacological inhibition of *IGF2* isoform 2 levels or activity could recapitulate the protective effect of the rs149483638 variant.

To assess potential negative effects of isoform 2 perturbation, we screened available datasets containing information on humans homozygous for the A allele of rs149483638. In the Exome Aggregation Consortium database (32) (ExAC, http://exac.broadinstitute.org/), we observed that there were 240 AA homozygotes (isoform 2 knockouts) within the Latin American population, all of whom were free of severe clinically recognized pediatric diseases. Furthermore, within the discovery and replication cohorts, we identified 293 AA homozygous individuals for whom clinical history of other diseases and fertility records were available and compared them to up to 6,407 GG homozygous individuals. We found that A allele homozygotes show reduced risk for T2D (OR=0.63, p=0.004) but do not exhibit increased prevalence of other diseases, and have indistinguishable reproductive performance based on number of children and percentage of individuals with children (Supplementary Table 6, Supplementary Figure 9). We also performed a phenome-wide association analysis in the Genetic Epidemiology Research on Aging (GERA) cohort, which revealed that rs149483638 is not associated with increased risk for any of the 18 available relevant medical conditions (Supplementary Table 7, Figure 5). Together, these data suggest that reduced activity or levels of *IGF2* isoform 2 does not a have a negative impact on general health or fertility.

For Peer Review Only

Discussion

Human genetics has proved successful in uncovering genetic risk factors for both Mendelian and complex diseases. In addition to identifying individuals at increased risk for disease, knowledge of genetic variants that influence disease risk can be translated for clinical impact by identifying new potential therapeutic targets and improving the success rate of drug development. In particular, loss-of-function variants that are protective for disease are attractive drug development targets, as their reduced function can be mimicked by pharmacological inhibition (41). Through whole-genome screening of protein-coding variants in Latin Americans, we have identified a protective T2D variant that disrupts a protein-coding exon of *IGF2*, leading to lower *IGF2* isoform 2 expression. Even in the absence of the protective variant, lower expression of *IGF2* isoform 2 is observed in non-T2D subjects compared with T2D subjects, and correlates with lower HbA1c levels in non-diabetic individuals. Importantly, we found no genetic evidence that loss of *IGF2* isoform 2 has a major negative impact on human health or reproduction. These findings suggest that reducing *IGF2* isoform 2 levels could provide therapeutic benefit for patients with T2D without adverse side-effects.

While a role for IGF2 in T2D and related glycemic traits has been previously suggested, our findings validate *IGF2* as a gene relevant to T2D pathophysiology in human populations. Recent human genetic studies in Latino and African American populations identified T2D risk associations at the *INS-IGF2* locus (6; 8); however, the effector gene responsible for the modified T2D risk was not identified. Here, we identified a protective signal for which the most probable causal variant is functionally validated as having an impact on *IGF2* isoform 2 expression.

It is unclear why the allele frequency of rs149483638 shows so much variability across populations. Even within different Latin American populations, the MAF ranges from ~5% in Puerto Ricans to ~23% in Peruvians. In Mexicans, the frequency increases with the percentage of Native American

For Peer Review Only

ancestry. Future analyses should clarify whether this protective variant underwent positive selection or its variable frequency is a result of genetic drift (54).

IGF2 has previously been implicated in T2D due to well-established metabolic functions of isoform 1. IGF2 is a peptide hormone with 47% amino acid sequence identity to insulin that regulates growth and metabolism through binding with insulin receptor, insulin-like growth factor 1 receptor (IGF1R) and insulin-like growth factor 2 receptor (IGFR2) (42). IGF2 regulates fetal development and differentiation, and has an important role in embryonic growth (43). In the adult, *IGF2* is expressed in several tissues, with highest levels in liver, where it is synthesized and released into the periphery. In the pancreas, IGF2 promotes □-cell proliferation and survival (44). Dysregulation of *IGF2* expression has been reported in several metabolic diseases, including growth disorders (45), obesity (46), and diabetes (47).

In mice, which only express isoform 1, *Igf2* inactivation promotes brown pre-adipocyte differentiation, protecting from insulin resistance (48). On the other hand, *Igf2* isoform 1 overexpression in murine pancreatic □ cells causes □-cell dysfunction that ultimately leads to hyperglycemia (49; 50). In humans, IGF2 has also been indirectly implicated in T2D due to the association of variants in IGF2 mRNA binding protein 2 (encoded by *IGF2BP2*) with T2D (51; 52). The protein encoded by *IGF2BP2*, IMP2, regulates *IGF2* mRNA levels, and mice deficient for *Igf2bp2* are resistant to diet-induced obesity and show higher glucose tolerance and insulin sensitivity (52; 53). While these studies support a potential role for IGF2 in T2D pathogenesis, they are based on the function of IGF2 isoform 1, while the function(s) of IGF2 isoform 2 -the isoform implicated in our study- remain unknown. Future studies are needed to elucidate how IGF2 isoform 2 differs from isoform 1 in its regulation and its effects on human cellular metabolism and physiology.

Though the molecular and cellular links between reduced expression of *IGF2* isoform 2, glucose regulation, and T2D have not yet been established, the observations reported here suggest inhibition

For Peer Review Only

of *IGF2* isoform 2 might be a potential strategy for prevention or treatment of T2D. Previously identified loss-of-function mutations that cause beneficial metabolic phenotypes have spurred the development of recently approved drugs. One such example is that of sodium-glucose transporter 2 (SGLT2) inhibitors, a new class of oral anti-diabetic agents that lower blood glucose and, for at least one agent in the class, also reduce cardiovascular events among individuals at high risk for such events (55). These agents mimic physiology observed in familial renal glucosuria, a condition caused by loss-of-function mutations in *SLC5A2*, which encodes SGLT2 (56). Therefore, perturbation of *IGF2* isoform 2, which protects against T2D without observable adverse phenotypes in humans, has potential for development as a novel metabolic therapy. Our findings that reduced expression of *IGF2* isoform 2 is correlated with lower prevalence of T2D in individuals who do not carry the protective allele further suggests that such a treatment could provide therapeutic benefit beyond Latin American populations.

Alternative splicing is a major source of proteome diversity, and miss-splicing of genes is a cause of several Mendelian diseases (57). Here, we demonstrate that disruption of a canonical splice-acceptor site is also associated with altered risk of a complex disease. This mechanism of variant action suggests a specific pharmacological strategy for T2D, namely, inducing exon skipping of *IGF2* exon 2 to prevent *IGF2* isoform 2 expression. Indeed, induction of exon-skipping through use of modified antisense oligonucleotides has been successfully applied as a molecular therapy for a form of Duchenne muscular dystrophy caused by a stop-coding mutation in exon 51 of *DMD* gene, and two drugs based on this idea are currently in advanced clinical trials (58). Overall, our identification of a T2D-protective splice variant in IGF2 suggests that modulating *IGF2* isoform splicing, possibly in accessible hepatic tissue, could be a possible strategy for preventing or treating T2D, and opens a new line of investigation to characterize the mechanism through which disruption of *IGF2* isoform 2 protects against T2D.

Conclusions

For Peer Review Only

IGF2 loss-of-function T2D protective variant

Genetic and functional evidence suggest a specific *IGF2* isoform as functionally relevant for the T2D physiology. Loss-of-function of this isoform is associated with reduced risk of T2D and shows no evidence of increased risk for other diseases, highlighting this isoform as a potential therapeutic target for T2D. Our results open a new line of investigation to characterize the mechanism through which disruption of *IGF2* isoform 2 protects against T2D.

For Peer Review Only

Author contributions

J.M.M, K.E., A.H.C, H.M.M, L.R., H.E.A, C.P.J., R.A.D., D.M.L., C.L.H, R.E.H., G.B., M.B., J.B., R.D., D.M., C.G.V., C.A.H., C.A.A.S, T.T.L., J.F., S.B.R.J., L.O., D.A., J.C.F. conceived, planned and oversaw the study. J.M.M, K.O., J.F., A.H.C, H.M.M, designed, performed, and analyzed most experiments. L.C., S.R., V.A., P.F., A.W., C.H., performed additional statistical analyses. B.G.P., M.H., S.G., L.C, C.H. J.F., P.F., J.M.M., performed the genotyping and quality control of the data. J.M.M, S.B.G., D.T., G.A.W., S.B.R.J., designed and performed the phenome-wide analysis. R.G.L., Z.D., J.M.M, S.B.R.J, performed, analyzed, and oversaw the in vitro splicing experiments, A.D, K.T., J.M.M, S.A.M. performed and oversaw the expression assays in human samples. T.K., I.M., J.F. analyzed the RNA-seq data. R.L.H., M.L.O.S, R.R.G., M.R.T., Y.Y.K., H.G.O., F.C.C., F.,B.O., S.P., C.G.V., C.A.H., C.A.A.S, T.T.L., L.O. K.A.J., R.S., J.L., C.Z., A.M.H., E.J.C., E.M.C., C.C.C, M.E.G.V., I.C.B., L.M.H., D.G.V., U.A., L.R.W., L.L.M., O.A.C., L.R., S.I.A., X.S., J.E.C., W.K., provided patient samples and genetic data. L.C., N.P.B, M.L.C, provided administrative, technical, or material support. The T2D-GENES Consortium and DPP Research Group provided genetic data. J.M.M, K.E., H.M.M., G.W.A., R. L. H., J.F., S.B.R.J., D. A., J.C.F. wrote and edited the manuscript.

Acknowledgements

Dr. Jose C. Florez is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

J.M.M. declares no conflicts of interest; R.G.L. declares no conflicts of interest; A.D. declares no conflicts of interest; Z.D. declares no conflicts of interest; K.E. declares no conflicts of interest; T.T. declares no conflicts of interest; A.H.C. declares no conflicts of interest; H.M.M. declares no conflicts of interest; K.A.J. declares no conflicts of interest; R.L.H. declares no conflicts of interest; G.A.W. declares no conflicts of interest; I.M. declares no conflicts of interest; L.C. declares no

For Peer Review Only

conflicts of interest; V.A. declares no conflicts of interest; M.L.O.S. declares no conflicts of interest; R.R.G. declares no conflicts of interest; M.R.T. declares no conflicts of interest; Y.S.K. declares no conflicts of interest; H.G.O. declares no conflicts of interest; F.C.C. declares no conflicts of interest; F.B.O. declares no conflicts of interest; L.C. declares no conflicts of interest; S.P. declares no conflicts of interest; P.F. declares no conflicts of interest; A.W. declares no conflicts of interest; S.B.G. declares no conflicts of interest; C.H. declares no conflicts of interest; S.R. declares no conflicts of interest; K.T. declares no conflicts of interest; J.L. declares no conflicts of interest; C.Z. declares no conflicts of interest; A.M.H. declares no conflicts of interest; E.J.C. declares no conflicts of interest; E.M.C. declares no conflicts of interest; C.C.C. declares no conflicts of interest; M.E.G.V. declares no conflicts of interest; I.C.B. declares no conflicts of interest; L.M.H. declares no conflicts of interest; D.G.V. declares no conflicts of interest; U.A. declares no conflicts of interest; L.R.W. declares no conflicts of interest; L.L.M. declares no conflicts of interest; O.A.C. declares no conflicts of interest; M.H. declares no conflicts of interest; S.G. declares no conflicts of interest; M.L.C. declares no conflicts of interest; C.R.M. declares no conflicts of interest; S.I.A. declares no conflicts of interest; X.S. declares no conflicts of interest; J.E.C. declares no conflicts of interest; C.P.J. declares no conflicts of interest; R.A.D. declares no conflicts of interest; D.M.L. declares no conflicts of interest; C.L.H. declares no conflicts of interest; G.I.B. declares no conflicts of interest; M.B. declares no conflicts of interest; J.B. declares no conflicts of interest; R.D. declares no conflicts of interest; R.S. declares no conflicts of interest; D.M. declares no conflicts of interest; J.Fe. declares no conflicts of interest; S.A.M. declares no conflicts of interest; D.T. declares no conflicts of interest; W.C.K. declares no conflicts of interest; L.J.B. declares no conflicts of interest; N.B. declares no conflicts of interest; C.G.V. declares no conflicts of interest; C.A.H. declares no conflicts of interest; C.A.A.S. declares no conflicts of interest; T.T.L. declares no conflicts of interest; J.Fl. declares no conflicts of interest; S.B.R.J. declares no conflicts of interest; L.O. declares no conflicts of interest; D.A. declares no conflicts of interest. J.C.F received consulting honoraria from Merck and from Boehringer-Ingelheim.

This work was conducted as part of the Slim Initiative for Genomic Medicine, a joint U.S.-Mexico project funded by the Carlos Slim Foundation. The UNAM/INCMNSZ diabetes study was supported by Consejo Nacional de Ciencia y Tecnología grants 138826, 128877, CONACyT-SALUD 2009-01-115250, and a grant from Dirección General de Asuntos del Personal Académico, UNAM, IT 214711. The Diabetes in Mexico Study was supported by Consejo Nacional de Ciencia y Tecnología grant number 86867 and by Carlos Slim Foundation. The Mexico City Diabetes Study was supported by National Institutes of Health grant number R01HL24799 from the National Heart, Lung and Blood Institute and by the Consejo Nacional de Ciencia y Tenologia grants numbers: 2092, M9303, F677-M9407, 251M and 2005-C01-14502, SALUD 2010-2-151165. The Multiethnic Cohort was supported by National Institutes of Health grants CA54281 and CA063464. A.L.W. is supported by National Institutes of Health Ruth L. Kirschstein National Research Service Award number F32HG005944. The DMS2 cohort and the visceral adipose tissue and liver samples collection were supported by Consejo Nacional de Ciencia y Tecnología grants number SALUD-233970 and 223019 respectively. The Pima longitudinal study is supported by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases. The Diabetes Prevention Program Research Group is supported by R01 DK072041 and by the Intramural Research Program of NIDDK and by the Indian Health Service. The VUMC Hormone Assay and Analytical Services Core is supported by NIH grants DK059637 and DK020593.

Josep M. Mercader was supported by Sara Borrell Fellowship from the Instituto Carlos III, grant SEV-2011-00067 of Severo Ochoa Program, and EMBO short term fellowship, EFSD/Lilly research fellowship and Beatriu de Pinós fellowship from the Agency for Management of University and Research Grants (AGAUR).

The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health. Additional funds were provided by the NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. Donors were enrolled at Biospecimen Source Sites

For Peer Review Only

funded by NCI\SAIC-Frederick, Inc. (SAIC-F) subcontracts to the National Disease Research Interchange (10XS170), Roswell Park Cancer Institute (10XS171), and Science Care, Inc. (X10S172). The Laboratory, Data Analysis, and Coordinating Center (LDACC) was funded through a contract (HHSN268201000029C) to The Broad Institute, Inc. Biorepository operations were funded through an SAIC-F subcontract to Van Andel Institute (10ST1035). Additional data repository and project management were provided by SAIC-F (HHSN261200800001E). The Brain Bank was supported by a supplements to University of Miami grants DA006227 & DA033684 and to contract N01MH000028. Statistical Methods development grants were made to the University of Geneva (MH090941 & MH101814), the University of Chicago (MH090951, MH090937, MH101820, MH101825), the University of North Carolina-Chapel Hill (MH090936 & MH101819), Harvard University (MH090948), Stanford University (MH101782), Washington University St Louis (MH101810), and the University of Pennsylvania (MH101822). Researchers of the DMS2 study thank Olaf Iván Corro Labra and José Luis de Jesus García Ruíz from the "Comisión Nacional para el Desarrollo de los Pueblos Indígenas" for their support in sample collection, for which they were not compensated. We also acknowledge Saúl Cano-Colín for his technical assistance in the genotyping of rs149483638 variant. We also acknowledge Vicky Kaur for her technical assistance in collecting the plasma samples for measuring IGF2 circulating levels. We also thank Joan Bacardí for his assistance in the preparation figures.

This paper is dedicated to the memories of our colleagues Laura Riba, Hanna Abboud and Brian Henderson.

References

1. World Health Organization: Global status report on noncommunicable diseases 2010. World Health Organization 2011;
2. Kohner EM, Barry PJ: Prevention of blindness in diabetic retinopathy. Diabetologia 1984;26:173-179
3. Mogensen CE: Preventing end-stage renal disease. Diabet Med 1998;15 Suppl 4:S51-56
4. Laakso M: Hyperglycemia and cardiovascular disease in type 2 diabetes. Diabetes 1999;48:937-942
5. DIAbetes Genetics Replication Meta-analysis Consortium, Asian Genetic Epidemiology Network Type 2 Diabetes Consortium, South Asian Type 2 Diabetes Consortium, Mexican American Type 2 Diabetes Consortium, Type 2 Diabetes Genetic Exploration by Nex-generation sequencing in muylti-Ethnic Samples Consortium, Mahajan A, Go MJ, Zhang W, Below JE, Gaulton KJ, Ferreira T, Horikoshi M, Johnson AD, Ng MC, Prokopenko I, Saleheen D, Wang X, Zeggini E, Abecasis GR, Adair LS, Almgren P, Atalay M, Aung T, Baldassarre D, Balkau B, Bao Y, Barnett AH, Barroso I, Basit A, Been LF, Beilby J, Bell GI, Benediktsson R, Bergman RN, Boehm BO, Boerwinkle E, Bonnycastle LL, Burtt N, Cai Q, Campbell H, Carey J, Cauchi S, Caulfield M, Chan JC, Chang LC, Chang TJ, Chang YC, Charpentier G, Chen CH, Chen H, Chen YT, Chia KS, Chidambaram M, Chines PS, Cho NH, Cho YM, Chuang LM, Collins FS, Cornelis MC, Couper DJ, Crenshaw AT, van Dam RM, Danesh J, Das D, de Faire U, Dedoussis G, Deloukas P, Dimas AS, Dina C, Doney AS, Donnelly PJ, Dorkhan M, van Duijn C, Dupuis J, Edkins S, Elliott P, Emilsson V, Erbel R, Eriksson JG, Escobedo J, Esko T, Eury E, Florez JC, Fontanillas P, Forouhi NG, Forsen T, Fox C, Fraser RM, Frayling TM, Froguel P, Frossard P, Gao Y, Gertow K, Gieger C, Gigante B, Grallert H, Grant GB, Grrop LC, Groves CJ, Grundberg E, Guiducci C, Hamsten A, Han BG, Hara K, Hassanali N, Hattersley AT, Hayward C, Hedman AK, Herder C, Hofman A, Holmen OL, Hovingh K, Hreidarsson AB, Hu C, Hu FB, Hui J, Humphries SE, Hunt SE, Hunter DJ, Hveem K, Hydrie ZI, Ikegami H, Illig T, Ingelsson E, Islam M, Isomaa B, Jackson AU, Jafar T, James A, Jia W, Jockel KH, Jonsson A, Jowett JB, Kadowaki T, Kang HM, Kanoni S, Kao WH, Kathiresan S, Kato N, Katulanda P, Keinanen-Kiukaanniemi KM, Kelly AM, Khan H, Khaw KT, Khor CC, Kim HL, Kim S, Kim YJ, Kinnunen L, Klopp N, Kong A, Korpi-Hyovalti E, Kowlessur S, Kraft P, Kravic J, Kristensen MM, Krithika S, Kumar A, Kumate J, Kuusisto J, Kwak SH, Laakso M, Lagou V, Lakka TA, Langenberg C, Langford C, Lawrence R, Leander K, Lee JM, Lee NR, Li M, Li X, Li Y, Liang J, Liju S, Lim WY, Lind L, Lindgren CM, Lindholm E, Liu CT, Liu JJ, Lobbens S, Long J, Loos RJ, Lu W, Luan J, Lyssenko V, Ma RC, Maeda S, Magi R, Mannisto S, Matthews DR, Meigs JB, Melander O, Metspalu A, Meyer J, Mirza G, Mihailov E, Moebus S, Mohan V, Mohlke KL, Morris AD, Muhleisen TW, Muller-Nurasyid M, Musk B, Nakamura J, Nakashima E, Navarro P, Ng PK, Nica AC, Nilsson PM, Njolstad I, Nothen MM, Ohnaka K, Ong TH, Owen KR, Palmer CN, Pankow JS, Park KS, Parkin M, Pechlivanis S, Pedersen NL, Peltonen L, Perry JR, Peters A, Pinidiyapathirage JM, Platou CG, Potter S, Price JF, Qi L, Radha V, Rallidis L, Rasheed A, Rathman W, Rauramaa R, Raychaudhuri S, Rayner NW, Rees SD, Rehnberg E, Ripatti S, Robertson N, Roden M, Rossin EJ, Rudan I, Rybin D, Saaristo TE, Salomaa V, Saltevo J, Samuel M, Sanghera DK, Saramies J, Scott J, Scott LJ, Scott RA, Segre AV, Sehmi J, Sennblad B, Shah N, Shah S, Shera AS, Shu XO, Shuldiner AR, Sigurdsson G, Sijbrands E, Silveira A, Sim X, Sivapalaratnam S, Small KS, So WY, Stancakova A, Stefansson K, Steinbach G, Steinthorsdottir V, Stirrups K, Strawbridge RJ, Stringham HM, Sun Q, Suo C, Syvanen AC, Takayanagi R, Takeuchi F, Tay WT, Teslovich TM, Thorand B, Thorleifsson G, Thorsteinsdottir U, Tikkanen E, Trakalo J, Tremoli E, Trip MD, Tsai FJ, Tuomi T, Tuomilehto J, Uitterlinden AG, Valladares-Salgado A, Vedantam S, Veglia F, Voight BF, Wang C, Wareham NJ, Wennauer R, Wickremasinghe AR, Wilsgaard T, Wilson JF, Wiltshire S, Winckler W, Wong TY, Wood AR, Wu JY, Wu Y, Yamamoto K, Yamauchi T, Yang M, Yengo L, Yokota M, Young R, Zabaneh D, Zhang F, Zhang R, Zheng W, Zimmet PZ, Altshuler D, Bowden DW, Cho YS, Cox NJ, Cruz M, Hanis CL, Kooner

J, Lee JY, Seielstad M, Teo YY, Boehnke M, Parra EJ, Chambers JC, Tai ES, McCarthy MI, Morris AP: Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat Genet 2014;46:234-244
6. SIGMA Type 2 Diabetes Consortium, Williams AL, Jacobs SB, Moreno-Macias H, Huerta-Chagoya A, Churchhouse C, Marquez-Luna C, Garcia-Ortiz H, Gomez-Vazquez MJ, Burtt NP, Aguilar-Salinas CA, Gonzalez-Villalpando C, Florez JC, Orozco L, Haiman CA, Tusie-Luna T, Altshuler D: Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. Nature 2014;506:97-101
7. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segre AV, Steinthorsdottir V, Strawbridge RJ, Khan H, Grallert H, Mahajan A, Prokopenko I, Kang HM, Dina C, Esko T, Fraser RM, Kanoni S, Kumar A, Lagou V, Langenberg C, Luan J, Lindgren CM, Muller-Nurasyid M, Pechlivanis S, Rayner NW, Scott LJ, Wiltshire S, Yengo L, Kinnunen L, Rossin EJ, Raychaudhuri S, Johnson AD, Dimas AS, Loos RJ, Vedantam S, Chen H, Florez JC, Fox C, Liu CT, Rybin D, Couper DJ, Kao WH, Li M, Cornelis MC, Kraft P, Sun Q, van Dam RM, Stringham HM, Chines PS, Fischer K, Fontanillas P, Holmen OL, Hunt SE, Jackson AU, Kong A, Lawrence R, Meyer J, Perry JR, Platou CG, Potter S, Rehnberg E, Robertson N, Sivapalaratnam S, Stancakova A, Stirrups K, Thorleifsson G, Tikkanen E, Wood AR, Almgren P, Atalay M, Benediktsson R, Bonnycastle LL, Burtt N, Carey J, Charpentier G, Crenshaw AT, Doney AS, Dorkhan M, Edkins S, Emilsson V, Eury E, Forsen T, Gertow K, Gigante B, Grant GB, Groves CJ, Guiducci C, Herder C, Hreidarsson AB, Hui J, James A, Jonsson A, Rathmann W, Klopp N, Kravic J, Krjutskov K, Langford C, Leander K, Lindholm E, Lobbens S, Mannisto S, Mirza G, Muhleisen TW, Musk B, Parkin M, Rallidis L, Saramies J, Sennblad B, Shah S, Sigurethsson G, Silveira A, Steinbach G, Thorand B, Trakalo J, Veglia F, Wennauer R, Winckler W, Zabaneh D, Campbell H, van Duijn C, Uitterlinden AG, Hofman A, Sijbrands E, Abecasis GR, Owen KR, Zeggini E, Trip MD, Forouhi NG, Syvanen AC, Eriksson JG, Peltonen L, Nothen MM, Balkau B, Palmer CN, Lyssenko V, Tuomi T, Isomaa B, Hunter DJ, Qi L, Wellcome Trust Case Control C, Meta-Analyses of G, Insulin-related traits Consortium I, Genetic Investigation of ATC, Asian Genetic Epidemiology Network-Type 2 Diabetes C, South Asian Type 2 Diabetes C, Shuldiner AR, Roden M, Barroso I, Wilsgaard T, Beilby J, Hovingh K, Price JF, Wilson JF, Rauramaa R, Lakka TA, Lind L, Dedoussis G, Njolstad I, Pedersen NL, Khaw KT, Wareham NJ, Keinanen-Kiukaanniemi SM, Saaristo TE, Korpi-Hyovalti E, Saltevo J, Laakso M, Kuusisto J, Metspalu A, Collins FS, Mohlke KL, Bergman RN, Tuomilehto J, Boehm BO, Gieger C, Hveem K, Cauchi S, Froguel P, Baldassarre D, Tremoli E, Humphries SE, Saleheen D, Danesh J, Ingelsson E, Ripatti S, Salomaa V, Erbel R, Jockel KH, Moebus S, Peters A, Illig T, de Faire U, Hamsten A, Morris AD, Donnelly PJ, Frayling TM, Hattersley AT, Boerwinkle E, Melander O, Kathiresan S, Nilsson PM, Deloukas P, Thorsteinsdottir U, Groop LC, Stefansson K, Hu F, Pankow JS, Dupuis J, Meigs JB, Altshuler D, Boehnke M, McCarthy MI, Replication DIG, Meta-analysis C: Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet 2012;44:981-990
8. Ng MC, Shriner D, Chen BH, Li J, Chen WM, Guo X, Liu J, Bielinski SJ, Yanek LR, Nalls MA, Comeau ME, Rasmussen-Torvik LJ, Jensen RA, Evans DS, Sun YV, An P, Patel SR, Lu Y, Long J, Armstrong LL, Wagenknecht L, Yang L, Snively BM, Palmer ND, Mudgal P, Langefeld CD, Keene KL, Freedman BI, Mychaleckyj JC, Nayak U, Raffel LJ, Goodarzi MO, Chen YD, Taylor HA, Jr., Correa A, Sims M, Couper D, Pankow JS, Boerwinkle E, Adeyemo A, Doumatey A, Chen G, Mathias RA, Vaidya D, Singleton AB, Zonderman AB, Igo RP, Jr., Sedor JR, Consortium F, Kabagambe EK, Siscovick DS, McKnight B, Rice K, Liu Y, Hsueh WC, Zhao W, Bielak LF, Kraja A, Province MA, Bottinger EP, Gottesman O, Cai Q, Zheng W, Blot WJ, Lowe WL, Pacheco JA, Crawford DC, e MC, Consortium D, Grundberg E, Mu TC, Rich SS, Hayes MG, Shu XO, Loos RJ, Borecki IB, Peyser PA, Cummings SR, Psaty BM, Fornage M, Iyengar SK, Evans MK, Becker DM, Kao WH, Wilson JG, Rotter JI, Sale MM, Liu S, Rotimi CN, Bowden DW, Consortium ME-aotDiAA: Meta-analysis of genome-wide association studies in African Americans provides insights into the genetic architecture of type 2 diabetes. PLoS Genet 2014;10:e1004517

9. Kim YJ, Go MJ, Hu C, Hong CB, Kim YK, Lee JY, Hwang JY, Oh JH, Kim DJ, Kim NH, Kim S, Hong EJ, Kim JH, Min H, Kim Y, Zhang R, Jia W, Okada Y, Takahashi A, Kubo M, Tanaka T, Kamatani N, Matsuda K, consortium M, Park T, Oh B, Kimm K, Kang D, Shin C, Cho NH, Kim HL, Han BG, Lee JY, Cho YS: Large-scale genome-wide association studies in East Asians identify new genetic loci influencing metabolic traits. Nat Genet 2011;43:990-995
10. The Wellcome Trust Case Control Consortium: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661-678
11. Flannick J, Florez JC: Type 2 diabetes: genetic data sharing to advance complex disease research. Nat Rev Genet 2016;17:535-549
12. Fuchsberger C, Flannick J, Teslovich TM, Mahajan A, Agarwala V, Gaulton KJ, Ma C, Fontanillas P, Moutsianas L, McCarthy DJ, Rivas MA, Perry JR, Sim X, Blackwell TW, Robertson NR, Rayner NW, Cingolani P, Locke AE, Fernandez Tajes J, Highland HM, Dupuis J, Chines PS, Lindgren CM, Hartl C, Jackson AU, Chen H, Huyghe JR, van de Bunt M, Pearson RD, Kumar A, Muller-Nurasyid M, Grarup N, Stringham HM, Gamazon ER, Lee J, Chen Y, Scott RA, Below JE, Chen P, Huang J, Go MJ, Stitzel ML, Pasko D, Parker SC, Varga TV, Green T, Beer NL, Day-Williams AG, Ferreira T, Fingerlin T, Horikoshi M, Hu C, Huh I, Ikram MK, Kim BJ, Kim Y, Kim YJ, Kwon MS, Lee J, Lee S, Lin KH, Maxwell TJ, Nagai Y, Wang X, Welch RP, Yoon J, Zhang W, Barzilai N, Voight BF, Han BG, Jenkinson CP, Kuulasmaa T, Kuusisto J, Manning A, Ng MC, Palmer ND, Balkau B, Stancakova A, Abboud HE, Boeing H, Giedraitis V, Prabhakaran D, Gottesman O, Scott J, Carey J, Kwan P, Grant G, Smith JD, Neale BM, Purcell S, Butterworth AS, Howson JM, Lee HM, Lu Y, Kwak SH, Zhao W, Danesh J, Lam VK, Park KS, Saleheen D, So WY, Tam CH, Afzal U, Aguilar D, Arya R, Aung T, Chan E, Navarro C, Cheng CY, Palli D, Correa A, Curran JE, Rybin D, Farook VS, Fowler SP, Freedman BI, Griswold M, Hale DE, Hicks PJ, Khor CC, Kumar S, Lehne B, Thuillier D, Lim WY, Liu J, van der Schouw YT, Loh M, Musani SK, Puppala S, Scott WR, Yengo L, Tan ST, Taylor HA, Jr., Thameem F, Wilson G, Sr., Wong TY, Njolstad PR, Levy JC, Mangino M, Bonnycastle LL, Schwarzmayr T, Fadista J, Surdulescu GL, Herder C, Groves CJ, Wieland T, Bork-Jensen J, Brandslund I, Christensen C, Koistinen HA, Doney AS, Kinnunen L, Esko T, Farmer AJ, Hakaste L, Hodgkiss D, Kravic J, Lyssenko V, Hollensted M, Jorgensen ME, Jorgensen T, Ladenvall C, Justesen JM, Karajamaki A, Kriebel J, Rathmann W, Lannfelt L, Lauritzen T, Narisu N, Linneberg A, Melander O, Milani L, Neville M, Orho-Melander M, Qi L, Qi Q, Roden M, Rolandsson O, Swift A, Rosengren AH, Stirrups K, Wood AR, Mihailov E, Blancher C, Carneiro MO, Maguire J, Poplin R, Shakir K, Fennell T, DePristo M, Hrabe de Angelis M, Deloukas P, Gjesing AP, Jun G, Nilsson P, Murphy J, Onofrio R, Thorand B, Hansen T, Meisinger C, Hu FB, Isomaa B, Karpe F, Liang L, Peters A, Huth C, O'Rahilly SP, Palmer CN, Pedersen O, Rauramaa R, Tuomilehto J, Salomaa V, Watanabe RM, Syvanen AC, Bergman RN, Bharadwaj D, Bottinger EP, Cho YS, Chandak GR, Chan JC, Chia KS, Daly MJ, Ebrahim SB, Langenberg C, Elliott P, Jablonski KA, Lehman DM, Jia W, Ma RC, Pollin TI, Sandhu M, Tandon N, Froguel P, Barroso I, Teo YY, Zeggini E, Loos RJ, Small KS, Ried JS, DeFronzo RA, Grallert H, Glaser B, Metspalu A, Wareham NJ, Walker M, Banks E, Gieger C, Ingelsson E, Im HK, Illig T, Franks PW, Buck G, Trakalo J, Buck D, Prokopenko I, Magi R, Lind L, Farjoun Y, Owen KR, Gloyn AL, Strauch K, Tuomi T, Kooner JS, Lee JY, Park T, Donnelly P, Morris AD, Hattersley AT, Bowden DW, Collins FS, Atzmon G, Chambers JC, Spector TD, Laakso M, Strom TM, Bell GI, Blangero J, Duggirala R, Tai ES, McVean G, Hanis CL, Wilson JG, Seielstad M, Frayling TM, Meigs JB, Cox NJ, Sladek R, Lander ES, Gabriel S, Burtt NP, Mohlke KL, Meitinger T, Groop L, Abecasis G, Florez JC, Scott LJ, Morris AP, Kang HM, Boehnke M, Altshuler D, McCarthy MI: The genetic architecture of type 2 diabetes. Nature 2016;536:41-47
13. Plenge RM, Scolnick EM, Altshuler D: Validating therapeutic targets through human genetics. Nat Rev Drug Discov 2013;12:581-594
14. Estrada K, Aukrust I, Bjorkhaug L, Burtt NP, Mercader JM, Garcia-Ortiz H, Huerta-Chagoya A, Moreno-Macias H, Walford G, Flannick J, Williams AL, Gomez-Vazquez MJ, Fernandez-Lopez JC, Martinez-Hernandez A, Centeno-Cruz F, Mendoza-Caamal E, Revilla-Monsalve C, Islas-

Andrade S, Cordova EJ, Soberon X, Gonzalez-Villalpando ME, Henderson E, Wilkens LR, Le Marchand L, Arellano-Campos O, Ordonez-Sanchez ML, Rodriguez-Torres M, Rodriguez-Guillen R, Riba L, Najmi LA, Jacobs SB, Fennell T, Gabriel S, Fontanillas P, Hanis CL, Lehman DM, Jenkinson CP, Abboud HE, Bell GI, Cortes ML, Boehnke M, Gonzalez-Villalpando C, Orozco L, Haiman CA, Tusie-Luna T, Aguilar-Salinas CA, Altshuler D, Njolstad PR, Florez JC, MacArthur DG: Association of a low-frequency variant in HNF1A with type 2 diabetes in a Latino population. JAMA 2014;311:2305-2314

15. Hara K, Fujita H, Johnson TA, Yamauchi T, Yasuda K, Horikoshi M, Peng C, Hu C, Ma RC, Imamura M, Iwata M, Tsunoda T, Morizono T, Shojima N, So WY, Leung TF, Kwan P, Zhang R, Wang J, Yu W, Maegawa H, Hirose H, consortium D, Kaku K, Ito C, Watada H, Tanaka Y, Tobe K, Kashiwagi A, Kawamori R, Jia W, Chan JC, Teo YY, Shyong TE, Kamatani N, Kubo M, Maeda S, Kadowaki T: Genome-wide association study identifies three novel loci for type 2 diabetes. Human molecular genetics 2014;23:239-246

16. Moltke I, Grarup N, Jorgensen ME, Bjerregaard P, Treebak JT, Fumagalli M, Korneliussen TS, Andersen MA, Nielsen TS, Krarup NT, Gjesing AP, Zierath JR, Linneberg A, Wu X, Sun G, Jin X, Al-Aama J, Wang J, Borch-Johnsen K, Pedersen O, Nielsen R, Albrechtsen A, Hansen T: A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. Nature 2014;512:190-193

17. Prevention CfDCa: National diabetes fact sheet: National estimates and general information on diabetes and prediabetes in the United States. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention 2011;

18. Villalpando S, de la Cruz V, Rojas R, Shamah-Levy T, Avila MA, Gaona B, Rebollar R, Hernandez L: Prevalence and distribution of type 2 diabetes mellitus in Mexican adult population: a probabilistic survey. Salud publica de Mexico 2010;52 Suppl 1:S19-26

19. Sigma Type 2 Diabetes Consortium, Estrada K, Aukrust I, Bjorkhaug L, Burtt NP, Mercader JM, Garcia-Ortiz H, Huerta-Chagoya A, Moreno-Macias H, Walford G, Flannick J, Williams AL, Gomez-Vazquez MJ, Fernandez-Lopez JC, Martinez-Hernandez A, Jimenez-Morales S, Centeno-Cruz F, Mendoza-Caamal E, Revilla-Monsalve C, Islas-Andrade S, Cordova EJ, Soberon X, Gonzalez-Villalpando ME, Henderson E, Wilkens LR, Le Marchand L, Arellano-Campos O, Ordonez-Sanchez ML, Rodriguez-Torres M, Rodriguez-Guillen R, Riba L, Najmi LA, Jacobs SB, Fennell T, Gabriel S, Fontanillas P, Hanis CL, Lehman DM, Jenkinson CP, Abboud HE, Bell GI, Cortes ML, Boehnke M, Gonzalez-Villalpando C, Orozco L, Haiman CA, Tusie-Luna T, Aguilar-Salinas CA, Altshuler D, Njolstad PR, Florez JC, MacArthur DG: Association of a low-frequency variant in HNF1A with type 2 diabetes in a Latino population. JAMA 2014;311:2305-2314

20. Cebola I, Rodriguez-Segui SA, Cho CH, Bessa J, Rovira M, Luengo M, Chhatriwala M, Berry A, Ponsa-Cobas J, Maestro MA, Jennings RE, Pasquali L, Moran I, Castro N, Hanley NA, Gomez-Skarmeta JL, Vallier L, Ferrer J: TEAD and YAP regulate the enhancer network of human embryonic pancreatic progenitors. Nature cell biology 2015;17:615-626

21. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR, Farnham PJ, Hirst M, Lander ES, Mikkelsen TS, Thomson JA: The NIH Roadmap Epigenomics Mapping Consortium. Nature biotechnology 2010;28:1045-1048

22. Moran I, Akerman I, van de Bunt M, Xie R, Benazra M, Nammo T, Arnes L, Nakic N, Garcia-Hurtado J, Rodriguez-Segui S, Pasquali L, Sauty-Colace C, Beucher A, Scharfmann R, van Arensbergen J, Johnson PR, Berry A, Lee C, Harkins T, Gmyr V, Pattou F, Kerr-Conte J, Piemonti L, Berney T, Hanley N, Gloyn AL, Sussel L, Langman L, Brayman KL, Sander M, McCarthy MI, Ravassard P, Ferrer J: Human beta cell transcriptome analysis uncovers lncRNAs that are tissue-specific, dynamically regulated, and abnormally expressed in type 2 diabetes. Cell Metab 2012;16:435-448

23. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR: STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15-21

24. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature methods 2008;5:621-628
25. The GTEx Consortium: Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 2015;348:648-660
26. DeLuca DS, Levin JZ, Sivachenko A, Fennell T, Nazaire MD, Williams C, Reich M, Winckler W, Getz G: RNA-SeQC: RNA-seq metrics for quality control and process optimization. Bioinformatics 2012;28:1530-1532
27. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, Sabatti C, Eskin E: Variance component model to account for sample structure in genome-wide association studies. Nat Genet 2010;42:348-354
28. Delaneau O, Marchini J, Zagury JF: A linear complexity phasing method for thousands of genomes. Nature methods 2012;9:179-181
29. Willer CJ, Li Y, Abecasis GR: METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26:2190-2191
30. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M: A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 2007;316:1341-1345
31. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, Zeggini E, Huth C, Aulchenko YS, Thorleifsson G, McCulloch LJ, Ferreira T, Grallert H, Amin N, Wu G, Willer CJ, Raychaudhuri S, McCarroll SA, Langenberg C, Hofmann OM, Dupuis J, Qi L, Segre AV, van Hoek M, Navarro P, Ardlie K, Balkau B, Benediktsson R, Bennett AJ, Blagieva R, Boerwinkle E, Bonnycastle LL, Bengtsson Bostrom K, Bravenboer B, Bumpstead S, Burtt NP, Charpentier G, Chines PS, Cornelis M, Couper DJ, Crawford G, Doney AS, Elliott KS, Elliott AL, Erdos MR, Fox CS, Franklin CS, Ganser M, Gieger C, Grarup N, Green T, Griffin S, Groves CJ, Guiducci C, Hadjadj S, Hassanali N, Herder C, Isomaa B, Jackson AU, Johnson PR, Jorgensen T, Kao WH, Klopp N, Kong A, Kraft P, Kuusisto J, Lauritzen T, Li M, Lieverse A, Lindgren CM, Lyssenko V, Marre M, Meitinger T, Midthjell K, Morken MA, Narisu N, Nilsson P, Owen KR, Payne F, Perry JR, Petersen AK, Platou C, Proenca C, Prokopenko I, Rathmann W, Rayner NW, Robertson NR, Rocheleau G, Roden M, Sampson MJ, Saxena R, Shields BM, Shrader P, Sigurdsson G, Sparso T, Strassburger K, Stringham HM, Sun Q, Swift AJ, Thorand B, Tichet J, Tuomi T, van Dam RM, van Haeften TW, van Herpt T, van Vliet-Ostaptchouk JV, Walters GB, Weedon MN, Wijmenga C, Witteman J, Bergman RN, Cauchi S, Collins FS, Gloyn AL, Gyllensten U, Hansen T, Hide WA, Hitman GA, Hofman A, Hunter DJ, Hveem K, Laakso M, Mohlke KL, Morris AD, Palmer CN, Pramstaller PP, Rudan I, Sijbrands E, Stein LD, Tuomilehto J, Uitterlinden A, Walker M, Wareham NJ, Watanabe RM, Abecasis GR, Boehm BO, Campbell H, Daly MJ, Hattersley AT, Hu FB, Meigs JB, Pankow JS, Pedersen O, Wichmann HE, Barroso I, Florez JC, Frayling TM, Groop L, Sladek R, Thorsteinsdottir U, Wilson JF, Illig T, Froguel P, van Duijn CM, Stefansson K, Altshuler D, Boehnke M, McCarthy MI, investigators M, Consortium G: Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet 2010;42:579-589
32. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won HH, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M,

McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG, Exome Aggregation C: Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;536:285-291
33. Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB: Rare variants create synthetic genome-wide associations. PLoS Biol 8:e1000294
34. Yasuda K, Miyake K, Horikawa Y, Hara K, Osawa H, Furuta H, Hirota Y, Mori H, Jonsson A, Sato Y, Yamagata K, Hinokio Y, Wang HY, Tanahashi T, Nakamura N, Oka Y, Iwasaki N, Iwamoto Y, Yamada Y, Seino Y, Maegawa H, Kashiwagi A, Takeda J, Maeda E, Shin HD, Cho YM, Park KS, Lee HK, Ng MC, Ma RC, So WY, Chan JC, Lyssenko V, Tuomi T, Nilsson P, Groop L, Kamatani N, Sekine A, Nakamura Y, Yamamoto K, Yoshida T, Tokunaga K, Itakura M, Makino H, Nanjo K, Kadowaki T, Kasuga M: Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. Nat Genet 2008;40:1092-1097
35. Hanson RL, Rong R, Kobes S, Muller YL, Weil EJ, Curtis JM, Nelson RG, Baier LJ: Role of Established Type 2 Diabetes-Susceptibility Genetic Variants in a High Prevalence American Indian Population. Diabetes 2015;64:2646-2657
36. Kvale MN, Hesselson S, Hoffmann TJ, Cao Y, Chan D, Connell S, Croen LA, Dispensa BP, Eshragh J, Finn A, Gollub J, Iribarren C, Jorgenson E, Kushi LH, Lao R, Lu Y, Ludwig D, Mathauda GK, McGuire WB, Mei G, Miles S, Mittman M, Patil M, Quesenberry CP, Jr., Ranatunga D, Rowell S, Sadler M, Sakoda LC, Shapero M, Shen L, Shenoy T, Smethurst D, Somkin CP, Van Den Eeden SK, Walter L, Wan E, Webster T, Whitmer RA, Wong S, Zau C, Zhan Y, Schaefer C, Kwok PY, Risch N: Genotyping Informatics and Quality Control for 100,000 Subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort. Genetics 2015;200:1051-1060
37. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, Nathan DM: Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. The New England journal of medicine 2002;346:393-403
38. Mitchell BD, Kammerer CM, Blangero J, Mahaney MC, Rainwater DL, Dyke B, Hixson JE, Henkel RD, Sharp RM, Comuzzie AG, VandeBerg JL, Stern MP, MacCluer JW: Genetic and environmental contributions to cardiovascular risk factors in Mexican Americans. The San Antonio Family Heart Study. Circulation 1996;94:2159-2170
39. Hunt KJ, Lehman DM, Arya R, Fowler S, Leach RJ, Goring HHH, Almasy L, Blangero J, Dyer TD, Duggirala R, Stern MP: Genome-wide linkage analyses of type 2 diabetes in Mexican Americans - The San Antonio Family Diabetes/Gallbladder Study. Diabetes 2005;54:2655-2662
40. Coletta DK, Schneider J, Hu SL, Dyer TD, Puppala S, Farook VS, Arya R, Lehman DM, Blangero J, DeFronzo RA, Duggirala R, Jenkinson CP: Genome-wide linkage scan for genes influencing plasma triglyceride levels in the Veterans Administration Genetic Epidemiology Study. Diabetes 2009;58:279-284
41. Harper AR, Nayee S, Topol EJ: Protective alleles and modifier variants in human health and disease. Nat Rev Genet 2015;16:689-701
42. Livingstone C, Borai A: Insulin-like growth factor-II: its role in metabolic and endocrine disease. Clin Endocrinol (Oxf) 2014;80:773-781
43. Morali OG, Jouneau A, McLaughlin KJ, Thiery JP, Larue L: IGF-II promotes mesoderm formation. Dev Biol 2000;227:133-145
44. Hill DJ, Strutt B, Arany E, Zaina S, Coukell S, Graham CF: Increased and persistent circulating insulin-like growth factor II in neonatal transgenic mice suppresses developmental apoptosis in the pancreatic islets. Endocrinology 2000;141:1151-1157
45. Sparago A, Cerrato F, Vernucci M, Ferrero GB, Silengo MC, Riccio A: Microdeletions in the human H19 DMR result in loss of IGF2 imprinting and Beckwith-Wiedemann syndrome. Nat Genet 2004;36:958-960

46. Frystyk J, Skjaerbaek C, Vestbo E, Fisker S, Orskov H: Circulating levels of free insulin-like growth factors in obese subjects: the impact of type 2 diabetes. Diabetes Metab Res Rev 1999;15:314-322
47. Estil les E, Tellez N, Soler J, Montanya E: High sensitivity of beta-cell replication to the inhibitory effects of interleukin-1beta: modulation by adenoviral overexpression of IGF2 in rat islets. J Endocrinol 2009;203:55-63
48. Poher AL, Altirriba J, Veyrat-Durebex C, Rohner-Jeanrenaud F: Brown adipose tissue activity as a target for the treatment of obesity/insulin resistance. Frontiers in physiology 2015;6:4
49. Devedjian JC, George M, Casellas A, Pujol A, Visa J, Pelegrin M, Gros L, Bosch F: Transgenic mice overexpressing insulin-like growth factor-II in beta cells develop type 2 diabetes. J Clin Invest 2000;105:731-740
50. Casellas A, Mallol C, Salavert A, Jimenez V, Garcia M, Agudo J, Obach M, Haurigot V, Vila L, Molas M, Lage R, Morro M, Casana E, Ruberte J, Bosch F: Insulin-like Growth Factor 2 Overexpression Induces Beta-cell Dysfunction and Increases Beta-cell Susceptibility to Damage. J Biol Chem 2015;
51. Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JR, Rayner NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ, Harries LW, Marchini JL, Owen KR, Knight B, Cardon LR, Walker M, Hitman GA, Morris AD, Doney AS, McCarthy MI, Hattersley AT: Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science 2007;316:1336-1341
52. Diabetes Genetics Initiative of Broad Institute of H, Mit LU, Novartis Institutes of BioMedical R, Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, Hughes TE, Groop L, Altshuler D, Almgren P, Florez JC, Meyer J, Ardlie K, Bengtsson Bostrom K, Isomaa B, Lettre G, Lindblad U, Lyon HN, Melander O, Newton-Cheh C, Nilsson P, Orho-Melander M, Rastam L, Speliotes EK, Taskinen MR, Tuomi T, Guiducci C, Berglund A, Carlson J, Gianniny L, Hackett R, Hall L, Holmkvist J, Laurila E, Sjogren M, Sterner M, Surti A, Svensson M, Svensson M, Tewhey R, Blumenstiel B, Parkin M, Defelice M, Barry R, Brodeur W, Camarata J, Chia N, Fava M, Gibbons J, Handsaker B, Healy C, Nguyen K, Gates C, Sougnez C, Gage D, Nizzari M, Gabriel SB, Chirn GW, Ma Q, Parikh H, Richardson D, Ricke D, Purcell S: Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 2007;316:1331-1336
53. Dai N, Zhao L, Wrighting D, Kramer D, Majithia A, Wang Y, Cracan V, Borges-Rivera D, Mootha VK, Nahrendorf M, Thorburn DR, Minichiello L, Altshuler D, Avruch J: IGF2BP2/IMP2-Deficient mice resist obesity through enhanced translation of Ucp1 mRNA and Other mRNAs encoding mitochondrial proteins. Cell Metab 2015;21:609-621
54. Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooper R, Kwiatkowski D, Ward R, Lander ES: Detecting recent positive selection in the human genome from haplotype structure. Nature 2002;419:832-837
55. Zinman B, Wanner C, Lachin JM, Fitchett D, Bluhmki E, Hantel S, Mattheus M, Devins T, Johansen OE, Woerle HJ, Broedl UC, Inzucchi SE, Investigators E-RO: Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes. The New England journal of medicine 2015;373:2117-2128
56. Santer R, Kinner M, Lassen CL, Schneppenheim R, Eggert P, Bald M, Brodehl J, Daschner M, Ehrich JH, Kemper M, Li Volti S, Neuhaus T, Skovby F, Swift PG, Schaub J, Klaerke D: Molecular analysis of the SGLT2 gene in patients with renal glucosuria. Journal of the American Society of Nephrology : JASN 2003;14:2873-2882
57. Scotti MM, Swanson MS: RNA mis-splicing in disease. Nat Rev Genet 2016;17:19-32
58. Kole R, Krieg AM: Exon skipping therapy for Duchenne muscular dystrophy. Adv Drug Deliv Rev 2015;87:104-107

IGF2 loss-of-function T2D protective variant

For Peer Review Only

Table 1. Study cohorts comprising the SIGMA type 2 diabetes exome chip project data set

Study	Sample location	Study design		n	Percent male	Age (years)	Age-of-onset (years)	BMI (kg/m^2)	Fasting plasma glucose (mmol/l)
UNAM/INCMNSZ Diabetes Study (UIDS)	Mexico City, Mexico	Prospective cohort	Controls	1164	41.3	55.4 (9.4)	-	28.2 (3.9)	4.8 (0.5)
			T2D cases	835	40.1	56.3 (12.4)	44.2 (11.4)	28.6 (4.6)	9.8 (4.5)
Diabetes in Mexico Study (DMS)	Mexico City, Mexico	Prospective cohort	Controls	486	25.3	52.6 (7.8)	-	28.1 (4.5)	5 (0.4)
			T2D cases	715	32.3	55.9 (11)	47.7 (10.4)	29 (5.6)	8.8 (3.9)
Mexico City Diabetes Study (MCDS)	Mexico City, Mexico	Prospective cohort	Controls	671	38.3	62.2 (7.7)	-	29.4 (4.6)	5 (0.6)
			T2D cases	315	40.3	63.9 (7.5)	54.7 (9.7)	30 (5.3)	8.8 (4)
Multiethnic Cohort (MEC)	Los Angeles, California, USA	Case-control	Controls	2285	48.5	59.2 (7)	-	26.6 (3.9)	-
			T2D cases	2187	47.6	59.1 (6.9)	-	29.9 (5.3)	-

The table shows sample location, study design, numbers of cases and controls, percent male participants, age, age-of-onset in cases, body mass index, and fasting plasma glucose in controls. Data are mean (SD).

Figure legends

Figure 1. Discovery and replication of the rs149483638 T2D protective variant. QQ plot for all common variants (**a**) and for Mexican population-specific variants (**b**). The plot shows the two most significant variants that have low frequency in Europeans, but higher frequency in the Mexican population. (**c**) Forest plot for the meta-analysis of rs149483638 variant in *IGF2*. We replicated the rs149483638 association in four independent data sets: 1,007 T2D cases and 917 controls of Hispanic origin from the T2D-GENES project (minor allele frequency [MAF]=0.12, odds ratio [OR]=0.98, p=0.3), 1,519 T2D cases and 1,680 controls of full-heritage American Indian ancestry from the Pima cohort (MAF=0.14, OR=0.68, $p=10^{-6}$), 427 cases and 751 controls of self-identified indigenous individuals from different ethnic groups in Mexico (DMS2 cohort) (MAF=0.36, OR=0.71, p=0.001) and 1,064 cases and 4,832 controls from the subset of cases from Latino ancestry (MAF=0.06, OR=0.82, p=0.11).

Figure 2. rs149483638 prevents splicing *in vitro*. (**a**) This variant is located at a canonical splice acceptor site, and is predicted to cause skipping of exon 2 of *IGF2* isoform 2. (**b**) 293T cells were transfected with *IGF2* minigenes containing the first three exons and two introns of the *IGF2* gene, and either allele of the rs149483638 C>T variant (G>A in the reverse strand) and cDNA was analyzed by droplet digital (ddPCR). This analysis revealed no expression of the *IGF2* exon 1-2 junction in cells transfected with the minigene containing the T2D-protective rs149483638 A allele. This was in contrast to the high levels of exon 1-2 splicing detected in cells transfected with the G allele. (**c**) One-dimensional plots of the ddPCR droplets plotted in (b). No *IGF2* transcript was detected in untransfected samples. *ACTB* was used as an internal control.

For Peer Review Only

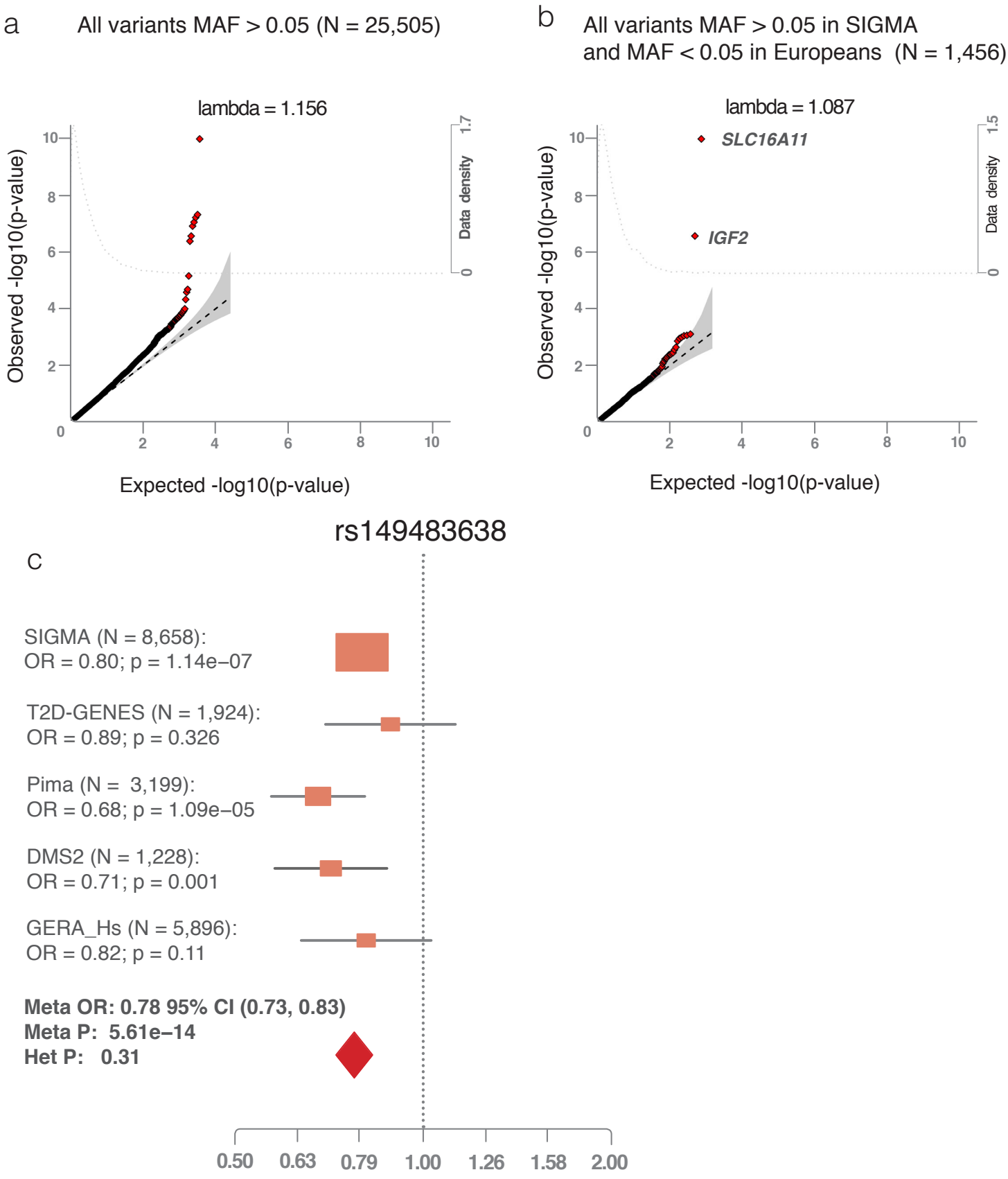
Figure 3. rs149483638 prevents splicing between exon 2 in liver and in adipose tissue. The dosage of the T2D protective A allele is correlated with lower expression of *IGF2* isoform 2 (as measured by expression levels of the exon 1-2 junction) in liver (n(GG) = 21, n(GA) = 9, n(AA) = 4) (**a**) and in adipose tissue (n(GG) = 83, n(GA) = 43, n(AA) = 5) (**b**).

Figure 4. Expression of *IGF2* isoform 2 with T2D and related. (a) Boxplots representing the expression of *IGF2* isoform 2 across T2D cases and controls in individuals homozygous for the G common allele. The linear model p-value represents the association between IGF2 isoform 2 expression, adjusted by age, body mass index, and sex. (**b**) *IGF2* isoform 2 positively correlates with higher plasma glycated hemoglobin (HbA1c) in non-diabetic participants. The grey area limited by the dashed red lines represent the 95% confidence interval of the slope of the linear regression. *Patients with HbA1c above 6.5% were non-T2D subjects according to the diagnostic criteria of Mexico at the time of extraction, as HbA1c was not considered a criterion in Mexico at the time of extraction. Therefore, none of the subjects were receiving any lowering glucose treatment. **For clarity, since the genotype is strongly associated with isoform 2 expression, only individuals carrying the GG genotype are plotted in (**a**) and (**b**).

Figure 5. Phenome-wide analysis of rs149483638 variant. The protective variant was tested for association across 18 different disease traits previously categorized in the subsample of GERA cohort of Latino ancestry (5,896 individuals). While the rs149483638 variant was associated with reduced risk of T2D, there was no significant association seen for other 18 conditions. IBS: Irritable bowel syndrome; Mac. Degen.: Macular degeneration; Psychiatric: any psychiatric condition; PVD: Peripheral vascular disease; Stress: acute reaction to stress. Association analyses were done by logistic regression analyses, considering additive model, and correcting for age, BMI, sex, and the first two principal components to correct for population stratification.

For Peer Review Only

Figure 1


a
All variants MAF > 0.05 (N = 25,505)
lambda = 1.156
Observed -log10(p-value)
Expected -log10(p-value)
Data density
b
All variants MAF > 0.05 in SIGMA
and MAF < 0.05 in Europeans (N = 1,456)
lambda = 1.087
SLC16A11
IGF2
Observed -log10(p-value)
Expected -log10(p-value)
Data density
rs149483638
c
SIGMA (N = 8,658):
OR = 0.80; p = 1.14e–07
T2D-GENES (N = 1,924):
OR = 0.89; p = 0.326
Pima (N = 3,199):
OR = 0.68; p = 1.09e–05
DMS2 (N = 1,228):
OR = 0.71; p = 0.001
GERA_Hs (N = 5,896):
OR = 0.82; p = 0.11
Meta OR: 0.78 95% CI (0.73, 0.83)
Meta P: 5.61e–14
Het P: 0.31
0.50 0.63 0.79 1.00 1.26 1.58 2.00

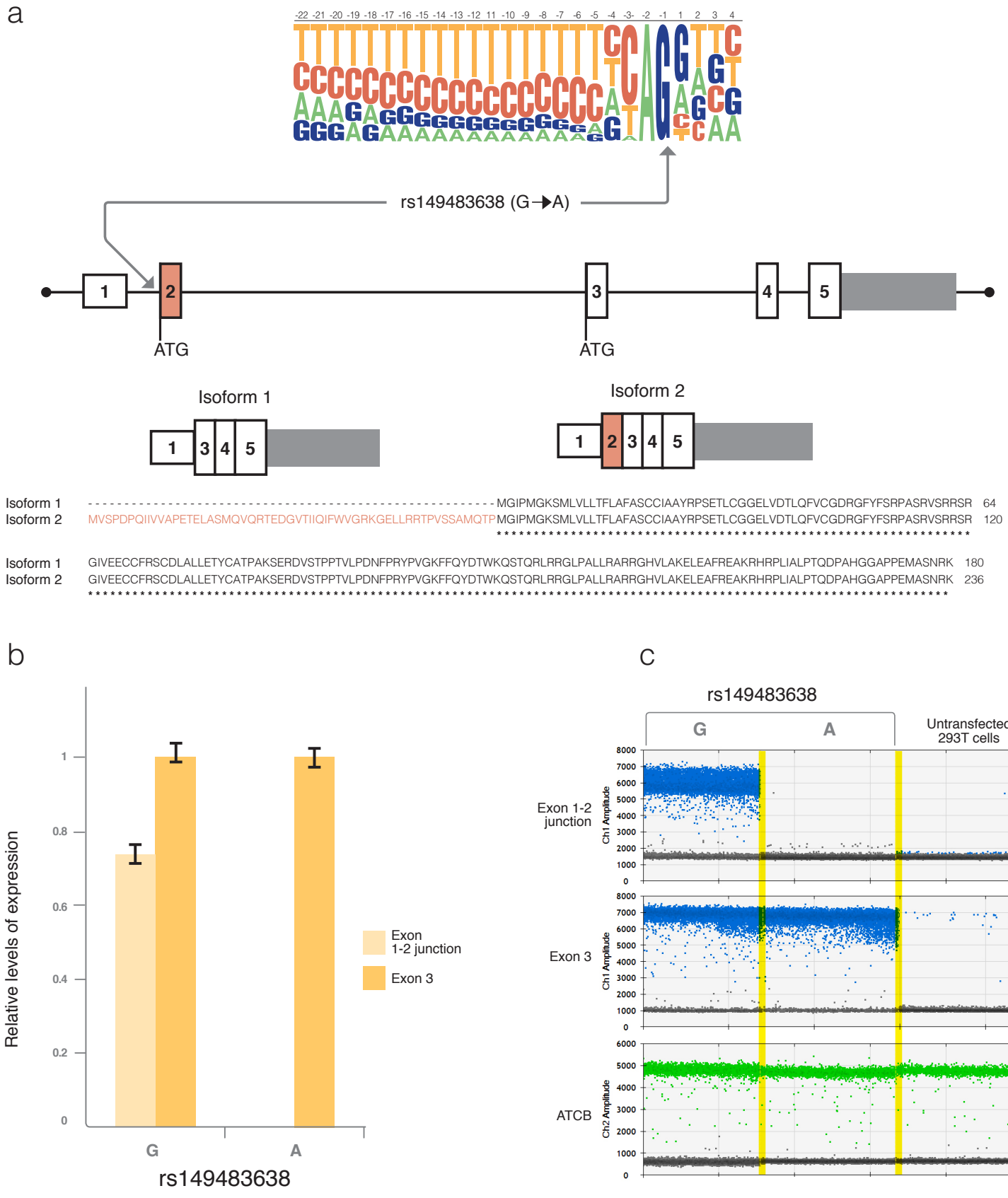
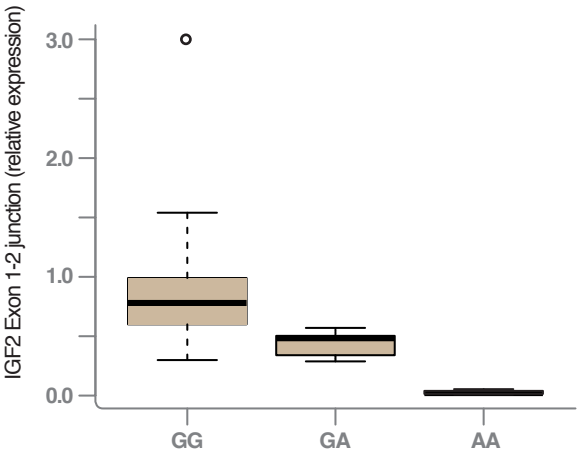

a
rs149483638 (G→A)
ATG
ATG
Isoform 1
Isoform 2
Isoform 1 MGIPMGKSMLVLLTFLAFASCCIAAYRPSETLCGGELVDTLQFVCGDRGFYFSRPASRVSRRSR 64
Isoform 2 MVSPDPQIIVVAPETELASMQVQRTEDGVTIIQIFWVGRKGELLRRTPVSSAMQTPMGIPMGKSMLVLLTFLAFASCCIAAYRPSETLCGGELVDTLQFVCGDRGFYFSRPASRVSRRSR 120
Isoform 1 GIVEECCFRSCDLALLETYCATPAKSERDVSTPPTVLPDNFPRYPVGKFFQYDTWKQSTQRLRRGLPALLRARRGHVLAKELEAFREAKRHRPLIALPTQDPAHGGAPPEMASNRK 180
Isoform 2 GIVEECCFRSCDLALLETYCATPAKSERDVSTPPTVLPDNFPRYPVGKFFQYDTWKQSTQRLRRGLPALLRARRGHVLAKELEAFREAKRHRPLIALPTQDPAHGGAPPEMASNRK 236
b
Relative levels of expression
Exon 1-2 junction
Exon 3
G
A
rs149483638
c
rs149483638
G
A
Untransfected 293T cells
Exon 1-2 junction
Exon 3
ATCB
Ch1 Amplitude
Ch2 Amplitude

Figure 3

a

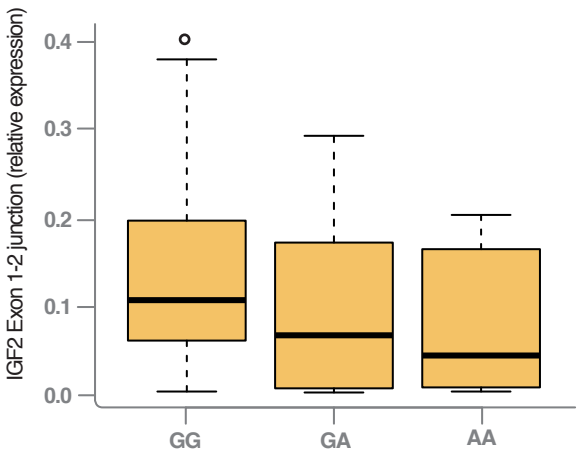
Liver (n = 34)
rho = –0.75; spearman P-value = 3e–07



rs149483638

b

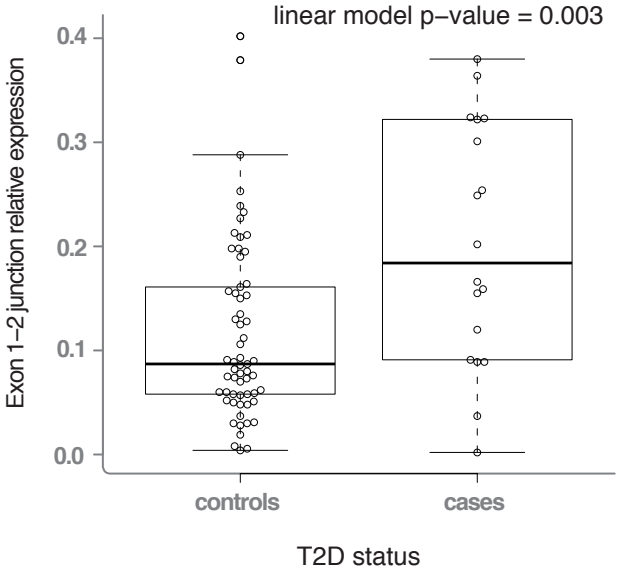
Adipose (n = 131)
rho = –0.22; spearman P-value = 0.011



rs149483638

a

Exon 1–2 junction in adipose tissue by Type 2 Diabetes status



b

HbA1c vs exon 1–2 junction adipose tissue expression in controls GG; (n = 46)

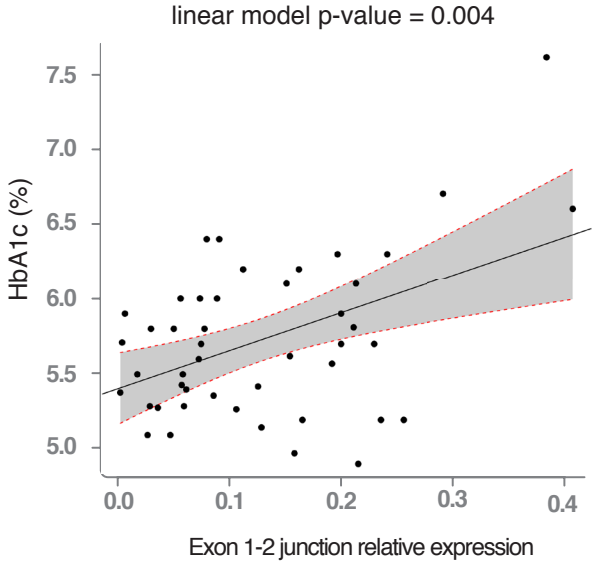
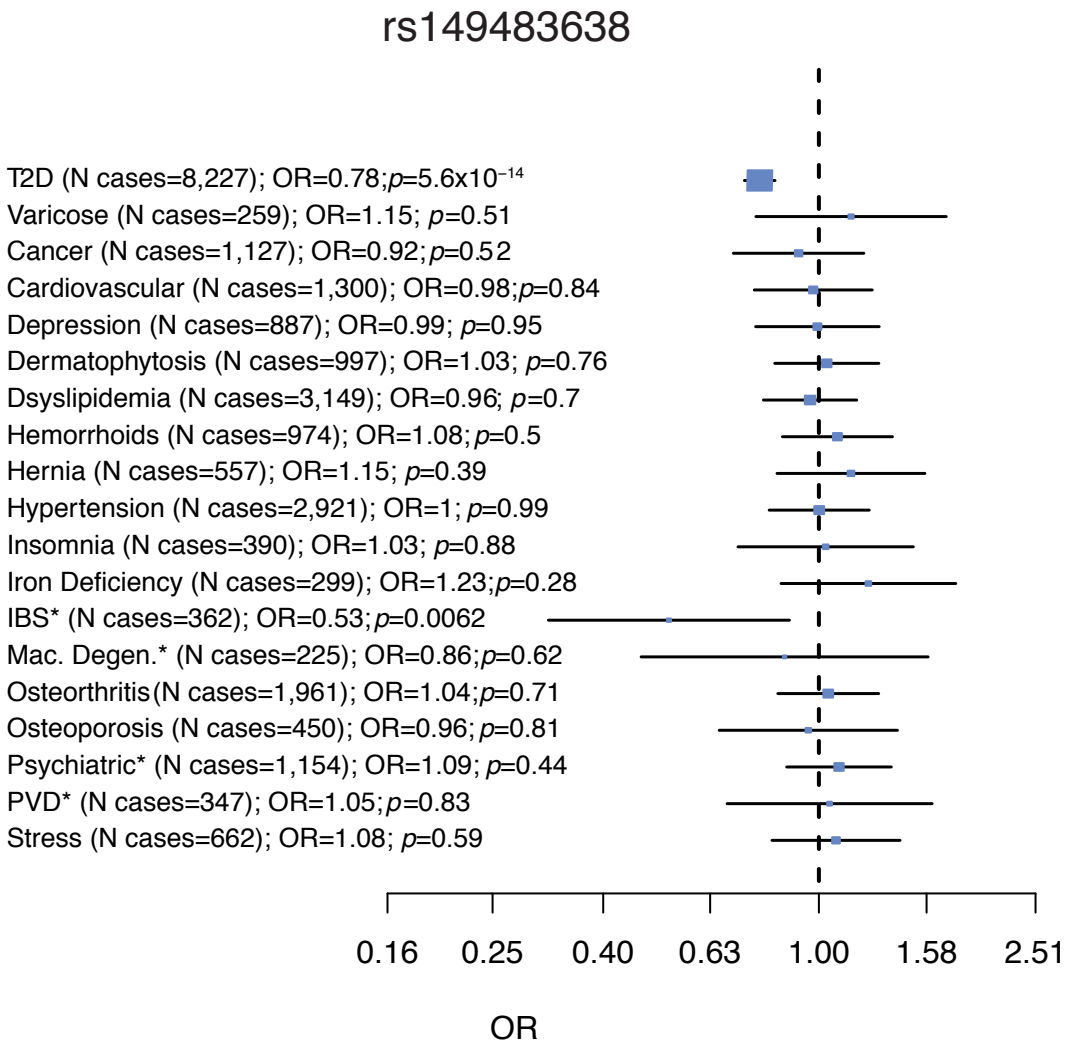


Figure 5


rs149483638
T2D (N cases=8,227); OR=0.78;p=5.6x10⁻¹⁴
Varicose (N cases=259); OR=1.15; p=0.51
Cancer (N cases=1,127); OR=0.92;p=0.52
Cardiovascular (N cases=1,300); OR=0.98;p=0.84
Depression (N cases=887); OR=0.99; p=0.95
Dermatophytosis (N cases=997); OR=1.03; p=0.76
Dsyslipidemia (N cases=3,149); OR=0.96; p=0.7
Hemorrhoids (N cases=974); OR=1.08;p=0.5
Hernia (N cases=557); OR=1.15; p=0.39
Hypertension (N cases=2,921); OR=1; p=0.99
Insomnia (N cases=390); OR=1.03; p=0.88
Iron Deficiency (N cases=299); OR=1.23;p=0.28
IBS* (N cases=362); OR=0.53;p=0.0062
Mac. Degen.* (N cases=225); OR=0.86;p=0.62
Osteorthritis(N cases=1,961); OR=1.04;p=0.71
Osteoporosis (N cases=450); OR=0.96;p=0.81
Psychiatric* (N cases=1,154); OR=1.09; p=0.44
PVD* (N cases=347); OR=1.05;p=0.83
Stress (N cases=662); OR=1.08; p=0.59
0.16
0.25
0.40
0.63
1.00
1.58
2.51
OR

ONLINE SUPPLEMENTARY MATERIAL

A loss-of-function splice acceptor variant in *IGF2* is protective for type 2 diabetes

Running Title: *IGF2* loss-of-function type 2 diabetes protective variant.

The SIGMA T2D Genetics Consortium

Josep M Mercader[1,2,3], Rachel G. Liao[1]*, Avery Davis[4,5,6]*, Zachary Dymek[1]*, Karol Estrada[1,7,8], Taru Tukiainen[4,6,7], Alicia Huerta-Chagoya[9], Hortensia Moreno-Macías[9,10], Kathleen A. Jablonski[11], Robert L. Hanson[12], Geoffrey A. Walford[1,2,8], Ignasi Moran[13], Ling Chen[1,2], Vineeta Agarwala[6], María Luisa Ordoñez-Sánchez[9], Rosario Rodríguez-Guillen[9], Maribel Rodríguez-Torres[9], Yayoi Segura-Kato[9], Humberto García-Ortiz[14], Federico Centeno-Cruz[14], Francisco Barajas-Olmos[14], Lizz Caulkins[1], Sobha Puppala[15], Pierre Fontanillas[6], Amy Williams[16], Sílvia Bonàs-Guarch[3], Chris Hartl[6], Stephan Ripke[5,7,17], Diabetes Prevention Program Research Group[¢], Katherine Tooley[4,5,6], Jacqueline Lane[6,18,19], Carlos Zerrweck[20], Angélica Martínez-Hernández[14], Emilio J. Córdova[14], Elvia Mendoza-Caamal[14], Cecilia Contreras-Cubas[14], María E. González-Villalpando[21], Ivette Cruz-Bautista[22], Liliana Muñoz-Hernández[22], Donaji Gómez-Velasco[22], Ulises Alvirde[22], Brian E. Henderson[23], Lynne R. Wilkens[24], Loic Le Marchand[24], Olimpia Arellano-Campos[22], Laura Riba[22], Maegan Harden[25], Broad Genomics Platform[25], Stacey Gabriel[25], T2D-GENES Consortium[¢], Hanna E. Abboud[26], Maria L. Cortes[27], Cristina Revilla-Monsalve[28], Sergio Islas-Andrade[28], Xavier Soberon[14], Joanne E. Curran[29], Christopher P. Jenkinson[30], Ralph A. DeFronzo[31], Donna M. Lehman[32], Craig L. Hanis[33], Graeme I. Bell[34,35], Michael Boehnke[36], John Blangero[29], Ravindranath Duggirala[30], Richa Saxena[6,18,19], Daniel MacArthur[6,7,8], Jorge Ferrer[13,37,38], Steven A. McCarroll[4,5,6], David Torrents[3,39], William C. Knowler[12], Leslie J. Baier[12], Noel Burtt[1], Clicerio González-Villalpando[21], Christopher A. Haiman[24], Carlos A. Aguilar-Salinas[22], Teresa Tusié-Luna[9], Jason Flannick[1,2,40], Suzanne B.R. Jacobs[1,2], Lorena Orozco[14], David Altshuler[2,4,6,8,18,40,41], Jose C. Florez[1,2,8,#]

[¢]Members of the consortia are provided in Appendix S1.

*These authors contributed equally to this work.

#To whom correspondence should be addressed.

1. Programs in Metabolism and Medical & Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA.
2. Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA.
3. Barcelona Supercomputing Center (BSC). Joint BSC-CRG-IRB Research Program in Computational Biology, 08034 Barcelona, Spain.
4. Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA.

5. Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.
6. Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA.
7. Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.
8. Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA.
9. Unidad de Biología Molecular y Medicina Genómica, I.d.I.B., UNAM/ Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán. Mexico City, México. Instituto de Investigaciones Biomédicas, UNAM Unidad de Biología Molecular y Medicina Genómica, UNAM/INCMNSZ, Coyoacán, 04510 Mexico City, Mexico.
10. Universidad Autónoma Metropolitana, Tlalpan 14387, Mexico City, Mexico.
11. The Biostatistics Center, George Washington University, Rockville, MD, 20852, USA.
12. Phoenix Epidemiology and Clinical Research Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Phoenix, Arizona, 85004, USA.
13. Department of Medicine, Imperial College London, London W12 0NN, United Kingdom.
14. Instituto Nacional de Medicina Genómica, Tlalpan, 14610, Mexico City, Mexico.
15. Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX, USA.
16. Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, New York, USA.
17. Department of Psychiatry and Psychotherapy, Charité–Universitätsmedizin Berlin, Campus Mitte, 10117 Berlin, Germany.
18. Center for Genomic Medicne, Massachusetts General Hospital, Boston, Massachusetts, USA.
19. Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA.
20. Clínica de Integral de Cirugía para la Obesidad y Enfermedades Metabólicas, Hospital General Tláhuac, Secretaría de Salud del CDMX. México City.
21. Centro de Estudios en Diabetes, Unidad de Investigacion en Diabetes y Riesgo Cardiovascular, Centro de Investigacion en Salud Poblacional, Instituto Nacional de Salud Publica, Mexico City, Mexico.
22. Departamento de Endocrinología y Metabolismo. Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City.
23. Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California,USA.
24. Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii, USA.
25. The Genomics Platform, The Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA.
26. Department of Medicine,University of Texas Health Science Center at San Antonio, San Antonio, Texas, USA.
27. Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA.
28. Unidad de Investigación Médica en Enfermedades Metabólicas, CMN SXXI, Instituto Mexicano del Seguro Social, Mexico City, México.
29. South Texas Diabetes and Obesity Institute, School of Medicine, University of Texas Rio Grande Valley, Brownsville, TX, USA.
30. South Texas Diabetes and Obesity Institute, School of Medicine, University of Texas Rio Grande Valley, Edinburg, TX, USA.
31. Division of Diabetes, Department of Medicine, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA.

32. Departments of Medicine and Cellular & Structural Biology, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA.
33. Human Genetics Center, University of Texas Health Science Center at Houston, Houston, Texas 77030, USA.
34. Department of Medicine, University of Chicago, Chicago, Illinois 60637, USA.
35. Department of Human Genetics, University of Chicago, Chicago, Illinois 60637, USA.
36. Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan 48109, USA.
37. Genomic Programming of Beta-cells Laboratory, Institut d'Investigacions August Pi i Sunyer (IDIBAPS), 08036 Barcelona, Spain.
38. CIBER de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), 08036 Barcelona, Spain.
39. Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.
40. Department of Molecular Biology, Harvard Medical School, Boston, Massachusetts, USA.
41. Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

Corresponding author:

Jose C. Florez, M. D. Ph. D.
Chief, Diabetes Unit
Massachusetts General Hospital
Associate Professor of Medicine
Harvard Medical School
Institute Member
Broad Institute
Diabetes Unit, Department of Medicine
Center for Genomic Medicine
Richard B. Simches Research Center
Massachusetts General Hospital
185 Cambridge Street, CPZN 5.250
Boston, MA 02114
Office: 617-643-3308
Fax: 617-726-5735
Email: jcflorez@mgh.harvard.edu

Supplementary Figure legends

Supplementary Figure 1: Genotype integration and imputation strategy. The sample was divided into two cohorts according to the technologies used to ascertain the genetic variants. A subset of the samples (Dataset 1) was genotyped on the OMNI2.5 (*1*)and exome chip arrays. Dataset 2 was also ascertained by whole-exome sequencing and was used as a Mexican specific reference panel (*2*). We imputed both the variants of the whole-exome-sequencing Mexican specific reference panel and the 1000G (phase 3, release June 2014) variants into the samples that did not have whole-exome sequencing information. We only imputed 1000G (phase 3) variants into the samples that had whole-exome sequencing, OMNI2.5 and exome chip genotypes. We then performed the association testing separately of each dataset and meta-analyzed both results.

Supplementary Figure 2: Characterization of the *IGF2-INS-TH* locus. Regional plot is shown for the *IGF2-INS-TH* and *KCNQ1*, without conditioning (a). Conditioning on rs139647931 and rs2237897 *KCNQ1* variants revealed two additional independent signals, rs4929965 and rs149483638 (b). 95% credible set when conditioning on the two *KCNQ1* variants and rs4929965 reveals the splice acceptor site variant (rs149483638) as top variant (c). 95% credible set when conditioning on rs149483638 and the *KCNQ1* variants (d). Point colors indicate the R-squared with the index SNP, marked in purple.

Supplementary Figure 3: Synthetic association plot. Whole-exome sequencing data from 3,732 individuals was integrated with OMNI2.5 and exome array genotypes as described in Supplementary Figure 1. All the variants with minor allele frequency (MAF)<0.001 were removed from the analysis. Low-frequency variants (0.001<MAF<0.05) were sequentially added into the model starting from the most significant variant. The first line represents the odds ratio (OR) and the 95% confidence interval for the association of the rs149484648 variant. Each of the following lines represent the estimated OR and 95% confidence interval of the rs149483638 after adding, in a sequential manner each of the low-frequency variants into the model. As shown in the plot, the rs149483638 association was did not disappear when adding up to 50 low-frequency variants into the model, which suggest that this signal is not driven by the presence of low-frequency variants strongly associated with the disease.

Supplementary Figure 4: Forest plot for the meta-analysis of rs10770141 variant, near *TH*. Odds ratios for the meta-analyses are represented with a diamond.

Supplementary Figure 5. Gene expression of isoform 2, as measured by expression of exon 2 of *IGF2* across tissues and cell lines. Expression across T2D- relevant tissues extracted from GTEx (left panel). Gene expression was also analyzed in different embryonic cell lines, including embryonic cell-line derived pancreatic precursors (right panel). *human embryonic stem cells presented here were included among the cell lines approved by NIH in 2010 to preempt the use of federal funds to generate new cell lines from human embryos.

Supplementary Figure 6. Expression level of *IGF2* in human pancreatic progenitors and other tissues. RNA-seq data showing the expression of *IGF2* isoforms in various adult human tissues (lung, muscle, adrenal, heart, adipose, breast, colon, brain, prostate and kidney), hESC-derived cell lines (mesenchymal, mesendoderm, trophoblast and neuronal progenitors) and hESC-derived pancreatic progenitors. The yellow box highlights the location of the *IGF2* exon of interest, which is highly expressed in human pancreatic progenitors, as opposed to most other tissues and cell lines. All tracks are scaled to a maximum of 20 RPKM.

Supplementary Figure 7. The dosage of the T2D protective A allele was not correlated with expression of total *IGF2* expression (as measured by expression levels of the exon 3) in liver (**a**) and in adipose tissue (**b**). The expression of all IGF2 (as measured by exon 3 expression) was not correlated with T2D (**c**) status or HbA1c in non-diabetic individuals (**d**).

Supplementary Figure 8. The dosage of the T2D protective A allele is not associated with total circulating IGF2.

Supplementary Figure 9. Association of human homozygous “knockouts” for isoform 2 of *IGF2* (AA homozygous) with other diseases or clinical outcomes. While “homozygous knock-outs” were associated with ~40% reduced risk for T2D, there was no evidence of increased risk for other diseases in individuals homozygous for the A allele.

Supplementary Note (online)

Detailed description of study participants

Diabetes in Mexico Study (DMS):

Individuals were enrolled in the study, recruited from two tertiary level institutions (IMSS and ISSSTE) located in Mexico City. The diagnosis of T2D was made based on ADA criteria. 811 unrelated healthy subjects older than 45 years and with fasting glucose levels below 100 mg/dL were classified as controls. 569 unrelated individuals, older than 18 years, with either previous T2D diagnosis or fasting glucose levels above 125 mg/dL were included as T2D cases. Individuals with fasting glycemia between 100-125 mg/dL were excluded. Informed consent was obtained from all participants. The study was conducted with the approval of the Ethics and Research Committees of all institutions involved. Genomic DNA was purified from whole blood samples using a modified salting-out precipitation method (Gentra Puregene, Qiagen Systems, Inc., Valencia, CA, USA).

Clinical history of these individuals was manually reviewed and incidence of difference diseases was used for the phenome-wide association analysis.

Mexico City Diabetes Study (MCDS):

The Mexico City Diabetes Study is a population based prospective investigation. All 35-64 years of age men and non-pregnant women residing in the study site (low income neighborhoods equivalent to 6 census tracks with a total population of 15,000 inhabitants) were interviewed and invited to participate in the study. We had a response rate of 67% for the initial exam. Diagnostic criteria for type 2 diabetes were recommended by the ADA. Fasting glucose 126 mg/dL or more or 2 hr post 75 gr of glucose load 200 or more. If a participant was diagnosed as diabetic by a physician and was under pharmacologic therapy for diabetes he was considered as diabetic regardless the blood glucose levels. The study was conducted with the approval of the Ethics and Research Committees

For Peer Review Only

of all institutions. Informed consent was obtained from all participants. Genomic DNA was extracted from whole blood using the QIAmp 96 DNA Blood Kit (12) (Qiagen, Cat. No. 51162).

Multiethnic Cohort (MEC):

The MEC consists of 215,251 men and women in Hawaii and Los Angeles, and comprises mainly five self-reported racial/ethnic populations: African Americans, Japanese Americans, Latinos, Native Hawaiians and European Americans (3). Between 1993 and 1996, adults between 45 and 75 years old were enrolled by completing a 26-page, self-administered questionnaire asking detailed information about dietary habits, demographic factors, level of education, personal behaviors, and history of prior medical conditions (e.g., diabetes). Potential cohort members were identified through Department of Motor Vehicles drivers' license files, voter registration files and Health Care Financing Administration data files. In 2001, a short follow-up questionnaire was sent to update information on dietary habits, as well as to obtain information about new diagnoses of medical conditions since recruitment. Between 2003 and 2007, we re-administered a modified version of the baseline questionnaire. All questionnaires inquired about history of diabetes, without specification as to type (1 vs. 2). Between 1995 and 2004, blood specimens were collected from ~67,000 MEC participants at which time a short questionnaire was administered to update certain exposures, and collect current information about medication use.

Cohort members in California are linked each year to the California Office of Statewide Health Planning and Development (OSHPD) hospitalization discharge database, which consists of mandatory records of all in-patient hospitalizations at most acute-care facilities in California. Records include information on the principal diagnosis plus up to 24 other diagnoses (coded according to ICD-9), including T1D and T2D. In Hawaii cohort members have been linked with the diabetes care registries for subjects with Hawaii Medical Service Association (HMSA) and Kaiser Permanente Hawaii (KPH) health plans (~90% of the Hawaii population has one of these two plans). Information from these additional databases has been utilized to assess the percentage of

T2D controls (as defined below) with undiagnosed T2D, as well as the percentage of identified diabetes cases with T1D rather than T2D. Based on the OSHPD database <3% of T2D cases had a previous diagnosis of T1D. We did not use these sources to identify T2D cases because they did not include information on diabetes medications, one of our inclusion criteria for cases (see below).

In the MEC, diabetic cases were defined using the following criteria: (a) a self-report of diabetes on the baseline questionnaire, 2nd questionnaire or 3rd questionnaire; and (b) self-report of taking medication for T2D at the time of blood draw; and (c) no diagnosis of T1D in the absence of a T2D diagnosis from the OSHPD (California Residents). Controls were defined as: (a) no self-report of diabetes on any of the questionnaires while having completed a minimum of 2 of the 3 (~80% of controls returned all 3 questionnaires); and (b) no use of medications for T2D at the time of blood draw; and (c) no diabetes diagnosis (type 1 or 2) from the OSHPD, HMSA or KPH registries. To preserve DNA for genetic studies of cancer in the MEC, subjects with an incident cancer diagnosis at time of selection for this study were excluded. Controls were frequency matched to cases on sex, ethnicity and age at entry into the cohort (5-year age groups) and for Latinos, place of birth (U.S. vs. Mexico, South or Central America), oversampling African American, Native Hawaiian and European American controls to increase statistical power. Many of the T2D variants have also been evaluated in studies of cancer in the MEC which allowed for inclusion of additional controls who met the criteria above.

Altogether, this study included 2,231 T2D cases and 2,607 controls of Latin American ethnicity. Informed consent was obtained from all participants. The study was conducted with the approval of the Ethics and Research Committees of all institutions. Genomic DNA extraction was done using Qiagen from buffy coat.

UNAM/INCMNSZ Diabetes Study (UIDS):

Cases were recruited at the outpatient diabetes clinic of the Department of Endocrinology and Metabolism of the Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán (INCMNSZ). All Mexican-mestizo individuals were invited to participate in the study. Diagnosis of type 2 diabetes was done following the American Diabetes Association criteria, i.e., fasting plasma glucose values ≥126 mg/dL, current treatment with a hypoglycemic agent, or casual glucose values ≥200 mg/dL.

Control subjects were recruited from a cohort of adults aged 45 years or older among government employees, blue collar workers and subjects seeking for attention in medical units for any condition besides those considered as exclusion criteria (see below). Normoglycemic status was defined as having a fasting plasma glucose concentration < 100 mg/dL and no previous history of hyperglycemia, gestational diabetes or use of metformin.

Patients were interviewed following a standardized questionnaire; it included the medical history, a previously validated, three days food record and a physical activity registry. In addition a blood sample (after 9-12 hours of fasting) was obtained. The questionnaire included demographic, socio-economic and medical history of the patients and their family. Blood pressure, height, waist circumference and weight must be measured in the same visit. For taking blood pressure, systolic and diastolic pressure were recorded using a mercury sphygmomanometer; subjects remained seated and at rest for five minutes before measuring.

Inclusion criteria: Men or women aged 25 years or older, with BMI greater than 20 but lower than 40 kg/m2.

Exclusion criteria: Diabetes, coronary heart disease, stroke, transient ischemic attack, lower limb amputations, alcoholism (more than 10 servings of alcohol per week) or any disease that in opinion of the researcher may limit life expectancy to less than 2 years. Subjects that planned to move out of town permanently during the next three years were also excluded. Pregnant women,

individuals with drug addictions, the use of systemic corticosteroids in pharmacologic doses (intravenous, oral or injectable, including injections in the joints) were exclusion criteria also. Replacement dosage of systemic corticosteroids (up 7.5 mg/day of prednisone or 30 mg/day of hydrocortisone or its equivalent; as well as inhaled or topical corticosteroids) was allowed into the study. Other exclusion criteria were: active liver disease (defined as AST (SGOT) or ALT (SGPT) > 2.0x upper limit of the normal range, alkaline phosphatase (ALK-P) > 1.5x upper limit of the normal range or total bilirubin > 1.5x upper limit of the normal range), significant renal dysfunction (defined as serum creatinine > 1.7 upper limit of the normal range or nephrotic syndrome), any history of malignancy (except for basal cell skin carcinoma) and uncontrolled depression or psychosis.

Informed consent was obtained from all participants. The study was conducted with the approval of the Ethics and Research Committees of all institutions. Genomic DNA was extracted from whole blood using the QIAmp 96 DNA Blood Kit (12) (Qiagen, Cat. No. 51162).

Pima Native Americans

Diabetes was diagnosed by 1997 American Diabetes Association criteria; details for this cohort have been described previously (4). rs149483638 was analyzed in 3,199 full heritage Pima Indians selected from a longitudinal study (4). These individuals included 1,847 women and 1,352 men; mean age at examination was 40.6 (±16.5) years, and 1519 individuals (47%) had diabetes. The frequency of the A allele of rs149483638 was 0.149 (0.169 in non-T2D and 0.135 in T2D). Mean (SD) age was 40.8 (16.4) (49.5 [13.0] for affected, 33.0 [14.4] for unaffected). Mean (SD) age at onset for affected was 35.9 (12.5). Mean (SD) maximum BMI observed in the longitudinal study was 37.6 (8.7) (38.9 [8.7] for affected; 36.3 [8.5] for unaffected). Mean (SD) fasting glucose: 134.1 (69.0) mg/dL (181.1 [78.5] for affected; 94.1 [9.6] for unnaffeted).

Genotypes were assessed by the BeadXpress system (Illumina, San Diego, CA) according to the manufacturer's instructions. The association between genotype and diabetes was assessed under an additive model by logistic regression model with control for age, sex, birth year, and proportion of Amerindian ancestry (estimated using 45 ancestry informative markers). The model was fit by the generalized estimating equation method to account for familial dependence among siblings.

The Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples T2D-GENES Consortium (T2D-GENES):

All the exons were sequenced in 12,294 additional individuals as part of the whole-exome sequencing studies performed through the Genetics of Type 2 Diabetes (GoT2D) and Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples (T2D-GENES) consortia. Individuals were selected spanning 5 ethnicities: European (the FUSION study (5)[FUSION], the METSIM study (6)[METSIM], KORA-gen (7)[KORA], the WTCCC/UKT2D consortium (8; 9)and the UK Adult Twin Registry (10)[UKT2D], as well as Ashkenazi individuals recruited from the metropolitan New York region (11)[Ashkenazim] and small number of individuals from the Finnish [Botnia] and Swedish [Malmo] prospective cohorts used for the initial sequencing experiment (12-19)), African-American (the Jackson Heart Study (JHS) cohort [JHS] as well as additional individuals recruited from North Carolina, South Carolina, Georgia, Tennessee, or Virginia (20)[WFS]), South Asian (the London Life Sciences Prospective Population Study (LOLIPOP) (21; 22)[LOLIPOP] and Singapore Indian Eye Study (SINDI) (23)[Singapore Indians]), East Asian (the Korean Association REsource (KARE) (24)[KARE] as well as the Singapore Diabetes Cohort Study (SDCS) and Singapore Prospective Study Program (25-27)[Singapore Chinese]), and Hispanic (the San Antonio Family Heart Study (FHS) (28), the San Antonio Family Diabetes/Gallbladder Study (SAFDGS) (29), the Veterans Administration Genetic Epidemiology Study (VAGES) (30), the Family Investigation of Nephropathy and Diabetes (FIND) (31), San Antonio component [San Antonio], and individuals from Starr County, TX (32)[Starr

For Peer Review Only

County]). Data generation and processing was performed in an identical fashion as for the initial sequencing experiment, although target capture was performed with the Agilent SureSelect Human All Exon platform rather than a custom hybrid capture array.

Diabetes in Mexico Study 2 (DMS2):

This cohort included 1228 unrelated volunteers from different ethnic groups along Mexico (Tarahumara, Yaqui, Mayo, Mixteco, Náhuatl, Otomí, Chinanteco, Mixe, Zapoteco, Mazateco, Totonaco, Huasteco, Maya, Kanjobal, Mame, Poptijacalteco, Kaqchikel, Tojolabal, Chontal, Huave). Inclusion criteria were that they identified themselves as indigenous, both parents and their four grandparents speak the same native language and were born in the same community. The diagnosis of T2D was made based on ADA criteria. Seven hundred eighty-three unrelated healthy subjects older than 45 years and with fasting glucose levels below 100 mg/dL were classified as controls. Four hundred forty-five unrelated individuals, older than 18 years, with either previous T2D diagnosis or fasting glucose levels above 125 mg/dL were included as T2D cases. Individuals with fasting glucose levels between 100-125 mg/dL were excluded. Informed consent was obtained from all participants. The study was conducted with the approval of the Ethics and Research Committees of all institutions involved. Genomic DNA was purified from whole blood samples using a modified salting-out precipitation method (Gentra Puregene, Qiagen Systems, Inc., Valencia, CA, USA). Genotyping of the rs149483638 variant was performed using a custom TaqMan SNP Genotyping Assay (Applied Biosystems, Foster City, CA, USA) and genotype of each sample was assigned automatically by SDS 2.3 software (Applied Biosystems, Foster City, CA, USA). For the genotyping quality control, 5% of samples were randomly selected and measured in duplicates. TaqMan probes: Allele ‘C’ VIC-CAAACTCTCCA[G]GAGATG. Allele ‘T’ FAM-CAAACTCTCCA[A]GAGATG. 5 Positive controls were added to all plates and verified that their genotype matched the expected.

Association analyses were performed by logistic regression adjusting for age, BMI, gender ant the

first two principal components derived from a panel of 96 ancestry informative markers (33).

Clinical history of these individuals was manually reviewed and incidence of difference diseases was used for the phenome-wide association analysis.

San Antonio Mexican American Family Studies (SAMAFS):

Genotypes for rs11564732, which showed an r2=0.89 with rs149483638 in our discovery sample, was carried out using the MassARRAY system (Sequenom, San Diego, CA). Variant assay primers were designed using Sequenom's online assay design tool in conjunction with their MassARRAY Assay Designer v4.0 software, to amplify ~100bp surrounding the variant for amplification in the MassEXTEND reaction. The MassARRAY Matrix Liquid Handler was used for automated preparation of reaction products which were then spotted onto 384-sample SpectroCHIP arrays using the MassARRAY Nanodispenser chip spotting station. Spotted arrays were loaded into the MassARRAY Analyzer 4 and sample genotypes determined by measuring the migration times, within a vacuum for each base at a specific locus (MALDI-TOF MS). Analysis of spectra and generation of genotypes was conducted using Sequenom's TyperAnalyzer software v4.0.21. Samples were from three Mexican American family studies from San Antonio, Texas: San Antonio Family Heart Study (28)(SAFHS); San Antonio Family Diabetes/Gallbladder Study (29)(SAFDGS); and the Veterans Administration Genetic Epidemiology Study (30). These studies are referred to as the San Antonio Mexican American Family Studies (SAMAFS). The rs11564732 genotypic data were available for 2,980 SAMAFS individuals (Mean age [± SD]=49.9 ± 15.6; Mean BMI [± SD]=32.3 ± 8.2; Females=61%; T2D=40%).

The association analysis was carried out using FBAT, assuming the additive model and using the residuals resulting from regressing out age, age^2 and BMI (34).

Diabetes Prevention Program

The DPP enrolled 3,234 US participants at high risk of developing diabetes (on the basis of overweight, increased fasting glucose and impaired glucose tolerance) and randomised them to placebo, metformin 850 mg twice daily or a lifestyle intervention aimed at ≥7% weight loss and ≥150 min of physical activity per week; a fourth arm of 585 participants initially randomised to troglitazone was terminated early because of concerns with hepatotoxicity (35). The main endpoint was development of diabetes confirmed by OGTT. The trial was conducted at 27 centers, all of which obtained individual Institutional Review Board approval. The DPP showed that participants treated with metformin or with a lifestyle intervention were 31% or 58% less likely to develop diabetes after an average of 3 years of follow-up, respectively (35). The 3,548 DPP participants presented here (2,994 who completed the trial in the placebo, metformin or lifestyle arms, plus 554 originally randomised to troglitazone) provided informed consent specific to genetic investigation. The distribution of self-reported ethnicities among participants in this genetic study was 56.4% white, 20.2% African American, 16.8% Hispanic, 4.3% Asian and 2.4% American Indian. The mean age was 51 years and mean BMI was 34.0 kg/m^2.

rs149483638 was genotyped as part of the Human Core Exome genotyping array from Illumina at the Genomics Platform at the Broad Institute. Genotyping and calling was performed as described below. Genotyping of rs149483638 was completed with high quality on the Human Core Exome genotyping arrays (99.9% call rate). The association between the outcome variables and the marker was tested using multivariable mixed model regression. All models were adjusted for age at randomization, sex, study site, and population stratification using principal components derived only from Hispanic and Amer-Indian subjects. Clinic was entered into the model as a random effect. In additional models, waist was added in addition to these covariates. Because of the low frequency of the minor allele in the other ethnic groups, only self-reported Hispanics (n=538) and American Indians (n=78) are included in the analysis. These two groups were analyzed together and then tested in stratified models as only the Hispanic group has individuals homozygous for the rare

For Peer Review Only

allele. Given the possibility that variants in *IGF2* exhibit a parent-of-origin effect, besides the additive model (per allele effect) we also tested the homozygous extremes model (comparing CC vs TT) in order to reduce the noise of heterozygous subjects in cases parent-of-origin is present.

The natural log of ISI and the IGR was used in the analysis because their distributions are skewed. Because these hypotheses represent confirmation of previous findings (and thus possess a high prior probability), a P value of 0.05 was considered statistically significant. Insulin sensitivity index (ISI), the reciprocal of insulin resistance by homeostasis model assessment, was calculated as 22.5/[fasting insulin x (fasting glucose/18.01)]. The mean ISI values were compared at baseline.

The Insulin Index was defined as [(insulin at 30 min) – (insulin at 0 min)]/[(glucose at 30 min) – (glucose at 0 min)]. The mean Ins Index values were compared at baseline.

Resource for Genetic Epidemiology Research on Adult Health and Aging (GERA)

GERA cohort data was obtained through dbGaP under accession phs000674.v1.p1 (36). The Resource for Genetic Epidemiology Research on Aging (GERA) Cohort was created by a RC2 "Grand Opportunity" grant that was awarded to the Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH) and the UCSF Institute for Human Genetics (AG036607; Schaefer/Risch, PIs). The RC2 project enabled genome-wide SNP genotyping (GWAS) to be conducted on a cohort of over 100,000 adults who are members of the Kaiser Permanente Medical Care Plan, Northern California Region (KPNC), and participating in its RPGEH. The resulting GERA cohort is 42% male, 58% female, and ranges in age from 18 to over 100 years old with an average age of 63 years at the time of the RPGEH survey (2007). The sample is ethnically diverse, generally well-educated with above average income. Approximately 69% of the participants are married or living with a partner. Length of membership in KPNC averages 23.5 years. UCSF and RPGEH investigators worked with the genomics company Affymetrix to design four custom microarrays for genotyping each of the four major race-ethnicity groups included in the GERA

Cohort, described in detail in Hoffmann et al., 2011a and 2011b. Following genotyping and quality control procedures, and after removal of invalid, discordant, or withdrawn samples, about 103,000 participants were successfully genotyped. The resulting genotypic data were linked to survey data and data abstracted from the electronic medical records. As described below, all RPGEH participants were mailed new consent forms with explicit discussion of the placement of data in the NIH-maintained dbGaP. About 77% of participants returned completed consent forms, resulting in a final sample size of 78,486 participants in the GERA Cohort with data for deposit into dbGaP.

A subset of Hispanic individuals (1064 cases and 4832 controls) from the GERA cohort, as potential carriers of the rs149483638 variant were separately QCed and analyzed. All genotyped datasets separately underwent the same 3-step quality control protocol using PLINK and included 2 stages of SNP removal and an intermediate stage of sample exclusion.

The exclusion criteria for genetic markers consisted on: proportion of missingness ≥ 0.05, Hardy-Weinberg Equilibrium p-value $\leq 1x10^{-20}$ for all the cohort. Only for the GERA cohort we considered a MAF of 0.001 as exclusion criteria because of the large sample size of this dataset. This protocol for genetic markers was performed twice, before and after sample exclusion.

For the individuals, we considered the following exclusion criteria: gender discordance, subject relatedness (pairs with ≥ 0.125 from which we removed the individual with the highest proportion of missingness), variant call rates ≥ 0.02 and population structure showing more than 4 standard deviations within the distribution of the study population according to the first seven principal components.

The presence of up to 18 medical conditions besides T2D was taken into account for the phenome-wide association analysis. The description of ICD9 codes that are included in each of the medical conditions can be found here (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/GetPdf.cgi?id=phd004308).

We performed a two-stage imputation procedure, which consisted in pre-phasing the genotypes into whole chromosome haplotypes followed by imputation itself. The pre-phasing was performed using the SHAPEIT2 (37)tool, IMPUTE2 for genotype imputation and the SNPTEST (https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html#introduction) tool for association testing. GWIMp-COMPSs can incorporate the contribution of several reference panels, and in this work we used 1000 Genomes (1000G) Phase3 haplotypes (October, 2014) (38). Association testing was performed by additive logistic regression using SNPTEST, and adjusting for the 7 derived principal components, age, and body mass index.

Exome chip SNP genotyping and quality control

The Genomics Platform at the Broad Institute (Cambridge, MA) received, QC'd and tracked DNA samples for Exome array processing. The exome array was designed in order to cover rare and low-frequency coding variants identified through whole-exome sequencing studies of 12,031 individuals from different populations including 362 individuals of Hispanic ancestry.

The samples were plated into 96-well plates that included a quality control sample for processing on the Illumina HumanExome BeadChip (Illumina, Inc. San Diego, CA) using manufacturer's protocols. The arrays were scanned using Illumina iScans. Genotypes were called using three different calling algorithms: Illumina GenCall, Z-call (39) and Birdsuite (http://www.broadinstitute.org/science/programs/medical-and-population-genetics/birdsuite/birdsuite-0).

Clusters were fit using the Birdseed algorithm to each genotyping plate independently. Genotypes with confidence below 99.9% were excluded from analysis (e.g. considered "missing" or "no-call" genotypes). Samples with low numbers of non-reference alleles (< ~20,000, depending on the cohort), low call rate (<99.3%) or unusually high heterozygosity (> ~0.05, depending on the cohort) were removed from subsequent analysis; thresholds were chosen based on visual inspection of the

sample distributions. Variants with low call rate (<99.2%) or mean confidence for alternative genotype calls (<99%) were also excluded from subsequent analysis.

RNA isolation from frozen tissue samples

RNA was extracted via the miRNeasy Mini Kit from Qiagen. This kit combines a phenol/guanidine-based lysis and a silica-membrane based purification. Tissue specimens were prepared and cut to 20-25 mg on a dry ice bath, then placed in 2 mL processing tubes containing QIAzol lysis reagent and a steel bead. Tubes were then placed in the TissueLyser for 5 min at 25Hz to lyse and homogenize the samples.

After homogenizing, tubes were incubated at room temperature for 5 min. 140 µl chloroform was then added to each tube containing homogenate. Samples incubated at room temperature for 2-3 min and were centrifuged at 12,000xg/4C for 15 min.

After centrifugation, the samples separated into 3 phases: an upper, colorless aqueous phase containing RNA, a white interphase, and a lower, red organic phase. The upper aqueous phase was carefully transferred to a new 1.5mL eppendorf tube (~350 µL). 525 µL of 100% ethanol was added to this phase and mixed thoroughly by pipetting up and down several times.

The entire sample, including any precipitate, was pipetted into RNeasy mini spin columns and then centrifuged for 15 s at 8000xg in order to collect and discard the flow-through.

350 µL Buffer RWT was pipetted into the RNeasy Mini spin column and centrifuged for 15s at 8000xg to wash. The flow-through was again discarded.

80 µL of DNase I diluted with buffer RDD was pipetted directly onto each column membrane and incubated at room temp for 15 minutes.

350 µL Buffer RWT was then pipetted onto the DNase I remaining on the RNeasy Mini spin column and centrifuged for 15s at 8000xg, flow through discarded

700 µL Buffer RWT was then added to the RNeasy Mini spin column, centrifuged for 15s at 8000xg, and the flow through discarded.

500µL Buffer RPE was then pipetted onto the RNeasy Mini spin column and centrifuged for 15s at 8000xg. The flow through was again discarded Again 500 µL Buffer RPE was pipetted onto the RNeasy Mini spin column and centrifuged for 2 min at 8000xg.

The RNeasy Mini spin columns were then placed into new 2mL collection tubes and centrifuged at full speed for 1 min. This allowed the membrane to fully dry out, ensuring no ethanol was carried over during RNA elution.

For each sample, the RNeasy Mini spin column was transferred to a new 1.5 ml eppendorf tube. 20 µL RNase-free water was then added directly to the spin column membrane. The tubes were then centrifuged for 1 min at 8000xg to elute the RNA. This step was then repeated a second time.

The RNA samples were then capped and incubated at 65˚C × 5 minutes to denature the RNA and then chilled immediately on wet ice for 2-3 minutes.

Supplementary text (online)

Fine mapping and credible set analysis of IGF2-INS-TH locus

In order to better characterize both the primary and the secondary signal, we performed a credible set analysis of the two regions, after integrating the exome chip results with Omni 2.5 Illumina array genotypes that were available for the majority of the samples (1), and exome-sequencing data that was available for 41% of the samples (2)genotyped by exome-chip. We imputed both exome-seq and 1000G (phase 3, release June 2014) variants into the samples that did not have whole-exome sequencing information, and only 1000G (phase 3) variants into the samples that had whole-exome sequencing, OMNI2.5 and exome chip genotypes. We then performed the association testing separately of each dataset and meta-analyzed both results (Supplementary Figure 1). We used this

data as input to perform the credible set analysis, following the previously described methodology (40). The splice site variant (rs149483638) ranked first according to its posterior probability, being the one with the highest prior of being the causal variant (Supplementary Figure 2).

When conditioning for these two *KCNQ1* variants and the rs149483638 in order to identify secondary signals, we identified another independent association with rs10770141 (OR=1.14; p=3×10^{-4}) located at the promoter region of TH gene. This variant was previously reported in a gene-centric meta-analysis (41). When meta-analyzing our data with those of this study, rs10770141 resulted in a novel GWAS significant signal, which was independent of the first splice-site variant in *IGF2* (OR=1.08 (1.05-1.11); p=1.1×10^{-9}) (Supplementary Figure 1).

We also used the fraction of the dataset for which we had exome-sequencing data in order to discard if rare variants with a strong effect size are the cause of a "synthetic association" (42), i. e. that rare variants with large effect size, are responsible for the association signal found in the common variant (rs149483638). For this purpose, we performed multiple regression analyses introducing the rare variants as covariates, to test if this caused a loss of the association signal in rs149483638. This analysis confirmed that we were not in front of a synthetic association, as at least 31 rare variants had to be excluded in order to dilute the signal of the rs149483638 variant (Supplementary Figure 3).

Credible set analysis

The credible set analysis were constructed as described in (40). Briefly, for each association analyses results, we computed an approximate Bayes factor for each variant,

$$r = \frac{0.04}{se^2 + 0.04}$$

$$z = beta/se$$

$$abf = \frac{\sqrt{(1-r)}}{\exp\left(-r \times \frac{z^2}{2}\right)}$$

assuming that the prior on beta is Gaussian and variance 0.04. Then, a posterior probability for each variant was computed dividing the ABF by the total number of variants in the region. Then the cumulative posterior probability was computed and all the variants in the 95% credible set interval were selected and included in the credible set. Assuming that we had two independent signals in the *IGF2* region, we computed a first credible set conditioning on the two *KCNQ1* T2D associated variants (rs139647931 and rs2237897) and the rs4929965, and a second credible set conditioning on rs149483638 variant and the two *KCNQ1* T2D associated variants (rs139647931 and rs2237897). We selected all the variants that showed an r-squared higher than 0.1 with the top variants in each of the regions to compute the credible set analysis.

For Peer Review Only

References

1. SIGMA Type 2 Diabetes Consortium, Williams AL, Jacobs SB, Moreno-Macias H, Huerta-Chagoya A, Churchhouse C, Marquez-Luna C, Garcia-Ortiz H, Gomez-Vazquez MJ, Burtt NP, Aguilar-Salinas CA, Gonzalez-Villalpando C, Florez JC, Orozco L, Haiman CA, Tusie-Luna T, Altshuler D: Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. Nature 2014;506:97-101

2. Estrada K, Aukrust I, Bjorkhaug L, Burtt NP, Mercader JM, Garcia-Ortiz H, Huerta-Chagoya A, Moreno-Macias H, Walford G, Flannick J, Williams AL, Gomez-Vazquez MJ, Fernandez-Lopez JC, Martinez-Hernandez A, Centeno-Cruz F, Mendoza-Caamal E, Revilla-Monsalve C, Islas-Andrade S, Cordova EJ, Soberon X, Gonzalez-Villalpando ME, Henderson E, Wilkens LR, Le Marchand L, Arellano-Campos O, Ordonez-Sanchez ML, Rodriguez-Torres M, Rodriguez-Guillen R, Riba L, Najmi LA, Jacobs SB, Fennell T, Gabriel S, Fontanillas P, Hanis CL, Lehman DM, Jenkinson CP, Abboud HE, Bell GI, Cortes ML, Boehnke M, Gonzalez-Villalpando C, Orozco L, Haiman CA, Tusie-Luna T, Aguilar-Salinas CA, Altshuler D, Njolstad PR, Florez JC, MacArthur DG: Association of a low-frequency variant in HNF1A with type 2 diabetes in a Latino population. JAMA 2014;311:2305-2314

3. Kolonel LN, Henderson BE, Hankin JH, Nomura AM, Wilkens LR, Pike MC, Stram DO, Monroe KR, Earle ME, Nagamine FS: A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. Am J Epidemiol 2000;151:346-357

4. Hanson RL, Rong R, Kobes S, Muller YL, Weil EJ, Curtis JM, Nelson RG, Baier LJ: Role of Established Type 2 Diabetes-Susceptibility Genetic Variants in a High Prevalence American Indian Population. Diabetes 2015;64:2646-2657

5. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M: A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 2007;316:1341-1345

6. Stancakova A, Javorsky M, Kuulasmaa T, Haffner SM, Kuusisto J, Laakso M: Changes in insulin sensitivity and insulin release in relation to glycemia and glucose tolerance in 6,414 Finnish men. Diabetes 2009;58:1212-1221

7. Wichmann HE, Gieger C, Illig T, Group MKS: KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen 2005;67 Suppl 1:S26-30

8. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, Zeggini E, Huth C, Aulchenko YS, Thorleifsson G, McCulloch LJ, Ferreira T, Grallert H, Amin N, Wu G, Willer CJ, Raychaudhuri S, McCarroll SA, Langenberg C, Hofmann OM, Dupuis J, Qi L, Segre AV, van Hoek M, Navarro P, Ardlie K, Balkau B, Benediktsson R, Bennett AJ, Blagieva R, Boerwinkle E, Bonnycastle LL, Bengtsson Bostrom K, Bravenboer B, Bumpstead S, Burtt NP, Charpentier G, Chines PS, Cornelis M, Couper DJ, Crawford G, Doney AS, Elliott KS, Elliott AL, Erdos MR, Fox CS, Franklin CS, Ganser M, Gieger C, Grarup N, Green T, Griffin S, Groves CJ, Guiducci C, Hadjadj S, Hassanali N, Herder C, Isomaa B, Jackson AU, Johnson PR, Jorgensen T, Kao WH, Klopp N, Kong A, Kraft P, Kuusisto J, Lauritzen T, Li M, Lieverse A, Lindgren CM, Lyssenko V, Marre M, Meitinger T, Midthjell K, Morken MA, Narisu N, Nilsson P, Owen KR, Payne F, Perry JR, Petersen AK, Platou C, Proenca C, Prokopenko I, Rathmann W, Rayner NW, Robertson NR, Rocheleau G, Roden M, Sampson MJ, Saxena R, Shields BM, Shrader P, Sigurdsson G, Sparso T, Strassburger K, Stringham HM, Sun Q, Swift AJ, Thorand B, Tichet J, Tuomi T, van Dam RM, van Haeften TW, van Herpt T, van Vliet-Ostaptchouk JV, Walters GB,

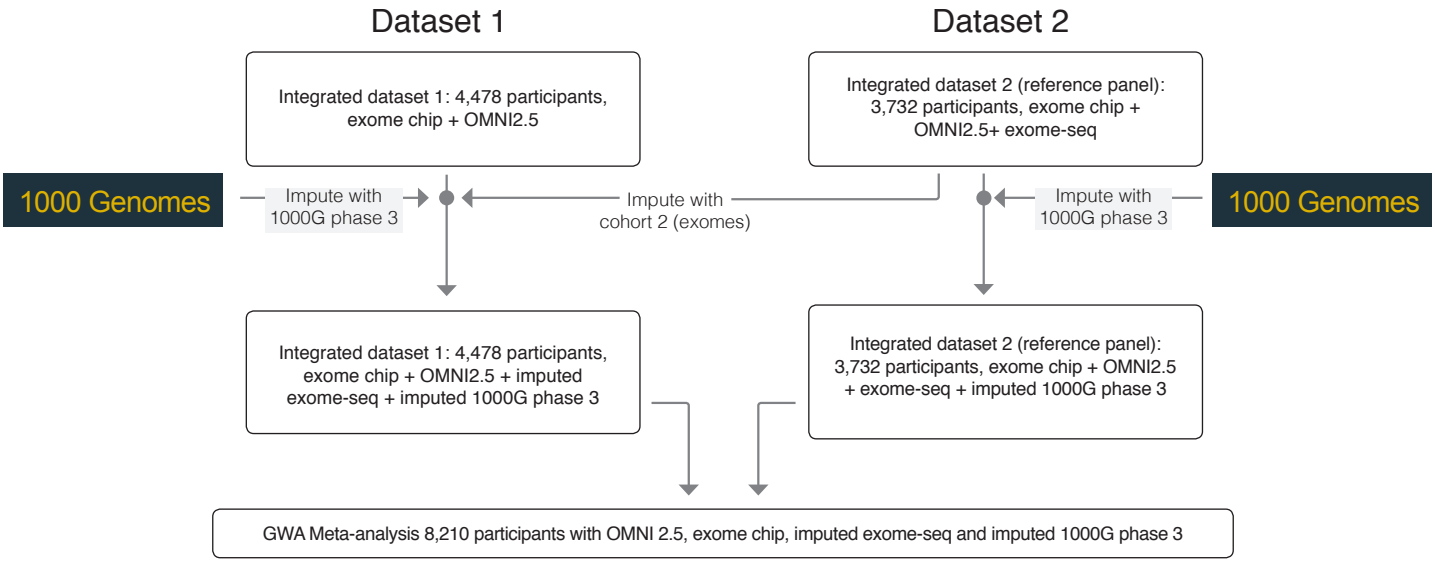
595 Weedon MN, Wijmenga C, Witteman J, Bergman RN, Cauchi S, Collins FS, Gloyn AL, Gyllensten U,
596 Hansen T, Hide WA, Hitman GA, Hofman A, Hunter DJ, Hveem K, Laakso M, Mohlke KL, Morris AD,
597 Palmer CN, Pramstaller PP, Rudan I, Sijbrands E, Stein LD, Tuomilehto J, Uitterlinden A, Walker M,
598 Wareham NJ, Watanabe RM, Abecasis GR, Boehm BO, Campbell H, Daly MJ, Hattersley AT, Hu FB,
599 Meigs JB, Pankow JS, Pedersen O, Wichmann HE, Barroso I, Florez JC, Frayling TM, Groop L, Sladek
600 R, Thorsteinsdottir U, Wilson JF, Illig T, Froguel P, van Duijn CM, Stefansson K, Altshuler D,
601 Boehnke M, McCarthy MI, investigators M, Consortium G: Twelve type 2 diabetes susceptibility
602 loci identified through large-scale association analysis. Nat Genet 2010;42:579-589
603 9. Wellcome Trust Case Control C: Genome-wide association study of 14,000 cases of seven
604 common diseases and 3,000 shared controls. Nature 2007;447:661-678
605 10. Spector TD, Williams FM: The UK Adult Twin Registry (TwinsUK). Twin research and human
606 genetics : the official journal of the International Society for Twin Studies 2006;9:899-906
607 11. Atzmon G, Hao L, Pe'er I, Velez C, Pearlman A, Palamara PF, Morrow B, Friedman E, Oddoux C,
608 Burns E, Ostrer H: Abraham's children in the genome era: major Jewish diaspora populations
609 comprise distinct genetic clusters with shared Middle Eastern Ancestry. Am J Hum Genet
610 2010;86:850-859
611 12. Berglund G, Elmstahl S, Janzon L, Larsson SA: The Malmo Diet and Cancer Study. Design and
612 feasibility. Journal of internal medicine 1993;233:45-51
613 13. Berglund G, Nilsson P, Eriksson KF, Nilsson JA, Hedblad B, Kristenson H, Lindgarde F: Long-term
614 outcome of the Malmo preventive project: mortality and cardiovascular morbidity. Journal of
615 internal medicine 2000;247:19-29
616 14. Bog-Hansen E, Lindblad U, Bengtsson K, Ranstam J, Melander A, Rastam L: Risk factor
617 clustering in patients with hypertension and non-insulin-dependent diabetes mellitus. The
618 Skaraborg Hypertension Project. Journal of internal medicine 1998;243:223-232
619 15. Groop L, Forsblom C, Lehtovirta M, Tuomi T, Karanko S, Nissen M, Ehrnstrom BO, Forsen B,
620 Isomaa B, Snickars B, Taskinen MR: Metabolic consequences of a family history of NIDDM (the
621 Botnia study): evidence for sex-specific parental effects. Diabetes 1996;45:1585-1593
622 16. Isomaa B, Forsen B, Lahti K, Holmstrom N, Waden J, Matintupa O, Almgren P, Eriksson JG,
623 Lyssenko V, Taskinen MR, Tuomi T, Groop LC: A family history of diabetes is associated with
624 reduced physical fitness in the Prevalence, Prediction and Prevention of Diabetes (PPP)-Botnia
625 study. Diabetologia 2010;53:1709-1713
626 17. Lindholm E, Agardh E, Tuomi T, Groop L, Agardh CD: Classifying diabetes according to the new
627 WHO clinical stages. European journal of epidemiology 2001;17:983-989
628 18. Lyssenko V, Jonsson A, Almgren P, Pulizzi N, Isomaa B, Tuomi T, Berglund G, Altshuler D,
629 Nilsson P, Groop L: Clinical risk factors, DNA variants, and the development of type 2 diabetes. The
630 New England journal of medicine 2008;359:2220-2232
631 19. Parker A, Meyer J, Lewitzky S, Rennich JS, Chan G, Thomas JD, Orho-Melander M, Lehtovirta M,
632 Forsblom C, Hyrkko A, Carlsson M, Lindgren C, Groop LC: A gene conferring susceptibility to type 2
633 diabetes in conjunction with obesity is located on chromosome 18p11. Diabetes 2001;50:675-680
634 20. Yu H, Bowden DW, Spray BJ, Rich SS, Freedman BI: Linkage analysis between loci in the renin-
635 angiotensin axis and end-stage renal disease in African Americans. Journal of the American Society
636 of Nephrology : JASN 1996;7:2559-2564
637 21. Chahal NS, Lim TK, Jain P, Chambers JC, Kooner JS, Senior R: Does subclinical atherosclerosis
638 burden identify the increased risk of cardiovascular disease mortality among United Kingdom
639 Indian Asians? A population study. American heart journal 2011;162:460-466
640 22. Chahal NS, Lim TK, Jain P, Chambers JC, Kooner JS, Senior R: Ethnicity-related differences in left
641 ventricular function, structure and geometry: a population study of UK Indian Asian and European
642 white subjects. Heart 2010;96:466-471

643 23. Lavanya R, Jeganathan VS, Zheng Y, Raju P, Cheung N, Tai ES, Wang JJ, Lamoureux E, Mitchell P,
644 Young TL, Cajucom-Uy H, Foster PJ, Aung T, Saw SM, Wong TY: Methodology of the Singapore
645 Indian Chinese Cohort (SICC) eye study: quantifying ethnic variations in the epidemiology of eye
646 diseases in Asians. Ophthalmic epidemiology 2009;16:325-336
647 24. Cho YS, Go MJ, Kim YJ, Heo JY, Oh JH, Ban HJ, Yoon D, Lee MH, Kim DJ, Park M, Cha SH, Kim JW,
648 Han BG, Min H, Ahn Y, Park MS, Han HR, Jang HY, Cho EY, Lee JE, Cho NH, Shin C, Park T, Park JW,
649 Lee JK, Cardon L, Clarke G, McCarthy MI, Lee JY, Lee JK, Oh B, Kim HL: A large-scale genome-wide
650 association study of Asian populations uncovers genetic factors influencing eight quantitative
651 traits. Nat Genet 2009;41:527-534
652 25. Tan CE, Emmanuel SC, Tan BY, Jacob E: Prevalence of diabetes and ethnic differences in
653 cardiovascular risk factors. The 1992 Singapore National Health Survey. Diabetes Care
654 1999;22:241-247
655 26. Hughes K, Yeo PP, Lun KC, Thai AC, Sothy SP, Wang KW, Cheah JS, Phoon WO, Lim P:
656 Cardiovascular diseases in Chinese, Malays, and Indians in Singapore. II. Differences in risk factor
657 levels. Journal of epidemiology and community health 1990;44:29-35
658 27. Hughes K, Aw TC, Kuperan P, Choo M: Central obesity, insulin resistance, syndrome X,
659 lipoprotein(a), and cardiovascular risk in Indians, Malays, and Chinese in Singapore. Journal of
660 epidemiology and community health 1997;51:394-399
661 28. Mitchell BD, Kammerer CM, Blangero J, Mahaney MC, Rainwater DL, Dyke B, Hixson JE, Henkel
662 RD, Sharp RM, Comuzzie AG, VandeBerg JL, Stern MP, MacCluer JW: Genetic and environmental
663 contributions to cardiovascular risk factors in Mexican Americans. The San Antonio Family Heart
664 Study. Circulation 1996;94:2159-2170
665 29. Hunt KJ, Lehman DM, Arya R, Fowler S, Leach RJ, Goring HHH, Almasy L, Blangero J, Dyer TD,
666 Duggirala R, Stern MP: Genome-wide linkage analyses of type 2 diabetes in Mexican Americans -
667 The San Antonio Family Diabetes/Gallbladder Study. Diabetes 2005;54:2655-2662
668 30. Coletta DK, Schneider J, Hu SL, Dyer TD, Puppala S, Farook VS, Arya R, Lehman DM, Blangero J,
669 DeFronzo RA, Duggirala R, Jenkinson CP: Genome-wide linkage scan for genes influencing plasma
670 triglyceride levels in the Veterans Administration Genetic Epidemiology Study. Diabetes
671 2009;58:279-284
672 31. Knowler WC, Coresh J, Elston RC, Freedman BI, Iyengar SK, Kimmel PL, Olson JM, Plaetke R,
673 Sedor JR, Seldin MF, Family Investigation of N, Diabetes Research G: The Family Investigation of
674 Nephropathy and Diabetes (FIND): design and methods. Journal of diabetes and its complications
675 2005;19:1-9
676 32. Hanis CL, Ferrell RE, Barton SA, Aguilar L, Garza-Ibarra A, Tulloch BR, Garcia CA, Schull WJ:
677 Diabetes among Mexican Americans in Starr County, Texas. American journal of epidemiology
678 1983;118:659-672
679 33. Kosoy R, Nassir R, Tian C, White PA, Butler LM, Silva G, Kittles R, Alarcon-Riquelme ME,
680 Gregersen PK, Belmont JW, De La Vega FM, Seldin MF: Ancestry informative marker sets for
681 determining continental origin and admixture proportions in common populations in America.
682 Human mutation 2009;30:69-78
683 34. Horvath S, Xu X, Laird NM: The family based association test method: strategies for studying
684 general genotype--phenotype associations. Eur J Hum Genet 2001;9:301-306
685 35. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, Nathan DM:
686 Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. The New
687 England journal of medicine 2002;346:393-403
688 36. Kvale MN, Hesselson S, Hoffmann TJ, Cao Y, Chan D, Connell S, Croen LA, Dispensa BP, Eshragh
689 J, Finn A, Gollub J, Iribarren C, Jorgenson E, Kushi LH, Lao R, Lu Y, Ludwig D, Mathauda GK, McGuire
690 WB, Mei G, Miles S, Mittman M, Patil M, Quesenberry CP, Jr., Ranatunga D, Rowell S, Sadler M,

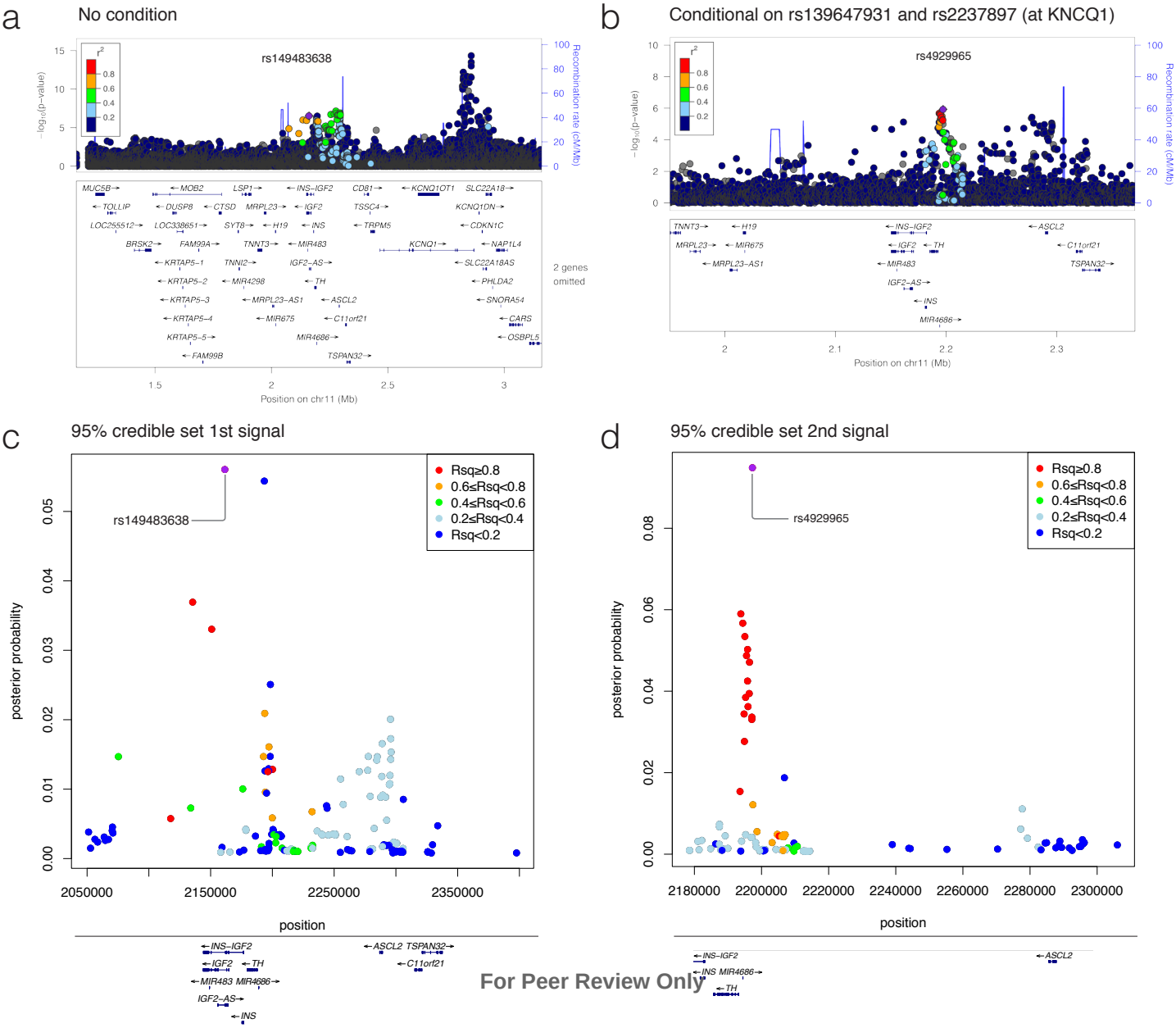
691 Sakoda LC, Shapero M, Shen L, Shenoy T, Smethurst D, Somkin CP, Van Den Eeden SK, Walter L,
692 Wan E, Webster T, Whitmer RA, Wong S, Zau C, Zhan Y, Schaefer C, Kwok PY, Risch N: Genotyping
693 Informatics and Quality Control for 100,000 Subjects in the Genetic Epidemiology Research on
694 Adult Health and Aging (GERA) Cohort. Genetics 2015;200:1051-1060
695 37. Delaneau O, Marchini J, Zagury JF: A linear complexity phasing method for thousands of
696 genomes. Nature methods 2012;9:179-181
697 38. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO,
698 Marchini JL, McCarthy S, McVean GA, Abecasis GR: A global reference for human genetic variation.
699 Nature 2015;526:68-74
700 39. Goldstein JI, Crenshaw A, Carey J, Grant GB, Maguire J, Fromer M, O'Dushlaine C, Moran JL,
701 Chambert K, Stevens C, Swedish Schizophrenia C, Consortium AAS, Sklar P, Hultman CM, Purcell S,
702 McCarroll SA, Sullivan PF, Daly MJ, Neale BM: zCall: a rare variant caller for array-based
703 genotyping: genetics and population analysis. Bioinformatics 2012;28:2543-2545
704 40. Wakefield J: A Bayesian measure of the probability of false discovery in genetic epidemiology
705 studies. Am J Hum Genet 2007;81:208-227
706 41. Saxena R, Elbers CC, Guo Y, Peter I, Gaunt TR, Mega JL, Lanktree MB, Tare A, Castillo BA, Li YR,
707 Johnson T, Bruinenberg M, Gilbert-Diamond D, Rajagopalan R, Voight BF, Balasubramanyam A,
708 Barnard J, Bauer F, Baumert J, Bhangale T, Bohm BO, Braund PS, Burton PR, Chandrupatla HR,
709 Clarke R, Cooper-DeHoff RM, Crook ED, Davey-Smith G, Day IN, de Boer A, de Groot MC, Drenos F,
710 Ferguson J, Fox CS, Furlong CE, Gibson Q, Gieger C, Gilhuijs-Pederson LA, Glessner JT, Goel A, Gong
711 Y, Grant SF, Grobbee DE, Hastie C, Humphries SE, Kim CE, Kivimaki M, Kleber M, Meisinger C,
712 Kumari M, Langaee TY, Lawlor DA, Li M, Lobmeyer MT, Maitland-van der Zee AH, Meijs MF,
713 Molony CM, Morrow DA, Murugesan G, Musani SK, Nelson CP, Newhouse SJ, O'Connell JR,
714 Padmanabhan S, Palmen J, Patel SR, Pepine CJ, Pettinger M, Price TS, Rafelt S, Ranchalis J, Rasheed
715 A, Rosenthal E, Ruczinski I, Shah S, Shen H, Silbernagel G, Smith EN, Spijkerman AW, Stanton A,
716 Steffes MW, Thorand B, Trip M, van der Harst P, van der AD, van Iperen EP, van Setten J, van Vliet-
717 Ostaptchouk JV, Verweij N, Wolffenbuttel BH, Young T, Zafarmand MH, Zmuda JM, Look ARG,
718 consortium D, Boehnke M, Altshuler D, McCarthy M, Kao WH, Pankow JS, Cappola TP, Sever P,
719 Poulter N, Caulfield M, Dominiczak A, Shields DC, Bhatt DL, Zhang L, Curtis SP, Danesh J, Casas JP,
720 van der Schouw YT, Onland-Moret NC, Doevendans PA, Dorn GW, 2nd, Farrall M, FitzGerald GA,
721 Hamsten A, Hegele R, Hingorani AD, Hofker MH, Huggins GS, Illig T, Jarvik GP, Johnson JA, Klungel
722 OH, Knowler WC, Koenig W, Marz W, Meigs JB, Melander O, Munroe PB, Mitchell BD, Bielinski SJ,
723 Rader DJ, Reilly MP, Rich SS, Rotter JI, Saleheen D, Samani NJ, Schadt EE, Shuldiner AR, Silverstein
724 R, Kottke-Marchant K, Talmud PJ, Watkins H, Asselbergs FW, de Bakker PI, McCaffery J, Wijmenga
725 C, Sabatine MS, Wilson JG, Reiner A, Bowden DW, Hakonarson H, Siscovick DS, Keating BJ: Large-
726 scale gene-centric meta-analysis across 39 studies identifies type 2 diabetes loci. Am J Hum Genet
727 2012;90:410-425
728 42. Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB: Rare variants create synthetic
729 genome-wide associations. PLoS Biol 8:e1000294

730

Supplementary Figure 1

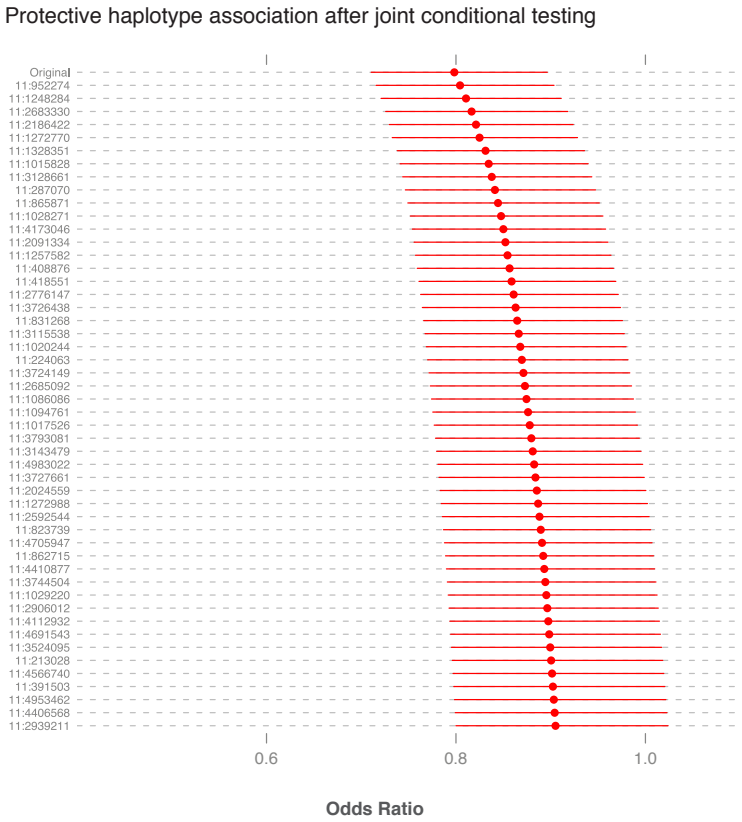


Supplementary Figure 2

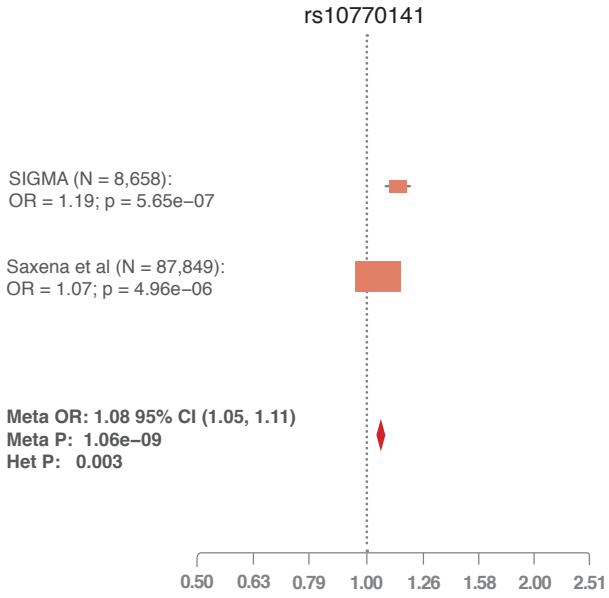


For Peer Review Only

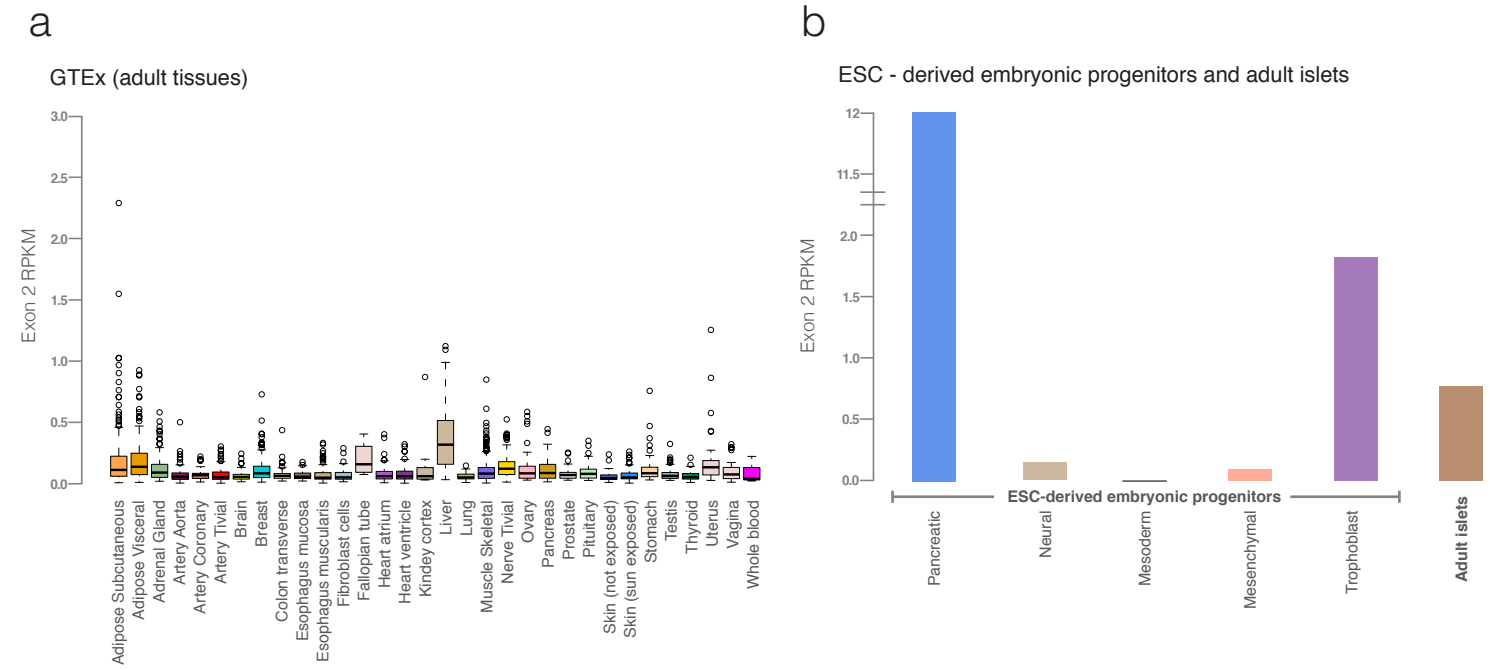
Supplementary Figure 3



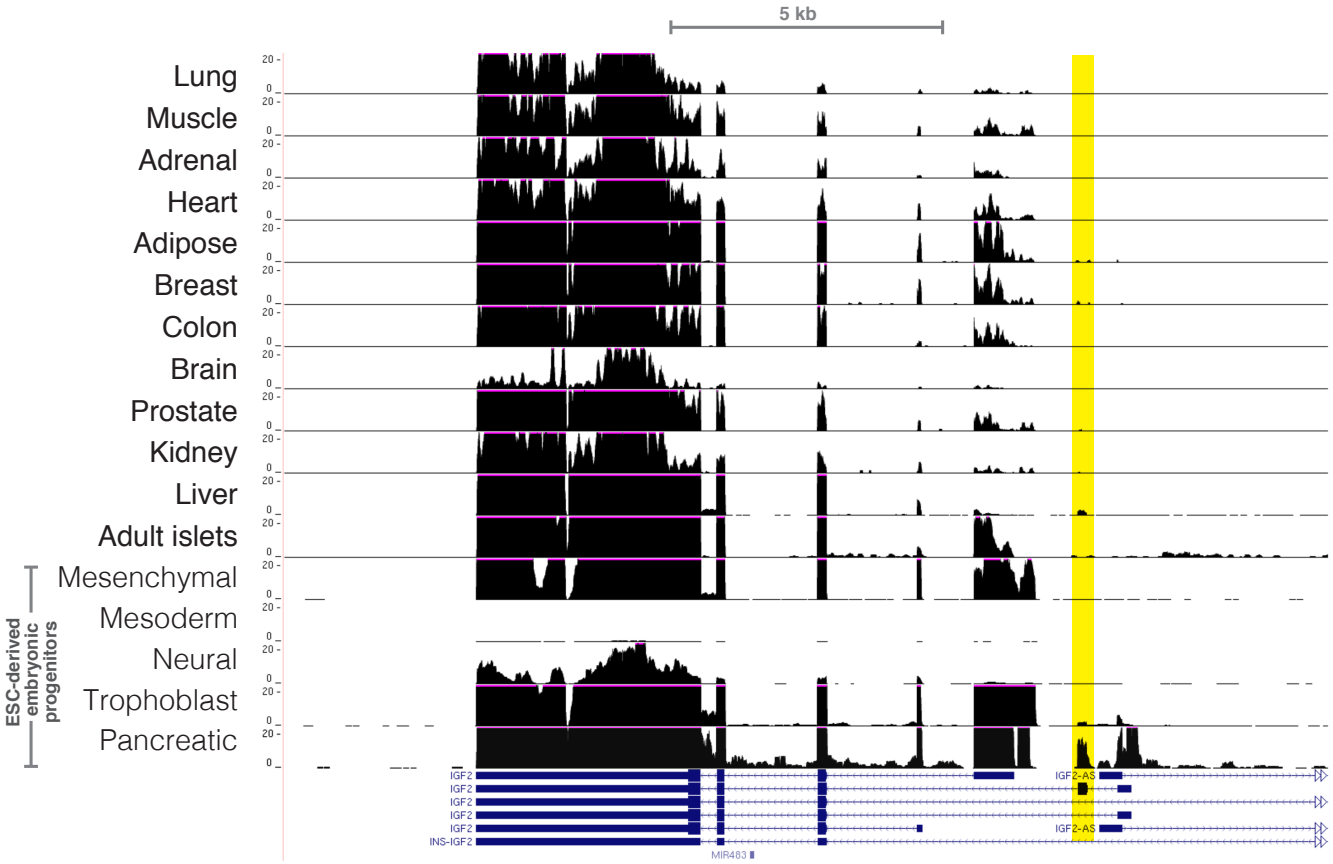
Supplementary Figure 4



Supplementary Figure 5



Supplementary Figure 6

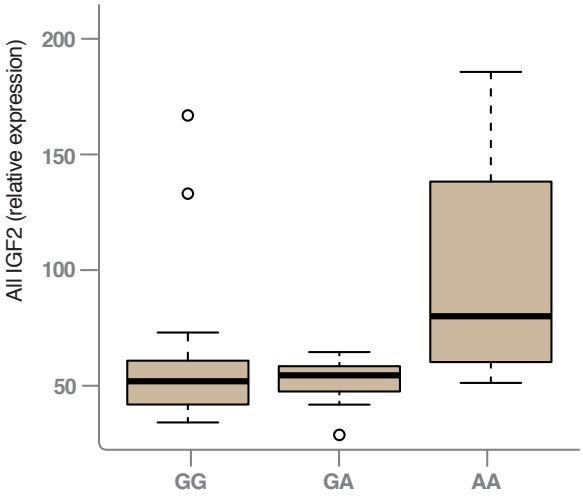


Supplementary Figure 7

a

Liver (n = 32)

rho = 0.26; spearman P-value = 0.1547

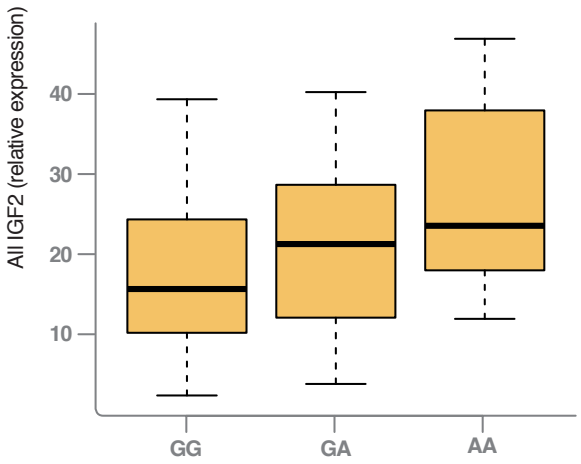


rs149483638

b

Adipose (n = 132)

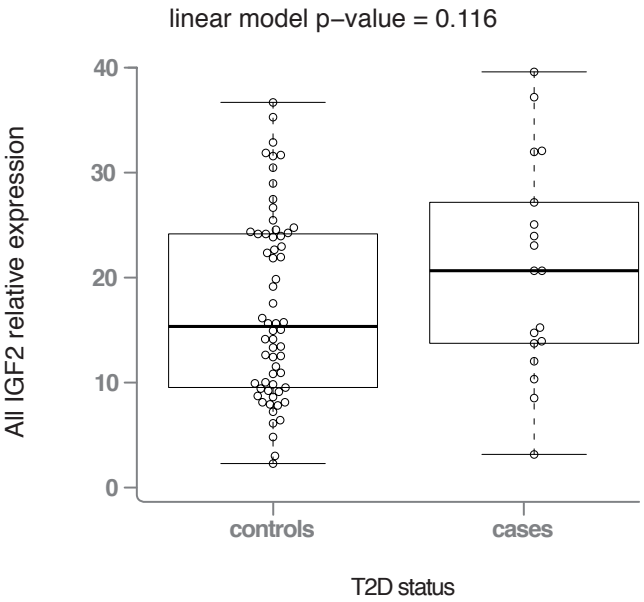
rho = 0.17; spearman P-value= 0.0553



rs149483638

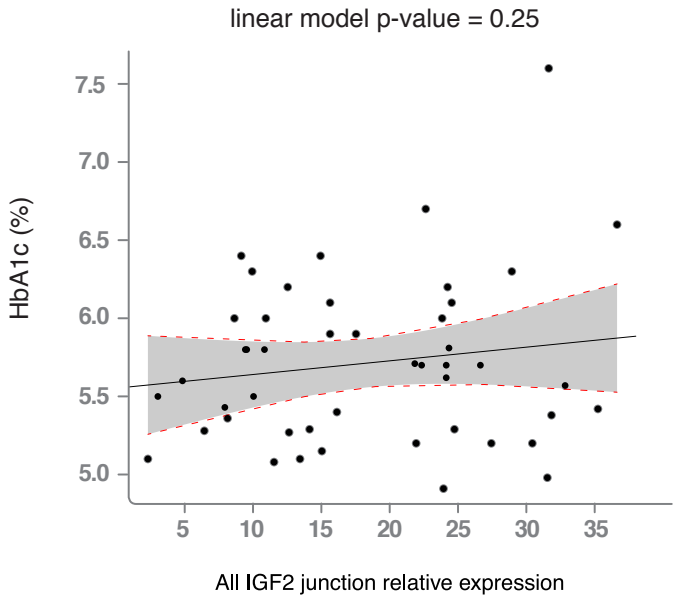
c

All IGF2 in adipose tissue by Type 2 Diabetes status in GG subjects; (n = 81)

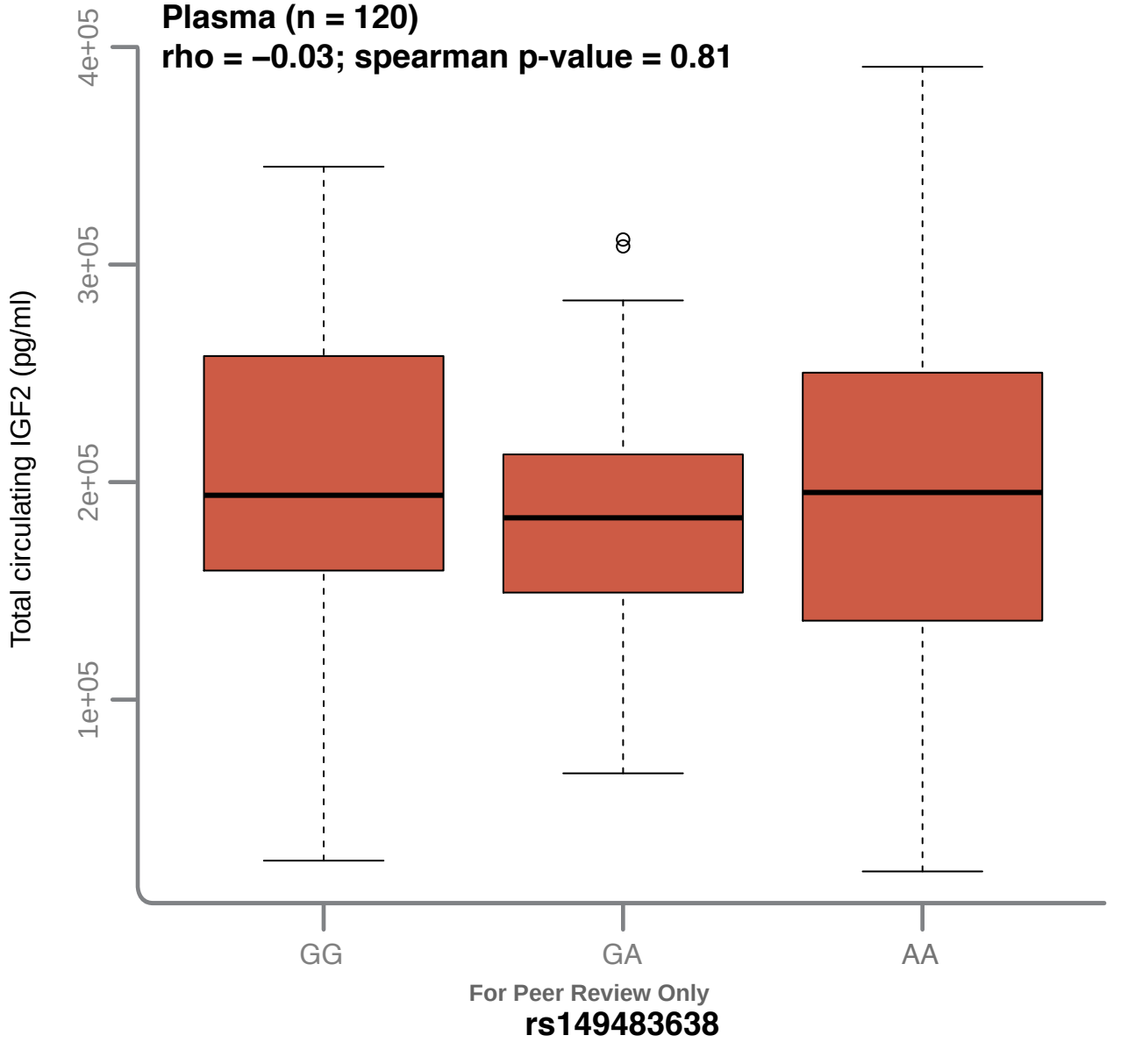


d

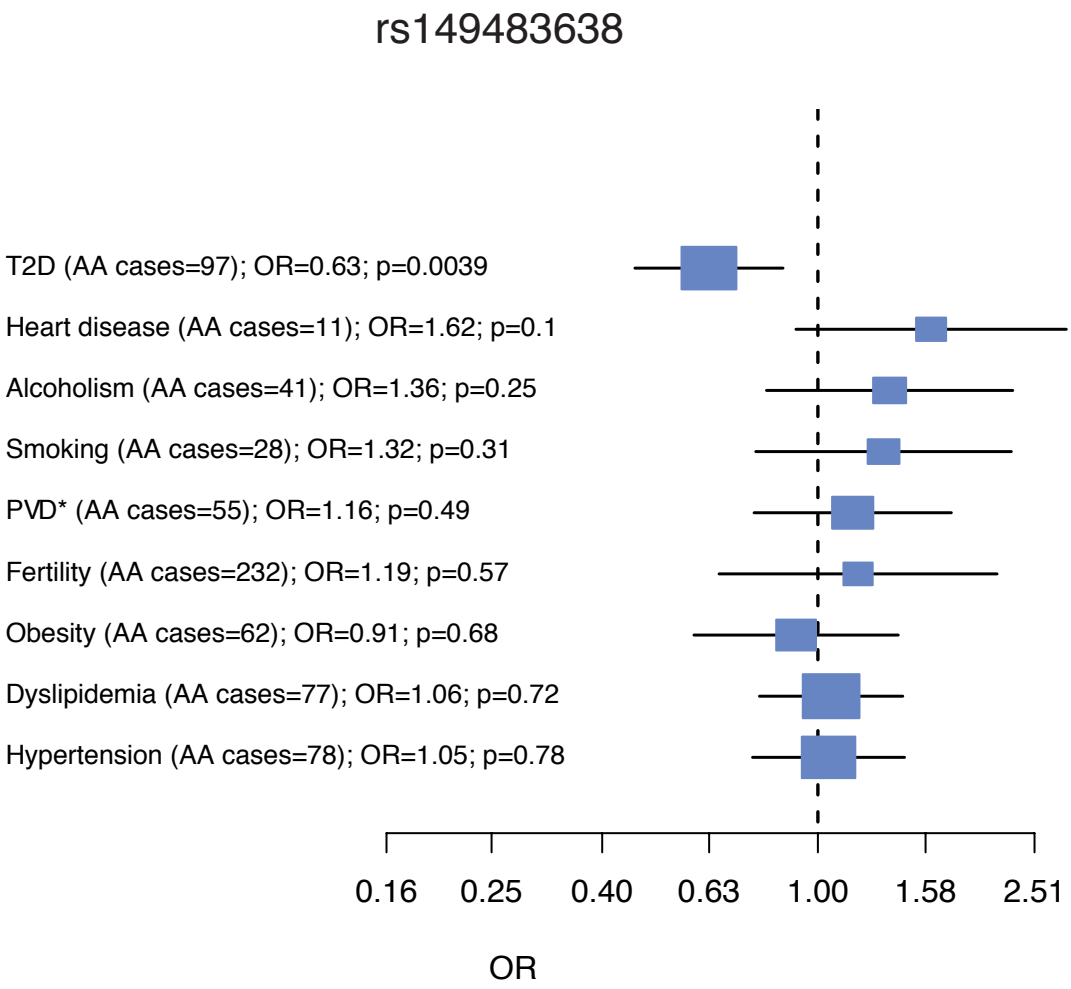
HbA1c vs all IGF2 adipose tissue expression in controls GG; (n = 47)



Supplementary Figure 8



For Peer Review Only


rs149483638
T2D (AA cases=97); OR=0.63; p=0.0039
Heart disease (AA cases=11); OR=1.62; p=0.1
Alcoholism (AA cases=41); OR=1.36; p=0.25
Smoking (AA cases=28); OR=1.32; p=0.31
PVD* (AA cases=55); OR=1.16; p=0.49
Fertility (AA cases=232); OR=1.19; p=0.57
Obesity (AA cases=62); OR=0.91; p=0.68
Dyslipidemia (AA cases=77); OR=1.06; p=0.72
Hypertension (AA cases=78); OR=1.05; p=0.78
0.16
0.25
0.40
0.63
1.00
1.58
2.51
OR

Supplementary Table 1: Top hits identified by exome chip association analysis. Odds ratios and p-values where derived by Wald's test, after adjusting for BMI, the first 10 principal components to adjust for population stratification and body mass index (BMI), age and sex. Directional consistency with previous studies is also represented.

rsid	Chr	Position	Closest Gene	Consequence	Reference Allele	Alternative Allele	MAF Affected	MAF Unnaffected	Number Effective Samples	Odds Ratio Wald Test adjusted	*p*-value Wald Test corrected	Odds ratio directionally consistent?
rs7903146	10	114758349	*TCF7L2*	intron variant	C	T	0.24	0.20	8622	1.38	4.76E-17	+
rs7901695	10	114754088	*TCF7L2*	intron variant	T	C	0.25	0.22	8622	1.31	5.03E-13	+
rs4506565	10	114756041	*TCF7L2*	intron variant	A	T	0.25	0.22	8617	1.31	6.63E-13	+
rs12243326	10	114788815	*TCF7L2*	intron variant	T	C	0.19	0.17	8622	1.31	6.23E-11	+
rs2237892	11	2839751	*KCNQ1*	intron variant	C	T	0.26	0.29	8622	0.80	2.46E-10	+
rs13342692	17	6946287	*SLC16A11*	missense variant	T	C	0.35	0.29	8622	1.24	4.75E-10	+
rs117767867	17	6946330	*SLC16A11*	missense variant	C	T	0.33	0.27	8620	1.22	7.33E-09	+
rs149483638	11	2161530	*INS-IGF2*	splice acceptor variant	C	T	0.17	0.18	8622	0.80	1.36E-07	novel
rs849134	7	28196222	*JAZF1*	intron variant	A	G	0.32	0.37	8622	0.84	2.63E-07	+
rs864745	7	28180556	*JAZF1*	intron variant	T	C	0.32	0.37	8610	0.85	3.97E-07	+
rs1635852	7	28189411	*JAZF1*	intron variant	T	C	0.32	0.37	8611	0.85	6.49E-07	+
rs2184898	10	119418104	*EMX2*	-	G	A	0.25	0.23	8613	1.20	1.12E-06	novel
rs2237895	11	2857194	*KCNQ1*	intron variant	A	C	0.47	0.43	8622	1.16	2.86E-06	+
rs1421085	16	53800954	*FTO*	intron variant	T	C	0.23	0.21	8621	1.18	1.48E-05	+
rs1558902	16	53803574	*FTO*	intron variant	T	A	0.23	0.21	8621	1.18	1.59E-05	+

Supplementary Table 2. Quantitative trait association results for rs149483638 in the discovery cohorts. P-values and effect sizes, and standard error of rs149483638-G allele with quantitative traits in non-diabetic individuals are represented.

Trait	Effective sample size	MAF	additive *p*-value	additive beta	additive S.E. beta
Fasting Glucose	2215	0.2043	0.3	0.0057	0.0053
Fasting Insulin	1515	0.2071	0.8	-0.0100	0.0381
2-hour glucose challenge	535	0.1998	0.3	0.0413	0.0392
Glycated Hemogoglobin	544	0.1998	0.1	0.0649	0.0418
Cholesterol	2210	0.2043	0.5	0.0249	0.0348
LDLC	1522	0.2066	0.9	-0.0050	0.0427
HDLC	1597	0.2053	0.8	-0.0105	0.0437
TG	2210	0.2043	0.8	0.0396	0.0297
BMI	4364	0.1862	0.2	0.0382	0.0272

LDLC: low-density lipoprotein cholesterol, HDLC: High-density lipoprotein cholesterol, TG: Triglycerides, BMI: Body Mass Index

Supplementary Table 3. Variants in linkage disequilibrium with rs149483638 (R-squared higher than 0.5) and their predicted functional impact

						Variant Effect Predictor (most severe consequence)		
rsid	**chr:position(bp)**	**Allele1**	**Allele2**	**R-squared with top variant**	**postProb**	**Consequence**	**IMPACT**	**SYMBOL**
rs149483638	chr11:2161530	T	C	1.000	0.056	splice_acceptor_variant	HIGH	IGF2
rs144656014	chr11:2135474	A	G	0.823	0.037	intergenic_variant	MODIFIER	-
rs11564732	chr11:2150895	T	C	0.852	0.033	3_prime_UTR_variant	MODIFIER	IGF2
rs10840490	chr11:2193817	C	G	0.619	0.021	upstream_gene_variant	MODIFIER	TH | MIR4686
rs34779113	chr11:2197234	G	GC	0.625	0.016	regulatory_region_variant	MODIFIER	TH | MIR4686
rs10840489	chr11:2192798	T	C	0.602	0.015	intron_variant	MODIFIER	TH | MIR4686
rs146043837	chr11:2075378	A	C	0.550	0.015	intergenic_variant	MODIFIER	-
rs187839678	chr11:2200156	T	C	0.841	0.013	regulatory_region_variant	MODIFIER	-
rs192912194	chr11:2196425	A	G	0.854	0.013	upstream_gene_variant	MODIFIER	TH | MIR4686
rs10840491	chr11:2194390	A	G	0.630	0.010	regulatory_region_variant	MODIFIER	TH | MIR4686
rs140996354	chr11:2232137	T	C	0.703	0.007	regulatory_region_variant	MODIFIER	-
rs11042982	chr11:2199963	C	G	0.603	0.006	regulatory_region_variant	MODIFIER	-
rs80089797	chr11:2117677	T	C	0.826	0.006	intergenic_variant	MODIFIER	-
rs6578993	chr11:2201163	T	C	0.595	0.003	intergenic_variant	MODIFIER	-
rs11043001	chr11:2203045	A	G	0.569	0.003	intergenic_variant	MODIFIER	-
rs7126800	chr11:2202668	A	C	0.586	0.002	intergenic_variant	MODIFIER	-
rs7925375	chr11:2191155	T	C	0.531	0.002	intron_variant	MODIFIER	TH | MIR4686
rs12224088	chr11:2217358	C	G	0.507	0.001	intergenic_variant	MODIFIER	-
rs147580690	chr11:2216970	CA	C	0.515	0.001	regulatory_region_variant	MODIFIER	-

Supplementary Table 4. Association results for each of the cohorts and meta-analysis using both inverse-variance fixed effects model and sample size model.

Dataset	N	OR (95% CI)	p-value
SIGMA	8,658	0.8 (0.74-0.87)	$1.14x10^{-7}$
Pima	3,199	0.68 (0.57-0.81)	$1.09x10^{-05}$
SAFS*	2,982	Z = -2.3	0.021
T2D-GENES-HS	1,924	0.89 (0.7-1.13)	0.326
DMS2	1,228	0.71 (0.58-0.88)	0.001
GERA	5,896	0.82 (0.64-1.05)	0.11
DPP*	616	HR=0.76 (0.49-1.20)	0.24

	OR (95%CI)	Z-score	P-value
METAL (IVFE)	0.78 (0.73-0.84)	-	$5.61x10^{-14}$
METAL (SAMPLE SIZE)	-	-7.53	$4.78x10^{-14}$

*SAFS where analyzed by a family-based association test (FBAT) and where not included in the inverse-variance fixed effects meta-analyses. Since diabetes incidence in the DPP was computed by Hazard Ratios (HR), the results from the DPP were only included in the sample size meta-analysis.

Supplementary Table 5. Family-based association analysis in the San Antonio Families Study

Marker	Allele	afreq	fam#	S-E(S)	Var(S)	Z	*p*
rs11564732**	A	0.13	283	-35.259	234.871	-2.301	0.02

*Residuals where computed after regressing out the age+age^2+BMI.
**rs11564732* was used as a proxy for rs149483638 (Rsq= 0.85, 10.6 kb). A allele in rs11564732 co-segregates with the T allele in rs149483638

Supplementary Table 6. Association of human knockouts for isoform 2 of *IGF2* (AA homozygous) with other diseases or clinical outcomes.

Clinical Outcome	OR	Odds Ratio (95%CI)	P	HetTest	cohorts tested*	total AA cases	total AA controls	total GG cases	total GG controls
T2D Status	0.63	0.63 (0.46-0.86)	0.0039	0.19	DMS1+DMS2+GERA	97	195	1534	4873
Dyslipidemia	1.06	1.06 (0.78-1.43)	0.72	0.69	DMS1+DMS2+GERA	77	210	3096	3281
Fertility	1.19	1.19 (0.66-2.14)	0.57	0.82	DMS1+DMS2	232	24	914	263
Heart disease	1.62	1.62 (0.91-2.87)	0.1	0.24	DMS1+DMS2+GERA	11	273	1200	5133
Hypertension	1.05	1.05 (0.76-1.44)	0.78	0.029	DMS1+DMS2+GERA	78	210	2970	3427
Obesity	0.91	0.91 (0.59-1.41)	0.68	0.13	DMS1+DMS2	62	178	320	644
Peripheral Vascular Disease	1.16	1.16 (0.76-1.76)	0.49	0.078	DMS1+DMS2	55	192	241	856
Smoking	1.32	1.32 (0.77-2.27)	0.31	0.79	DMS1+DMS2	28	220	182	923

*The results were meta-analyzed, when possible with all the available cohorts with that clinical outcome available. Association analyses were corrected for age, BMI, and sex, and the first two principal components to correct for population stratification. Inverse-variant fixed effects meta-analysis was performed. Only clinical conditions with at least 10 AA carriers are presented.

Supplementary table 7. Association of rs149483638 variant with other diseases assessed in the Genetic Epidemiology Research on Aging (GERA). Effect sizes are considering the A allele as effect allele.

Clinical condition	cases (n)	controls (n)	cases MAF	controls MAF	p-value	OR
Varicose veins	259	5637	0.07	0.06	0.506	1.15
Cancer	1127	4769	0.04	0.06	0.522	0.92
Cardiovascular disease	1300	4596	0.05	0.06	0.844	0.98
Depression	887	5009	0.05	0.06	0.954	0.99
Dermathophytosis	997	4899	0.07	0.06	0.760	1.03
Type 2 diabetes	8,227	12,966	0.17	0.18	$5.6x10^{-14}$	0.78
Dyslipidemia	3149	2747	0.06	0.06	0.701	0.96
Hemorrhoids	974	4922	0.06	0.06	0.500	1.08
Hernia abdominopelvic cavity	557	5339	0.06	0.06	0.388	1.15
Hypertensive Disease	2921	2975	0.06	0.06	0.988	1.00
Insomnia	390	5506	0.05	0.06	0.876	1.03
Iron deficiency anemias	299	5597	0.07	0.06	0.277	1.23
Irritable bowel syndrome	362	5534	0.04	0.06	0.006	0.53
Macular Degeneration	225	5671	0.04	0.06	0.616	0.86
Osteoarthritis	1961	3935	0.05	0.06	0.705	1.04
Osteoporosis	450	5446	0.05	0.06	0.810	0.96
Psychiatric: any	1154	4742	0.06	0.06	0.439	1.09
Peripheral Vascular Disease	347	5549	0.05	0.06	0.834	1.05
Acute reaction to stress	662	5234	0.06	0.06	0.591	1.08

* Association analyses were corrected for age, BMI, and sex, and the first two principal components to correct for population stratification. For type 2 diabetes, the meta-analysis results of the discovery and replication cohorts are presented.

The T2D-GENES Consortium:

The Broad Genomics Platform[1], Gonçalo Abecasis[2], Marcio Almeida[3], David Altshuler[4,5,6,7,8,9,10], Jennifer L. Asimit[11], Gil Atzmon[12], Mathew Barber[13], Nicola L. Beer[14], Graeme I. Bell[13,15], Jennifer Below[16], Tom Blackwell[2], John Blangero[3], Michael Boehnke[2], Donald W. Bowden[17,18,19,20], Noël Burtt[4], John Chambers[21,22,23], Han Chen[24], Peng Chen[25], Peter S.Chines[26], Sungkyoung Choi[27], Claire Churchhouse[4], Pablo Cingolani[28], Belinda K. Cornes[29], Nancy Cox[13,15], Aaron G. Day-Williams[11], Ravindranath Duggirala[3], Josée Dupuis[24], Thomas Dyer[3], Shuang Feng[2], Juan Fernandez-Tajes[30], Teresa Ferreira[30], Tasha E. Fingerlin[31], Jason Flannick[4,6], Jose Florez[4,6,7], Pierre Fontanillas[4], Timothy M. Frayling[32], Christian Fuchsberger[2], Eric R. Gamazon[15], Kyle Gaulton[30], Saurabh Ghosh, Anna Gloyn[14], Robert L. Grossman[15,33], Jason Grundstad[33], Craig Hanis[16], Allison Heath[33], Heather Highland[16], Momoko Hirokoshi[30], Ik-Soo Huh[27], Jeroen R. Huyghe[2], Kamran Ikram[34,29,35,36], Kathleen A. Jablonski[37], Young Jin Kim[38], Goo Jun[2], Norihiro Kato[39], Jayoun Kim[27], C. Ryan King[40], Jaspal Kooner[22,23,41], Min Seok Kwon[27], Hae Kyung Im[40], Markku Laakso[42] , Kevin Koi-Yau Lam[25], Jaehoon Lee[27], Selyeong Lee[27], Sungyoung Lee[27], Donna M. Lehman[43], Heng Li[4], Cecilia M. Lindgren[30], Xuanyao Liu[25,44], Oren E. Livne[13], Adam E. Locke[2], Anubha Mahajan[30], Julian B. Maller[30,45], Alisa K. Manning[4]. Taylor J. Maxwell[16], Alexander Mazoure[46], Mark I. McCarthy[30,14,47], James B. Meigs[7,48], Byungju Min[27], Karen L. Mohlke[49], Andrew Morris[50], Solomon Musani[51], Yoshihiko Nagai[46], Maggie C.Y. Ng[17,18], Dan Nicolae[13,15,52], Sohee Oh[27], Nicholette Palmer[17,18,19], Taesung Park[27], Toni I. Pollin[53], Inga Prokopenko[30,54], David Reich[4,5], Manuel A. Rivas[30], Laura J. Scott[2], Mark Seielstad[55], Yoon Shin Cho[56], E-Shyong Tai[34,25,57], Xueling Sim[2], Robert Sladek[46,58], Philip Smith[59], Ioanna Tachmazidou[11], Tanya M. Teslovich[2], Jason Torres[13,15], Vasily Trubetskoy[13,15], Sara M. Willems[60], Amy L. Williams[4,5], James G. Wilson[61], Steven Wiltshire[62], Sungho Won[63], Andrew R. Wood[32], Wang Xu[57], Yik Ying Teo[64,65,66,67,68], Joon Yoon[27], Jong-Young Lee[69], Matthew Zawistowski[2], Eleftheria Zeggini[11], Weihua Zhang[21], Sebastian Zöllner[2,70]

[1]The Genomics Platform, The Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA.

[2]Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan 48109, USA.

[3]Department of Genetics, Texas Biomedical Research Institute, San Antonio, Texas 78227, USA.

[4]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA.

[5]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA.

[6]Center for Human Genetic Research and Diabetes Research Center (Diabetes Unit), Massachusetts General Hospital, Boston 02114, Massachusetts, USA.

[7]Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA.

[8]Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.

[9]Department of Molecular Biology, Harvard Medical School, Boston, Massachusetts 02114, USA.

[10]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

[11]Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1HH, UK.

[12]Department of Medicine, Department of Genetics, Albert Einstein College of Medicine, Bronx, New York 10461, USA.

[13]Department of Human Genetics, University of Chicago, Chicago, Illinois 60637, USA.

[14]Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, OX3 7LJ, UK.

[15]Department of Medicine, University of Chicago, Chicago, Illinois 60637, USA.

[16]Human Genetics Center, University of Texas Health Science Center at Houston, Houston, Texas 77030, USA.

[17]Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston- Salem, North Carolina 27157, USA.

[18]Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, North Carolina 27157, USA.

[19]Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, North Carolina 27157, USA.

[20]Internal Medicine-Endocrinology, Wake Forest School of Medicine, Winston-Salem, North Carolina 27157, USA.

[21]Department of Epidemiology and Biostatistics, Imperial College London, London SW7 2AZ, UK.

[22]Imperial College Healthcare NHS Trust, London W2 1NY, UK.

[23]Ealing Hospital National Health Service (NHS) Trust, Middlesex UB1 3HW, UK.

[24]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts 02115, USA.

[25]Saw Swee Hock School of Public Health, National University of Singapore, Singapore 117597, Singapore.

[26]National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA.

[27]Seoul National University, Seoul 110-799, South Korea.

[28]McGill Centre for Bioinformatics, McGill University, Montréal, Quebec, H3G 0B1, Canada.

[29]Singapore Eye Research Institute, Singapore National Eye Centre, Singapore 168751, Singapore.

[30]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK.

[31]Department of Epidemiology, Colorado School of Public Health, Aurora, Colorado 80045, USA.

[32]Genetics of Complex Traits, University of Exeter Medical School, Exeter, EX4 4SB UK.

[33]Institute for Genomics and Systems Biology, University of Chicago, Chicago,

Illinois 60637, USA.

[34]Duke National University of Singapore Graduate Medical School, Singapore 169857, Singapore.

[35]Department of Ophthalmology, National University of Singapore and National University Health System, Singapore 119228, Singapore.

[36]Department of Ophthalmology, Erasmus Medical Center, Rotterdam 3000 CA, the Netherlands.

[37]The Biostatistics Center, George Washington University, Rockville, Maryland 20852, USA.

[38]Department of Neurology, Konkuk University School of Medicine, Seoul 143-701, South Korea.

[39]Department of Gene Diagnostics and Therapeutics, Research Institute, National Center for Global Health and Medicine, Tokyo 162-8655, Japan.

[40]Department of Health Studies, University of Chicago, Chicago, Illinois 60637, USA.

[41]National Heart and Lung Institute (NHLI), Imperial College London, Hammersmith Hospital, London W12 0HS, UK.

[42] Department of Medicine, University of Eastern Finland, Kuopio Campus and Kuopio University Hospital, FI-70211 Kuopio, Finland.

[43]Division of Clinical Epidemiology, Department of Medicine, University of Texas Health Science Center at San Antonio, San Antonio, Texas 78229, USA.

[44]Graduate School for Integrative Science and Engineering, National University of Singapore, Singapore 117456, Singapore.

[45]Department of Statistics, University of Oxford, Oxford, OX1 3TG UK.

[46]McGill University, Montréal, Québec H3A 0G4, Canada.

[47]Oxford NIHR Biomedical Research Centre, Churchill Hospital, Headington OX3 7LE, UK.

[48]General Medicine Division, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.

[49]Department of Genetics, University of North Carolina-Chapel Hill, Chapel Hill, North Carolina 27599, USA.

[50]Department of Genetic Medicine, Manchester Academic Health Sciences Centre, Manchester M13 9NT, UK.

[51]Department of Medicine, University of Mississippi Medical Center, Jackson, Mississippi 39126, USA.

[52]Department of Statistics, University of Chicago, Chicago, Illinois 60637, USA.

[53]Department of Medicine, Program for Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA.

[54]Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, 751 05 Uppsala, Sweden.

[55]University of California San Francisco, San Francisco, California 94143, USA.

[56]Department of Biomedical Science, Hallym University, Chuncheon, Gangwon-do, 200-702 South Korea.

[57]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117597, Singapore.

[58]Department of Medicine, Royal Victoria Hospital, Montréal, Québec H3A 1A1, Canada.

[59]National Institute of Diabetes and Digestive and Kidney Disease, National Institutes of Health, Bethesda, MD 20817, USA.

[60]Department of Genetic Epidemiology, Erasmus Medical Center, Rotterdam 3000 CA, the Netherlands.
[61]Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, Mississippi 39216, USA.
[62]Centre for Medical Research, Western Australian Institute for Medical Research, The University of Western Australia, Nedlands WA 6008, Australia.
[63]Chung-Ang University, Seoul 156-756, South Korea.
[64]Department of Epidemiology and Public Health, National University of Singapore, Singapore 117597, Singapore.
[65]Centre for Molecular Epidemiology, National University of Singapore, Singapore 117456, Singapore.
[66]Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore 138672, Singapore.
[67]Graduate School for Integrative Science and Engineering, National University of Singapore, Singapore 117456, Singapore.
[68]Department of Statistics and Applied Probability, National University of Singapore, Singapore 117546, Singapore.
[69]Center for Genome Science, Korea National Institute of Health, Osong Health Technology Administration Complex, Chungcheongbuk-do, 363-951, South Korea.
[70]Department of Psychiatry, University of Michigan, Ann Arbor, Michigan 48109, USA.

DPP Research Group Investigators:

Pennington Biomedical Research Center (Baton Rouge, LA)
George A. Bray, MD*
Iris W. Culbert, BSN, RN, CCRC**
Catherine M. Champagne, PhD, RD
Barbara Eberhardt, RD, LDN
Frank Greenway, MD
Fonda G. Guillory, LPN
April A. Herbert, RD
Michael L. Jeffirs, LPN
Betty M. Kennedy, MPA
Jennifer C. Lovejoy, PhD
Laura H. Morris, BS
Lee E. Melancon, BA, BS
Donna Ryan, MD
Deborah A. Sanford, LPN
Kenneth G. Smith, BS, MT
Lisa L. Smith, BS
Julia A. St.Amant, RTR
Richard T. Tulley, PhD
Paula C. Vicknair, MS, RD
Donald Williamson, PhD
Jeffery J. Zachwieja, PhD
University of Chicago **(Chicago, IL)**
Kenneth S. Polonsky, MD*
Janet Tobian, MD, PhD*
David Ehrmann, MD*

Margaret J. Matulik, RN, BSN**
Bart Clark, MD
Kirsten Czech, MS
Catherine DeSandre, BA
Ruthanne Hilbrich, RD
Wylie McNabb, EdD
Ann R. Semenske, MS, RD

Jefferson Medical College **(Philadelphia, PA)**

Jose F. Caro, MD*
Pamela G. Watson, RN, ScD*
Barry J. Goldstein, MD, PhD*
Kellie A. Smith, RN, MSN**
Jewel Mendoza, RN, BSN**
Renee Liberoni, MPH
Constance Pepe, MS, RD
John Spandorfer, MD

University of Miami **(Miami, FL)**

Richard P. Donahue, PhD*
Ronald B. Goldberg, MD*
Ronald Prineas, MD, PhD*
Patricia Rowe, MPA**
Jeanette Calles, MSEd
Paul Cassanova-Romero, MD
Hermes J. Florez, MD
Anna Giannella, RD, MS
Lascelles Kirby, MS
Carmen Larreal
Valerie McLymont, RN
Jadell Mendez
Juliet Ojito, RN
Arlette Perry, PhD
Patrice Saab, PhD

The University of Texas Health Science Center **(San Antonio, TX)**

Steven M. Haffner, MD, MPH*
Maria G. Montez, RN, MSHP, CDE**
Carlos Lorenzo, MD, PhD
Arlene Martinez, RN, BSN, CDE

University of Colorado **(Denver, CO)**

Richard F. Hamman, MD, DrPH*
Patricia V. Nash, MS**
Lisa Testaverde, MS**
Denise R. Anderson, RN, BSN
Larry B. Ballonoff, MD
Alexis Bouffard, MA,
B. Ned Calonge, MD, MPH
Lynne Delve
Martha Farago, RN
James O. Hill, PhD
Shelley R. Hoyer, BS
Bonnie T. Jortberg, MS, RD, CDE
Dione Lenz, RN, BSN
Marsha Miller, MS, RD
David W. Price, MD
Judith G. Regensteiner, PhD

Helen Seagle, MS, RD
Carissa M. Smith, BS
Sheila C. Steinke, MS
Brent VanDorsten, PhD
Joslin Diabetes Center **(Boston, MA)**
Edward S. Horton, MD*
Kathleen E. Lawton, RN**
Ronald A. Arky, MD
Marybeth Bryant
Jacqueline P. Burke, BSN
Enrique Caballero, MD
Karen M. Callaphan, BA
Om P. Ganda, MD
Therese Franklin
Sharon D. Jackson, MS, RD, CDE
Alan M. Jacobsen, MD
Lyn M. Kula, RD
Margaret Kocal, RN, CDE
Maureen A. Malloy, BS
Maryanne Nicosia, MS, RD
Cathryn F. Oldmixon, RN
Jocelyn Pan, BS, MPH
Marizel Quitingon
Stacy Rubtchinsky, BS
Ellen W. Seely, MD
Dana Schweizer, BSN
Donald Simonson, MD
Fannie Smith, MD
Caren G. Solomon, MD, MPH
James Warram, MD
VA Puget Sound Health Care System and University of Washington **(Seattle, WA)**
Steven E. Kahn, MB, ChB*
Brenda K. Montgomery, RN, BSN, CDE**
Wilfred Fujimoto, MD
Robert H. Knopp, MD
Edward W. Lipkin, MD
Michelle Marr, BA
Dace Trence, MD
University of Tennessee **(Memphis, TN)**
Abbas E. Kitabchi, PhD, MD, FACP*
Mary E. Murphy, RN, MS, CDE, MBA**
William B. Applegate, MD, MPH
Michael Bryer-Ash, MD
Sandra L. Frieson, RN
Raed Imseis, MD
Helen Lambeth, RN, BSN
Lynne C. Lichtermann, RN, BSN
Hooman Oktaei, MD
Lily M.K. Rutledge, RN, BSN
Amy R. Sherman, RD, LD
Clara M. Smith, RD, MHP, LDN
Judith E. Soberman, MD
Beverly Williams-Cleaves, MD
Northwestern University's Feinberg School of Medicine **(Chicago, IL)**

Boyd E. Metzger, MD*
Mariana K. Johnson, MS, RN**
Catherine Behrends
Michelle Cook, MS
Marian Fitzgibbon, PhD
Mimi M. Giles, MS, RD
Deloris Heard, MA
Cheryl K.H. Johnson, MS, RN
Diane Larsen, BS
Anne Lowe, BS
Megan Lyman, BS
David McPherson, MD
Mark E. Molitch, MD
Thomas Pitts, MD
Renee Reinhart, RN, MS
Susan Roston, RN, RD
Pamela A. Schinleber, RN, MS

***Massachusetts General Hospital* (Boston, MA)**

David M. Nathan, MD*
Charles McKitrick, BSN**
Heather Turgeon, BSN**
Kathy Abbott
Ellen Anderson, MS, RD
Laurie Bissett, MS, RD
Enrico Cagliero, MD
Jose C. Florez, MD, PhD+
Linda Delahanty, MS, RD
Valerie Goldman, MS, RD
Alexandra Poulos

***University of California-San Diego* (San Diego, CA)**

Jerrold M. Olefsky, MD*
Mary Lou Carrion-Petersen, RN, BSN**
Elizabeth Barrett-Connor, MD
Steven V. Edelman, MD
Robert R. Henry, MD
Javiva Horne, RD
Simona Szerdi Janesch, BA
Diana Leos, RN, BSN
Sundar Mudaliar, MD
William Polonsky, PhD
Jean Smith, RN
Karen Vejvoda, RN, BSN, CDE, CCRC

***St. Luke's-Roosevelt Hospital* (New York, NY)**

F. Xavier Pi-Sunyer, MD*
Jane E. Lee, MS**
David B. Allison, PhD
Nancy J. Aronoff, MS, RD
Jill P. Crandall, MD
Sandra T. Foo, MD
Carmen Pal, MD
Kathy Parkes, RN
Mary Beth Pena, RN
Ellen S. Rooney, BA
Gretchen E.H. Van Wye, MA

Kristine A. Viscovich, ANP
***Indiana University* (Indianapolis, IN)**
David G. Marrero, PhD*
Melvin J. Prince, MD*
Susie M. Kelly, RN, CDE**
Yolanda F. Dotson, BS
Edwin S. Fineberg, MD
John C. Guare, PhD
Angela M. Hadden
James M. Ignaut, MA
Marcia L. Jackson
Marion S. Kirkman, MD
Kieren J. Mather, MD
Beverly D. Porter, MSN
Paris J. Roach, MD
Nancy D. Rowland, BS, MS
Madelyn L. Wheeler, RD
***Medstar Research Institute* (Washington, DC)**
Robert E. Ratner, MD*
Gretchen Youssef, RD, CDE**
Sue Shapiro, RN, BSN, CCRC**
Catherine Bavido-Arrage, MS, RD, LD
Geraldine Boggs, MSN, RN
Marjorie Bronsord, MS, RD, CDE
Ernestine Brown
Wayman W. Cheatham, MD
Susan Cola
Cindy Evans
Peggy Gibbs
Tracy Kellum, MS, RD, CDE
Claresa Levatan, MD
Asha K. Nair, BS
Maureen Passaro, MD
Gabriel Uwaifo, MD
***University of Southern California/UCLA Research Center* (Alhambra, CA)**
Mohammed F. Saad, MD*
Maria Budget**
Sujata Jinagouda, MD**
Khan Akbar, MD
Claudia Conzues
Perpetua Magpuri
Kathy Ngo
Amer Rassam, MD
Debra Waters
Kathy Xapthalamous
***Washington University* (St. Louis, MO)**
Julio V. Santiago, MD* (deceased)
Samuel Dagogo-Jack, MD, MSc, FRCP, FACP*
Neil H. White, MD, CDE*
Samia Das, MS, MBA, RD, LD**
Ana Santiago, RD**
Angela Brown, MD
Edwin Fisher, PhD
Emma Hurt, RN

Tracy Jones, RN
Michelle Kerr, RD
Lucy Ryder, RN
Cormarie Wernimont, MS, RD

Johns Hopkins School of Medicine **(Baltimore, MD)**

Christopher D. Saudek, MD*
Vanessa Bradley, BA**
Emily Sullivan, MEd, RN**
Tracy Whittington, BS**
Caroline Abbas
Frederick L. Brancati, MD, MHS
Jeanne M. Clark, MD
Jeanne B. Charleston, RN, MSN
Janice Freel
Katherine Horak, RD
Dawn Jiggetts
Deloris Johnson
Hope Joseph
Kimberly Loman
Henry Mosley
Richard R. Rubin, PhD
Alafia Samuels, MD
Kerry J. Stewart, EdD
Paula Williamson

University of New Mexico **(Albuquerque, NM)**

David S. Schade, MD*
Karwyn S. Adams, RN, MSN**
Carolyn Johannes, RN, CDE**
Leslie F. Atler, PhD
Patrick J. Boyle, MD
Mark R. Burge, MD
Janene L. Canady, RN, CDE
Lisa Chai, RN
Ysela Gonzales, RN, MSN
Doris A. Hernandez-McGinnis
Patricia Katz, LPN
Carolyn King
Amer Rassam, MD
Sofya Rubinchik, MD
Willette Senter, RD
Debra Waters, PhD

Albert Einstein College of Medicine **(Bronx, NY)**

Harry Shamoon, MD*
Janet O. Brown, RN, MPH, MSN**
Elsie Adorno, BS
Liane Cox, MS, RD
Jill Crandall, MD
Helena Duffy, MS, C-ANP
Samuel Engel, MD
Allison Friedler, BS
Crystal J. Howard-Century, MA
Stacey Kloiber, RN
Nadege Longchamp, LPN
Helen Martinez, RN, MSN, FNP-C

Dorothy Pompi, BA
Jonathan Scheindlin, MD
Elissa Violino, RD, MS
Elizabeth Walker, RN, DNSc, CDE
Judith Wylie-Rosett, EdD, RD
Elise Zimmerman, RD, MS
Joel Zonszein, MD

University of Pittsburgh **(Pittsburgh, PA)**

Trevor Orchard, MD*
Rena R. Wing, PhD*
Gaye Koenning, MS, RD**
M. Kaye Kramer, BSN, MPH**
Susan Barr, BS
Miriam Boraz
Lisa Clifford, BS
Rebecca Culyba, BS
Marlene Frazier
Ryan Gilligan, BS
Susan Harrier, MLT
Louann Harris, RN
Susan Jeffries, RN, MSN
Andrea Kriska, PhD
Qurashia Manjoo, MD
Monica Mullen, MHP, RD
Alicia Noel, BS
Amy Otto, PhD
Linda Semler, MS, RD
Cheryl F. Smith, PhD
Marie Smith, RN, BSN
Elizabeth Venditti, PhD
Valarie Weinzierl, BS
Katherine V. Williams, MD, MPH
Tara Wilson, BA

University of Hawaii **(Honolulu, HI)**

Richard F. Arakaki, MD*
Renee W. Latimer, BSN, MPH**
Narleen K. Baker-Ladao, BS
Ralph Beddow, MD
Lorna Dias, AA
Jillian Inouye, RN, PhD
Marjorie K. Mau, MD
Kathy Mikami, BS, RD
Pharis Mohideen, MD
Sharon K. Odom, RD, MPH
Raynette U. Perry, AA

Southwest American Indian Centers **(Phoenix, AZ; Shiprock, NM; Zuni, NM)**

William C. Knowler, MD, DrPH*+
Norman Cooeyate**
Mary A. Hoskin, RD, MS**
Carol A. Percy, RN, MS**
Kelly J. Acton, MD, MPH
Vickie L. Andre, RN, FNP
Rosalyn Barber
Shandiin Begay, MPH

Peter H. Bennett, MB, FRCP
Mary Beth Benson, RN, BSN
Evelyn C. Bird, RD, MPH
Brenda A. Broussard, RD, MPH, MBA, CDE
Marcella Chavez, RN, AS
Tara Dacawyma
Matthew S. Doughty, MD
Roberta Duncan, RD
Cyndy Edgerton, RD
Jacqueline M. Ghahate
Justin Glass, MD
Martia Glass, MD
Dorothy Gohdes, MD
Wendy Grant, MD
Robert L. Hanson, MD, MPH
Ellie Horse
Louise E. Ingraham, MS, RD, LN
Merry Jackson
Priscilla Jay
Roylen S. Kaskalla
David Kessler, MD
Kathleen M. Kobus, RNC-ANP
Jonathan Krakoff, MD
Catherine Manus, LPN
Sara Michaels, MD
Tina Morgan
Yolanda Nashboo (deceased)
Julie A. Nelson, RD
Steven Poirier, MD
Evette Polczynski, MD
Mike Reidy, MD
Jeanine Roumain, MD, MPH
Debra Rowse, MD
Sandra Sangster
Janet Sewenemewa
Darryl Tonemah, PhD
Charlton Wilson, MD
Michelle Yazzie

George Washington University Biostatistics Center **(DPP Coordinating Center Rockville, MD)**

Raymond Bain, PhD*
Sarah Fowler, PhD*
Tina Brenneman**
Solome Abebe
Julie Bamdad, MS
Jackie Callaghan
Sharon L. Edelstein, ScM
Yuping Gao
Kristina L. Grimes
Nisha Grover
Lori Haffner, MS
Steve Jones
Tara L. Jones
Richard Katz, MD

John M. Lachin, ScD
Pamela Mucik
Robert Orlosky
James Rochon, PhD
Alla Sapozhnikova
Hanna Sherif, MS
Charlotte Stimpson
Marinella Temprosa, MS
Fredricka Walker-Murray

Central Biochemistry Laboratory **(Seattle, WA)**
Santica Marcovina, PhD, ScD*
Greg Strylewicz, PhD**
F. Alan Aldrich

Carotid Ultrasound
Dan O'Leary, MD*

CT Scan Reading Center
Elizabeth Stamm, MD*

Epidemiological Cardiology Research Center- Epicare **(Winston-Salem, NC)**
Pentti Rautaharju, MD, PhD*
Ronald J. Prineas, MD, PhD*/*
Teresa Alexander
Charles Campbell, MS
Sharon Hall
Yabing Li, MD
Margaret Mills
Nancy Pemberton, MS
Farida Rautaharju, PhD
Zhuming Zhang, MD

Nutrition Coding Center **(Columbia, SC)**
Elizabeth Mayer-Davis, PhD*
Robert R. Moran, PhD**

Quality of Well-Being Center **(La Jolla, CA)**
Ted Ganiats, MD*
Kristin David, MHP*
Andrew J. Sarkin, PhD*
Erik Groessl, PhD

NIH/NIDDK **(Bethesda, MD)**
R. Eastman, MD
Judith Fradkin, MD
Sanford Garfield, PhD

Centers for Disease Control & Prevention **(Atlanta, GA)**
Edward Gregg, PhD
Ping Zhang, PhD

University of Michigan **(Ann Arbor, MI)**
William H. Herman, MD, MPH

+Genetics Working Group
Jose C. Florez, MD, PhD[1, 2]
David Altshuler, MD, PhD[1, 2]
Liana K. Billings, MD[1]
Ling Chen, MS[1]
Maegan Harden, BS[2]
Robert L. Hanson, MD, MPH[3]
William C. Knowler, MD, DrPH[3]
Toni I. Pollin, PhD[4]

Alan R. Shuldiner, MD[4]
Kathleen Jablonski, PhD[5]
Paul W. Franks, PhD, MPhil, MS[6, 7, 8]
Marie-France Hivert, MD[9]

1=Massachusetts General Hospital
2=Broad Institute
3=NIDDK
4=University of Maryland
5=Coordinating Center
6=Lund University, Sweden
7=Umeå University, Sweden
8=Harvard School of Public Health
9=Université de Sherbrooke