SUPPLEMENTAL INFORMATION

This Supplemental Information file includes Acknowledgments, Supplemental Materials and Methods, Supplementary References and Supplemental Figure legends for the Figures S1-S22 below.

Acknowledgments

Supplemental Materials and Methods

Sample processing. Islets from 89 cadaver donors of European ancestry were provided by the Nordic Islet Transplantation Programme (http://www.nordicislets.org). All procedures were approved by the ethics committee at Lund University. Purity of islets was assessed by dithizone staining, while measurement of DNA content and estimate of the contribution of exocrine and endocrine tissue were assessed as previously described (6). The islets were cultured in CMRL 1066 (ICN Biomedicals) supplemented with 10 mM HEPES, 2 mM L-glutamine, 50 µg/ml gentamicin, 0.25 µg/ml Fungizone (GIBCO), 20 µg/ml ciprofloxacin (Bayer Healthcare), and 10 mM nicotinamide at 37 °C (5% CO2) for 1–9 days prior to RNA preparation. Total RNA was isolated with the AllPrep DNA/RNA Mini Kit following the manufacturer's instructions (Qiagen). RNA quality and concentration were measured using an Agilent 2100 bioanalyzer (Bio-Rad) and a Nanodrop ND-1000 (NanoDrop Technologies).

Microarray. Whole transcript microarray analysis was performed using GeneChip Human Gene 1.0 ST and processed with the standard Affymetrix protocol. The array data was then summarized and normalized with Robust Multi-array Analysis (RMA) method using the oligo package from BioConductor (52). Batch correction was done with COMBAT function from SVA package from BioConductor (53). Annotation was done using annotate package from BioConductor and hugene10sttranscriptcluster.db annotation data. Probesets were only kept if they matched uniquely to a gene in the latest hg19 human genome assembly. If more than one probeset matched a gene, one probeset at random was chosen in order to have only 1 probeset per gene. Finally, only probesets (or genes) mapped to the autosomes were kept.

RNA sequencing and analysis of gene and exon expression. Sample preparation was made using Illumina's TruSeq RNA Sample Preparation Kit according to their recommendations using 1 ug of high quality total RNA. The target insert size was 300 bp and it was sequenced using a paired end 101 bp protocol on the HiSeq2000 platform (Illumina). Quality assessment was made pre- and post-sample preparation on the 2100 Bioanalyzer (Agilent). Illumina Casava v.1.8.2 software was used for base calling. Paired-end 101 bp length output reads were aligned to the human reference genome (hg19) with TopHat v.2.0.2 (54) using Bowtie v.0.12.8 (55). The TopHat parameters explicitly used are tophat -p 30 -G genes.gtf --library-type fr-unstranded -r 100 -F 0.05 --microexon-search. The annotated RefSeq GTF transcript and fasta genome files were from UCSC and were downloaded from http://cufflinks.cbcb.umd.edu/igenomes.html. Gene expression was measured as the normalized sum of expression of all exons. Exons were defined as non-overlapping unique exonic units, as described previously (56). The dexseq_count python script (http://www-huber.embl.de/pub/DEXSeq/analysis/scripts/) was used by counting uniquely mapped reads in each exon. Gene and exon expression normalizations were then performed using the TMM

method (57), and further normalization was applied by adjusting the expression to gene or exon length, respectively. In addition, only the genes and exons that had reads mapped to them in at least 5% of the samples were kept. The Cufflinks tool v.1.3.0 (29) was used to detect novel gene loci. Novel intergenic gene loci were kept if they didn´t overlap any GENCODE v.12 gene (30), UCSC and Ensembl gene structures, had exon-exon junction reads mapped to them, had at least two exons with no Ns, and were expressed (non-null read coverage) in at least 5% of the samples. Coding potential of these novel intergenic loci was assessed with the CPAT tool (31).

Differential expression of genes and exons between normoglycemic and hyperglycemic islets. Samples were stratified based upon glucose tolerance estimated from HbA1c, i.e. donors with normal glucose tolerance (HbA1c < 6%, n=51), impaired glucose tolerance (IGT, 6% ≤ HbA1c < 6.5%, n=15), and T2D (HbA1c ≥ 6.5%, n=12) (63). A linear model adjusting for age and sex as implemented in the R Matrix eQTL package (58) was used to determine the expression of genes associated with glucose tolerance status. Genes were kept if both microarray and RNA-seq gene expression were nominally associated with HbA1c levels, with both nominal and permutation p-values < 0.05 (after performing 10,000 permutations). Known exons overlapping only one gene were classified as associated with HbA1c levels if the exon expression in RNA-seq was also confirmed at their exon-exon junction's expression level, with both nominal and permutation p-values < 0.01 (after performing 10,000 permutations), and p-value/permutation p-value ratio ≤ mean ratio + 1 s.d.. Since most of the lincRNAs were not probed on the expression array, the list of lincRNAs associated with HbA1c levels was taken only from the RNA-seq data at a threshold of FDR<5%, and expressed in at least 5% of our samples. The same threshold was applied for the novel gene loci detected. Of note, the lincRNAs we have reported are all known RefSeq genes with known gene structures and annotations.

Genotyping. Genotyping was performed on the Illumina HumanOmniExpress 12v1 C chips and genotype calling was done with the Illumina Genome studio software. All the samples passed standard genotype QC (quality control) metrics: sample call rate >98%, only European ancestry assessed by principal component analysis comparisons with HapMap populations, gender matched, no relatedness, and no genome-wide heterozygosity outliers. SNPs were removed if SNP call rate < 98% and Hardy-Weinberg equilibrium test p-values < 5.7×10^{-7}. Individual QC genotypes were imputed to 1000 Genomes data, using IMPUTE2 (61) and the March 2012 release of the 1000 Genomes Phase I panel (http://mathgen.stats.ox.ac.uk/impute/data_download_1000G_phase1_integrated.html). The program SHAPEIT (62) was used for the pre-phasing. Probabilistic genotypes were used for the subsequent analyses and after imputation, SNPs were filtered using a minor allele frequency (MAF) > 5% and an IMPUTE2 info value of >0.8.

cis-eQTL and cis-sQTL analysis. cis-eQTL and cis-sQTL analyses were carried out on samples from 89 individuals. Associations were computed between gene expression levels (eQTL), or exon expression levels (sQTL), and all SNPs within 250kb up- or downstream of each of these genes. We used a linear model adjusting for age and sex as implemented in the R Matrix eQTL package (58). Adjusting also for HbA1c did not significantly affect QTL results, so all the results are shown only with age and sex as covariates. The eQTLs and sQTLs were kept if the false discovery rate (FDR) was less than 1%, the QTL variants had rs IDs (for the sentinel variants), and if no smaller p-value was obtained after doing 10,000 permutations. A literature search (64-79) was performed to reveal whether the eQTLs observed in islets also were observed in other human tissues. Human pancreatic islets H3K4m3, FAIRE (moderate stringency FAIRE-seq site threshold from intersection of 3 islet samples) and DNase I hypersensitivity sites were annotated as such from recent studies (32-34). Evolutionarily conserved sites were defined has such if they were called conserved by both SiPhy (80) and GERP (81) programs, as annotated by Haploreg annotation tool (82). All enrichment analyses were carried out comparing the eQTL and sQTL SNPs plus SNPs in high LD ($r^2 > 0.8$) with them vs. all the SNPs tested, to avoid biasing enrichment to more densely genotype or imputed genomic regions.

Exome sequencing and Allelic expression imbalance (AEI). Exome sequencing was performed using Illumina exome sequencing protocols. To prepare the DNA for exome capture 1 ug of intact DNA was used as input for the TruSeq DNA sample preparation Kit v2 (Illumina), which was processed according to standard protocols. Briefly, DNA shearing was performed on the Covaris S2 with a target fragment size of 300 bp before end-repair, A-tailing and adaptor ligation. After DNA sample preparation, 500 ng of each sample was pooled together in libraries of a total of 5 samples before clustering with the TruSeq PE Cluster Kit v3 (Illumina). The libraries for 82 out of the 89 samples were then sequenced on the HiSeq2000 (Illumina) platform (paired end 101 bp protocol). Illumina Casava1.8.2 software was used for base calling. Paired-end reads were aligned to the human genome (hg19) with BWA v.0.6.2 (59) in paired-end mode with –q 10 as a set parameter. Duplicated aligned reads were removed by Picard v.1.58 (http://picard.sourceforge.net), reads were then realigned and quality base scores were recalibrated using GATK v.1.6.2 (60). SNP calling was also done with GATK with parameters -T UnifiedGenotyper -baq RECALCULATE only under the TruSeq Exome targeted regions, and excluding regions of known segmental duplications, structural variants and repeats. We further restricted SNP calling to biallelic SNPs, with read depth > 14X, MAPQ0 < 1, homozygosity runs < 3 bp, mapping quality > 30, and QD (QualByDepth) > 2. The RNA-seq reads from the same 82 samples were also aligned with BWA but in single-end mode with –q 15 as a set parameter and without removing potential duplicated reads. For each RNA-seq sample we called the genotypes that were detected as heterozygous SNPs in the exome sequencing. We then filtered out genomic positions where RNA-seq reads had less than 10X coverage and that both the reference and alternative alleles in the exome sequencing had less than 10X coverage. We then did a Fisher

exact test for the proportion of reference/alternative alleles in the exome sequencing vs. RNA-seq for each sample and kept only SNPs if the allelic imbalance was detected in at least 2 samples with a false discovery rate (FDR) p-value ≤ 0.01. False discovery rate (FDR) was calculated with the Benjamini & Hochberg method under the p.adjust function in R. Briefly, all the p-values retrieved from all the testable SNPs in each sample (after the filtering criteria written above) were sorted and FDR was applied to them for significance. We further filtered out SNPs overlapping known splice sites, that were not within RefSeq autosomal genes and were not present in dbSNP v.137 (with unique mapped position), as annotated by HaploReg (82). Genes with previous allelic imbalance or imprinting status were searched in literature (83, 84) (http://www.geneimprint.com/, http://www.otago.ac.nz/IGC).

RNA editing. RNA-seq reads from the 82 samples aligned with BWA were used for SNP calling with the same parameters and filters used for exome sequencing reads described above. For each exome sequenced sample we called the genotypes that were detected as SNPs in the RNA-seq data. RNA editing sites were called on autosomes in positions which were homozygous in the exome sequencing but heterozygous in the RNA-seq data in at least 2 samples. We further filtered out RNA editing variants with low quality and coverage < 15X; that were within +/- 10 bp of exon-exon junctions discovered in all 89 samples; overlapped known splice sites, more than one gene, present in dbSNP v.137, had HaplotypeScore > 13.0, ReadPosRankSum < -8.0, MQRankSum < -12.5, were within 100bp of each other; and were not in uniquely mapable 100mers regions. The RNA editing events were checked for novelty at DARNED database (85), a repository of RNA editing events in brain, blood and lymphoblastoid cell lines.

Sanger sequencing analysis. Validation of allelic imbalance and RNA editing was carried out by RT-PCR with subsequent Sanger sequencing. For reverse transcription SuperScript II RT was used with a mixture of random hexamer primers and dT$_{18}$ (Life Technologies); PCR was run using AmpliTaq Gold Master Mix (Life Technologies), and Sanger sequencing was performed by GATC Biotech. RNA editing was examined in the genes listed in Table S18; nucleotide position, primers used, and numbers of samples are indicated. Allelic imbalance was tested for the three variants listed in Table S19 in the number of heterozygous samples indicated. PCR was run using the programme: 6' 96° - [96° 15'' - 55° 30'' - 72° 45'']$_{50}$ - 4°∞. Sanger sequencing reads were analyzed with the Mutation Surveyor V3:97 software (SoftGenetics).

RNA Interference (siRNA) and insulin secretion assay. Clonal INS-1 832/13 ^{2}cells were cultured as previously described (86) and transfected using a mixture of DharmaFECT® 1 (Dharmacon; Life Technologies) and the respective siRNAs. Different sets of siRNA sequences were purchased with siRNA identification numbers: s178860 and s178858 (*TSPAN33*), s132856 and s132854 (*NT5E*), s161202 and s161203 (*PAK7*) and s146175 (*TMED6*) (Ambion). For control purposes, a previously described control sequence Silencer® Negative Control #2 from Ambion was used. Cells were cultured in medium for 72 hours at 37°C in a humidified atmosphere

containing 95% air and 5% CO_2 in the presence of 40 nM siRNA in 24-well cell culture microplates. Knockdown was assessed by RT-qPCR of the target genes as described above using the following Taqman® gene expression assays (Life Technologies): *TSPAN33* (Rn01500778_m1), *NT5E* (Rn00665212_m1), *PAK7* (Rn01746951_m1) and *TMED6* (Rn01432785_m1). After transfection insulin secretion measurements were performed. Confluent plates containing transfected INS1-832/13 cells were washed twice with 1 mL pre-warmed Secretion Assay Buffer (SAB), pH 7.2 (114 mM NaCl, 4.7 mM KCl, 1.2 mM KH_2PO_4, 1.16 mM $MgSO_4$, 20 mM HEPES, 2.5 mM $CaCl_2$, 25.5 mM $NaHCO_3$ and 0.2% Bovine Serum Albumin) containing 2.8 mM glucose. The cells were then pre-incubated for two hours in new 2 mL SAB with 2.8 mM glucose. Afterwards, separate wells were incubated for 1 hour in 1 mL SAB containing either 2.8 mM or 16.7 mM glucose. Secreted insulin was measured from supernatant using Coat-a-Count Insulin radioimmunoassay kit (Siemens) and the values were normalized using total protein content individually for each well (BCA protein assay kit, Thermo Scientific).

Flow cytometry of islets cells. Human islets were dissociated to single cell suspension using Accutase (Life Technology). Dissociated islet cells were fixed and permeabilised priori of flow cytometric analysis of intracellular insulin and glucagon using anti-insulin and anti- glucagon antibodies (R&D Systems) conjugated with R-phycoerythrin and allophycocyanin respectively by the Lightning-Link technology (Innova Bioscience, Cambridge, United Kingdom). Flow cytometry data were acquired on a CyAN ADP (Beckman Coulter) and analyzed using FlowJo software (TreeStar, Ashland, OR, USA).

Accession numbers. Clinical information on the 89 islet donors, gene and exon annotation files, raw and processed files for their islet array and RNA-seq mRNA expression are deposited at GEO under the accession number GSE50398.

Supplementary References

1. Scully T (2012) Diabetes in numbers. *Nature* **485**, S2-S3.

2. Morris AP, *et al.* (2012) Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* **44**, 981-990.

3. Scott RA, *et al.* (2012) Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat Genet* **44**, 991-1005.

4. Lyssenko V, *et al.* (2008) Clinical risk factors, DNA variants, and the development of type 2 diabetes. *N Engl J Med* **359**, 2220-2232.

5. Dimas AS, *et al.* (2014) Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes* **63**, 2158–2171.

6. Taneera J, *et al.* (2012) A systems genetics approach identifies genes and pathways for type 2 diabetes in human islets. *Cell Metab* **16**, 122-134.

7. Mahdi T, *et al.* (2012) Secreted frizzled-related protein 4 reduces insulin secretion and is overexpressed in type 2 diabetes. *Cell Metab* **16**, 625-633.

8. Dermitzakis ET (2012) Cellular genomics for complex traits. *Nat Rev Genet* **13**, 215-220.

9. Dorrell C, *et al.* (2011) Transcriptomes of the major human pancreatic cell types. *Diabetologia* **54**, 2832-2844.

10. Eizirik DL, *et al.* (2012) The human pancreatic islet transcriptome: expression of candidate genes for type 1 diabetes and the impact of pro-inflammatory cytokines. *PLoS Genet* **8**, e1002552.

11. Maffei A, *et al.* (2004) Identification of tissue-restricted transcripts in human islets. *Endocrinology* **145**, 4513–4521.

12. Gunton JE, *et al.* (2005) Loss of ARNT/HIF1ß mediates altered gene expression and pancreatic-islet dysfunction in human type 2 diabetes. *Cell* **122**, 337–349.

13. Kutlu B, *et al.* (2009) Detailed transcriptome atlas of the pancreatic beta cell. *BMC Med Genomics* **2**, 3.

14. Lyttle BM, *et al.* (2008) Transcription factor expression in the developing human fetal endocrine pancreas. *Diabetologia* **51**, 1169–1180.

15. Marselli L, *et al.* (2010) Gene expression profiles of beta-cell enriched tissue obtained by laser capture microdissection from subjects with type 2 diabetes. *PLoS One* **5**, e11499.

16. Morán I, *et al.* (2012) Human ß cell transcriptome analysis uncovers lncRNAs that are tissue-specific, dynamically regulated, and abnormally expressed in type 2 diabetes. *Cell Metab* **16**, 435-448.

17. Bramswig NC, *et al.* (2013) Epigenomic plasticity enables human pancreatic a to ß cell reprogramming. *J Clin Invest* **123**, 1275-1284.

18. Nica AC, *et al.* (2013) Cell-type, allelic and genetic signatures in the human pancreatic beta cell transcriptome. *Genome Res* **23**, 1554-1562.

19. Djebali S, et al. (2012) Landscape of transcription in human cells. *Nature* **489**, 101-108.

20. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* **18**, 1509-1517.

21. Ait-Lounis A, et al. (2010) The transcription factor Rfx3 regulates beta-cell differentiation, function, and glucokinase expression. *Diabetes* **59**, 1674-1685.

22. Aston-Mourney K, et al. (2007) Increased nicotinamide nucleotide transhydrogenase levels predispose to insulin hypersecretion in a mouse strain susceptible to diabetes. *Diabetologia* **50**, 2476–2485.

23. Manning AK, et al. (2012) A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet* **44**, 659-669.

24. Kooner JS, et al. (2011) Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nat Genet* **43**, 984-989.

25. Huyghe JR, et al. (2013) Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. *Nat Genet* **45**, 197-201.

26. Dupuis J, et al. (2010) New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* **42**, 105-116.

27. Cho YS, et al. (2011) Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. *Nat Genet* **44**, 67-72.

28. Soranzo N, et al. (2010) Common variants at 10 genomic loci influence hemoglobin A1(C) levels via glycemic and nonglycemic pathways. *Diabetes* **59**, 3229-3239.

29. Roberts A, Pimentel H, Trapnell C, Pachter L (2011) Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* **27**, 2325-2329.

30. Harrow J, et al. (2012) GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* **22**, 1760-1774.

31. Wang L, et al. (2013) CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res* **41**(6), e74.

32. ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74.

33. Stitzel ML, et al. (2010) Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci. *Cell Metab* **12**, 443-455.

34. Gaulton KJ, et al. (2010) A map of open chromatin in human pancreatic islets. *Nat Genet* **42**, 255-259.

35. 1000 Genomes Project Consortium (2010) A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061-1073.

36. Sekine N, et al. (1994) Low lactate dehydrogenase and high mitochondrial glycerol phosphate dehydrogenase in pancreatic beta-cells. Potential role in nutrient sensing. *J Biol Chem* **269**, 4895-4902.

37. Fierabracci A, Milillo A, Locatelli F, Fruci D (2012) The putative role of endoplasmic reticulum aminopeptidases in autoimmunity: insights from genomic-wide association studies. *Autoimmun Rev* **12**, 281-288.

38. Harashima S, *et al.* (2012) Sorting nexin 19 regulates the number of dense core vesicles in pancreatic ß-cells. *J Diab Invest* **3**, 52-61.

39. Andersson SA, *et al.* (2012) Reduced insulin secretion correlates with decreased expression of exocytotic genes in pancreatic islets from patients with type 2 diabetes. *Mol Cell Endocrinol* **364**, 36-45.

40. Gosmain Y, *et al.* (2012) Pax6 is crucial for ß-cell function, insulin biosynthesis, and glucose-induced insulin secretion. *Mol Endocrinol* **26**, 696-709.

41. Cunnington MS, Santibanez KM, Mayosi BM, Burn J, Keavney B (2010) Chromosome 9p21 SNPs Associated with Multiple Disease Phenotypes Correlate with ANRIL Expression. *PLoS Genet* **6**, e1000899.

42. Ge ZJ, *et al.* (2013) Maternal diabetes causes alterations of DNA methylation statuses of some imprinted genes in murine oocytes. *Biol Reprod* **88**, 117.

43. Pasquali L, *et al.* (2014) Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat Genet* **46**, 136-143.

44. Nogueira TC, *et al.* (2013) GLIS3, a Susceptibility Gene for Type 1 and Type 2 Diabetes, Modulates Pancreatic Beta Cell Apoptosis via Regulation of a Splice Variant of the BH3-Only Protein Bim. *PLoS Genet* **9**, e1003532.

45. Chen L (2013) Characterization and comparison of human nuclear and cytosolic editomes. *Proc Natl Acad Sci USA* **110**, 2741-2747.

46. Pickrell JK, Gilad Y, Pritchard JK (2012) Comment on "Widespread RNA and DNA sequence differences in the human transcriptome". *Science* **335**, 1302.

47. Flannick J, *et al.* (2014) Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. *Nat Genet* **46**, 357-63.

48. Hardy AB, *et al.* (2012) Effects of high-fat diet feeding on Znt8-null mice: differences between ß-cell and global knockout of Znt8. *Am J Physiol Endocrinol Metab* **302**, E1084-E1096.

49. Dayeh T, *et al.* (2014) Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion. *PLoS Genet* **10**, e1004160.

50. Jiang X, *et al.* (2010) The imprinted gene PEG3 inhibits Wnt signaling and regulates glioma growth. *J Biol Chem* **285**, 8472-8480.

51. Zhou Y, *et al.* (2012) Survival of pancreatic beta cells is partly controlled by a TCF7L2-p53-p53INP1-dependent pathway. *Hum Mol Genet* **21**, 196-207.

52. Carvalho BS, Irizarry RA (2010) A framework for oligonucleotide microarray preprocessing. *Bioinformatics* **26**, 2363-2367.

53. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882-883.

54. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111.

55. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25.

56. Anders S, Reyes A, Huber W (2012) Detecting differential usage of exons from RNA-seq data. *Genome Res* **22**, 2008-2017.

57. Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11**, R25.

58. Shabalin AA (2012) Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353-1358.

59. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760.

60. McKenna A, *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303.

61. Howie BN, Donnelly P, Marchini J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**, e1000529.

62. Delaneau O, Marchini J, Zagury JF (2012) A linear complexity phasing method for thousands of genomes. *Nat Methods* **9**, 179–181.

63. International Expert Committee (2009) International Expert Committee report on the role of the A1C assay in the diagnosis of diabetes. *Diabetes Care* **32**, 1327–1334.

64. Liang L *et al.* (2013) A cross-platform analysis of 14,177 expression quantitative trait loci derived from lymphoblastoid cell lines. *Genome Res* **23**, 716-726.

65. Grundberg E, *et al.* (2012) Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet* **44**, 1084-1089.

66. Montgomery SB, *et al.* (2010) Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* **464**, 773-777.

67. Lonsdale J, *et al.* (2013) The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-585.

68. Degner JF, *et al.* (2012) DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature* **482**, 390-394.

69. Schadt EE, *et al.* (2008) Mapping the genetic architecture of gene expression in human liver. *PLoS Biol* **6**, e107.

70. Myers AJ, *et al.* (2007) A survey of genetic human cortical gene expression. *Nat Genet* **39**, 1494-1499.

71. Stranger BE, *et al.* (2007) Population genomics of human gene expression. *Nat Genet* **39**, 1217-1224.

72. Veyrieras JB, *et al.* (2008) High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet* **4**, e1000214.

73. Pickrell JK, *et al.* (2010) Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* **464**, 768-772.

74. Gaffney DJ, *et al.* (2012) Dissecting the regulatory architecture of gene expression QTLs. *Genome Biol* **13**, R7.

75. Innocenti F, *et al.* (2011) Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. *PLoS Genet* **7**, e1002078.

76. Zeller T, *et al.* (2010) Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS One* **5**, e10693.

77. Dimas AS, *et al.* (2009) Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* **325**, 1246-1250.

78. Boyle AP, *et al.* (2012) Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* **22**, 1790-1797.

79. Ramos EM, *et al.* (2013) Phenotype-Genotype Integrator (PheGenI): synthesizing genome-wide association study (GWAS) data with existing genomic resources. *Eur J Hum Genet* **22**, 144-147.

80. Garber M, *et al.* (2009) Identifying novel constrained elements by exploiting biased substitution patterns. *Bioinformatics* **25**, 54-62.

81. Davydov EV, *et al.* (2010) Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol* **6**, e1001025.

82. Ward LD, Kellis M (2012) HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* **40**, D930-D934.

83. Morison IM, Paton CJ, Cleverley SD (2001) The imprinted gene and parent-of-origin effect database. *Nucleic Acids Res* **29**, 275-276.

84. Serre D, *et al.* (2008) Differential allelic expression in the human genome: a robust approach to identify genetic and epigenetic cis-acting mechanisms regulating gene expression. *PLoS Genet* **4**, e1000006.

85. Kiran A, Baranov PV (2010) DARNED: a DAtabase of RNa EDiting in humans. *Bioinformatics* **26**, 1772-1776.

86. Malmgren S, *et al.* (2009) Tight coupling between glucose and mitochondrial metabolism in clonal beta-cells is required for robust insulin secretion. *J Biol Chem* **284**, 32395-32404.

Supplemental Figure legends

Figure S1. Overview of study design and main results. Different *omics* platforms were used to assess a comprehensive spectrum of gene regulation in human pancreatic islets. RNA sequencing (RNA-seq) was used to detect known and novel genes expressed in at least 5% of the 89 samples, and known exons and novel genes associated with glucose tolerance status. Known genes were reported to associate with glucose tolerance status if both Expression microarrays and RNA-seq detect them at nominal and permutation p-value<0.05. SNPs genotyped in our islet samples and further imputed to the 1000 Genomes reference panel were used in combination with RNA-seq to detect expression quantitative trait loci (eQTL) for known and novel genes, and splicing QTL (sQTL) for known exons at a 1% false discovery rate (FDR) and 10,000 permutations. 35 genes had both eQTLs and associated linearly with glucose tolerance status. Allelic imbalance was detected by using Fisher exact test to compute significant deviations from the expected 50/50 allelic distribution when comparing Exome sequencing (Exome-seq) and RNA-seq for the same individuals (at 1% FDR and detected in at least 2 samples). RNA editing was also detected by comparing exome and RNA sequencing data after a stringent pipeline (Materials and Methods and Supplemental Information).

Figure S2. Distribution of RNA-seq expression of known genes on 89 human pancreatic islet samples. Density of reads mapped to RefSeq genes with red vertical bars separating the 4 quartiles of expression.

Figure S3. Correlation between RNA-seq and microarray data on 89 human pancreatic islet samples. Spearman correlation between the normalized expression of genes detected in both platforms (Materials and Methods and Supplemental Information).

Figure S4. Fraction of RefSeq genes, transcripts, exons and junctions detected by RNA-seq as a function of cumulative reads mapped to these features. Black vertical line marks the average number of reads per sample (38.2 million paired-end reads) mapped to the human genome.

Fig. S5. Co-expression analysis of *RASGRP1* (n=89). (A) *RASGRP1* vs. *GCG*. (B) *RASGRP1* vs. *INS*. (C) *RASGRP1* vs. *SST*. (D) *RASGRP1* vs. Glucose tolerance status.

Fig. S6. Co-expression analysis of *RFX3* (n=89). (A) *RFX3* vs. *GCG*. (B) *RFX3* vs. *INS*. (C) *RFX3* vs. *SST*. (D) *RFX3* vs. Glucose tolerance status.

Fig. S7. Co-expression analysis of *NNT* (n=89). (A) *NNT* vs. *GCG*. (B) *NNT* vs. *INS*. (C) *NNT* vs. *SST*. (D) *NNT* vs. Glucose tolerance status.

Fig. S8. Co-expression analysis with glucagon gene (n=89). (A) *SLC30A8* vs. *GCG*. (B) *PCSK1* vs. *GCG*. (C) *G6PC2* vs. *GCG*.

Figure S9. Novel gene locus (chr12:43,504,654-43,507,028) (A) with evidence of sequence conservation and transcription, (B) associated with HbA1c levels in human pancreatic islets (Normal n=51; IGT n=15; T2D n=12), (C) and under a region nominally significant associated with fasting glucose in a previous study (23). Arrow provides the location of this new transcribed locus.

Figure S10. RNA-seq eQTL fraction detected to be nominally significant in the microarray data, stratified by gene expression quartiles (1 being the lowest and 4 the highest quartile).

Figure S11. Power to detect eQTLs as a function of sample size. This plot is calculated with the java applet at http://homepage.stat.uiowa.edu/~rlenth/Power/

Figure S12. High linkage disequilibrium (LD) region around *ERAP2* eQTL sentinel SNP (rs2910686) shows nominal significance with fasting glucose in MAGIC database (23).

Figure S13. Example of an sQTL not detected at gene level. (A) The sQTL is not detected at the gene expression level (p-value >0.05), (B) and only usage of exon 12 of *BRD2* gene is associated with SNP rs114933220 (p-value = $9.5e^{-06}$). This gene has been linked to obesity and protection from type 2 diabetes.

Figure S14. Known type 2 diabetes (T2D) GWAS locus rs1535500 as eQTL for *KCNK17*. (A) The SNP rs1535500 is located in *KCNK16*, but (B) shows an eQTL effect on neighboring *KCNK17* gene (p-value = $1.2e^{-06}$) (C) and not on *KCNK16* (p-value > 0.05).

Figure S15. Example of allelic imbalance in the *MMP7* gene validated by Sanger sequencing. (A) IGV browser with RNA-seq reads from samples in which the allelic imbalance locus rs10502001 was validated by Sanger sequencing. (B) The genotype for rs10502001 is associated with depolarization-evoked insulin exocytosis.

Figure S16. Overview of the pipeline for detecting RNA editing events in 82 human pancreatic islet samples. From the initial 89 samples we only had enough DNA and RNA for doing both RNA and Exome sequencing in 82 samples. Each of these 82 samples was then processed through this pipeline. To minimize the false positive rate we report only the RNA editing events detected in at least two individuals (Materials and Methods and Supplemental Information).

Figure S17. Distribution of RNA editing events (Materials and Methods and Supplemental Information).

Figure S18. RNA editing events in human pancreatic islets validated by Sanger sequencing. From 9 randomly chosen editing events, we could validate 3 out of 6 of the A-to-G events by Sanger sequencing, but none of the 3 non A-to-G events.

Fig. S19. *RFX3* co-expression with glucokinase (*GCK*) (n=89).

Fig. S20. Islet purity for our 89 human pancreatic islet samples (assessed by dithizone staining) in relation to disease status. Kruskal-Wallis rank sum test was used to assess the association of gene expression with glucose tolerance status of the islet donors (Normal n=51; IGT n=15; T2D n=12).

Fig. S21. Expression of cell-type specific genes in relation to disease status. (A) *GCG* is an alpha-cell specific gene. (B) *MAFA* is a beta-cell specific gene. (C) *SST* is a delta-cell specific gene. (D) *AMY2A* is an exocrine specific gene. Kruskal-Wallis rank sum test was used to assess the association of gene expression with glucose tolerance status of the islet donors (Normal n=51; IGT n=15; T2D n=12).

Fig. S22. FACS beta/alpha cells ratio in relation to disease status of 49 islet donor samples (partially overlapped by our 89 islet donor samples used in our study). (A) Kruskal-Wallis rank sum test was used to assess the association of FACS beta/alpha cells ratio with glucose tolerance status (p-value = 0.1373) (Normal n=26; IGT n=14; T2D n=9). (B) The blue horizontal line separates the few T2D donors with high HbA1c (n=3, HbA1c ≥7.3%) that have an insulin/glucagon ratio less than any other Normal or IGT sample.

Fig. S1

89 human pancreatic islet donors

Expression microarray

RNA-seq gene expression & exon usage
MAF ≥ 5% samples (18 567 known & 445 novel genes)

Genotypic imputation
r^2 ≥ 0.8 & MAF ≥ 5% (6.2 M variants)

Exome-Seq

• **1619 known genes vs. HbA1c** (array & RNA-seq)
• **271 known genes with exons vs. HbA1c** (exon & junction data confirmation)
• **1 novel gene locus vs. HbA1c** (FDR<1% & 10k perm)

• **616 cis-eQTLs (known genes)**
• **371 cis-sQTLs (known exons)**
• **24 cis-eQTLs (novel gene loci)** (FDR<1% & 10k permutations)

35 known genes have eQTLs and associate with HbA1c

• **1102 allelic imbalance genes** (FDR<1% & ≥ 2 samples)
• **61 genes with RNA editing** (≥ 2 samples)

Fig. S2

Fig. S3

Mean expression values from RNA−seq and microarray (rho = 0.83)
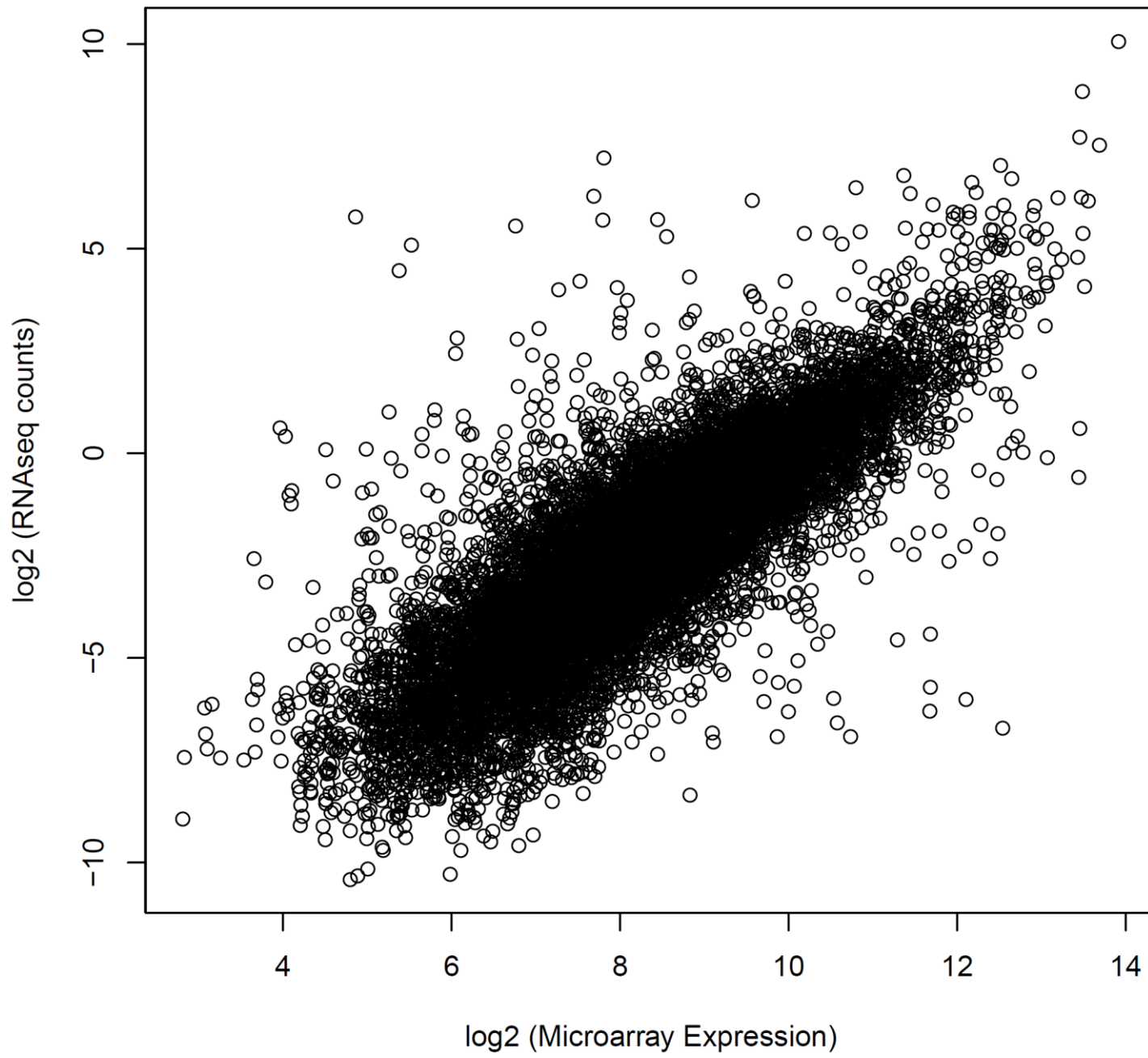
Detected features (compared with high-coverage sample)

genes
junctions
transcripts
exons

Fraction of detected features

Million reads

Fig. S5

A

log2 (GCG expression) vs log2 (RASGRP1 expression)

Spearman´s cor = 0.44
p−value = 1.827e−05

B

log2 (INS expression) vs log2 (RASGRP1 expression)

Spearman´s cor = −0.43
p−value = 3.25e−05

C

log2 (SST expression) vs log2 (RASGRP1 expression)

Spearman´s cor = 0.02
p−value = 0.826

D

log2 (RASGRP1 expression) vs Glucose tolerance status

p−value = 6.87e−06

Normal   IGT   T2D

Fig. S6

A

log2 (GCG expression) vs log2 (RFX3 expression)

Spearman´s cor = 0.50
p−value = 4.704e−07

B

log2 (INS expression) vs log2 (RFX3 expression)

Spearman´s cor = −0.51
p−value = 3.421e−07

C

log2 (SST expression) vs log2 (RFX3 expression)

Spearman´s cor = −0.18
p−value = 0.09

D

log2 (RFX3 expression) vs Glucose tolerance status

p−value = 0.03

Normal   IGT   T2D

Fig. S7

A

log2 (GCG expression) vs log2 (NNT expression)

Spearman´s cor = 0.50
p−value = 7.238e−07

B

log2 (INS expression) vs log2 (NNT expression)

Spearman´s cor = −0.63
p−value < 2.2e−16

C

log2 (SST expression) vs log2 (NNT expression)

Spearman´s cor = −0.27
p−value = 0.01

D

log2 (NNT expression) vs Glucose tolerance status

p−value = 0.001

Normal   IGT   T2D

Fig. S8

Fig. S9

Fig. S10

Fig. S11

Power to detect eQTLs

Beta (effect size)

Legend: 89, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000

Fig. S12

MAGIC_Manning_et_al_FastingGlucose_MainEffect_2012

Fig. S13

Fig. S14

Fig. S15

Fig. S16

Exome-seq reads — BWA → Human genome sequence (b37) ← BWA — RNA-seq reads

↓

SNP calling with GATK Unified Genotyper

↓

Discard dbSNPs and SNPs whose DNA and RNA share the same genotypes

Discard loci of known and novel splice sites, repeats, SVs, low qual, depth <15X, non homozygous DNA genotypes &overlapping more than one gene

↓

BLAT filter – Discard loci with multiple genome hits at high sequence id

↓

Discard SNPs with extreme editing degree

↓

65 potential RNA edited loci in at least two individuals

Fig. S17

Validation of RNA editing by Sanger sequencing

PPIA A → G chr7 44 842 099
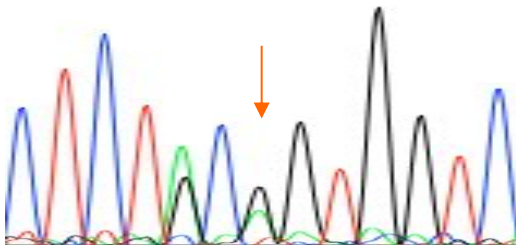
DDX58 A → G chr9 32 456 316
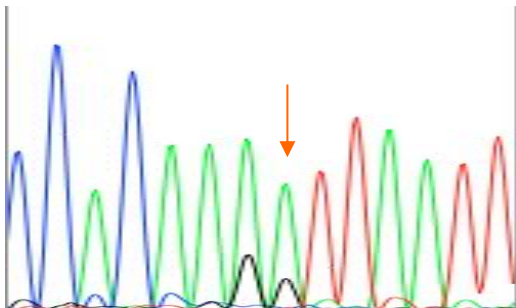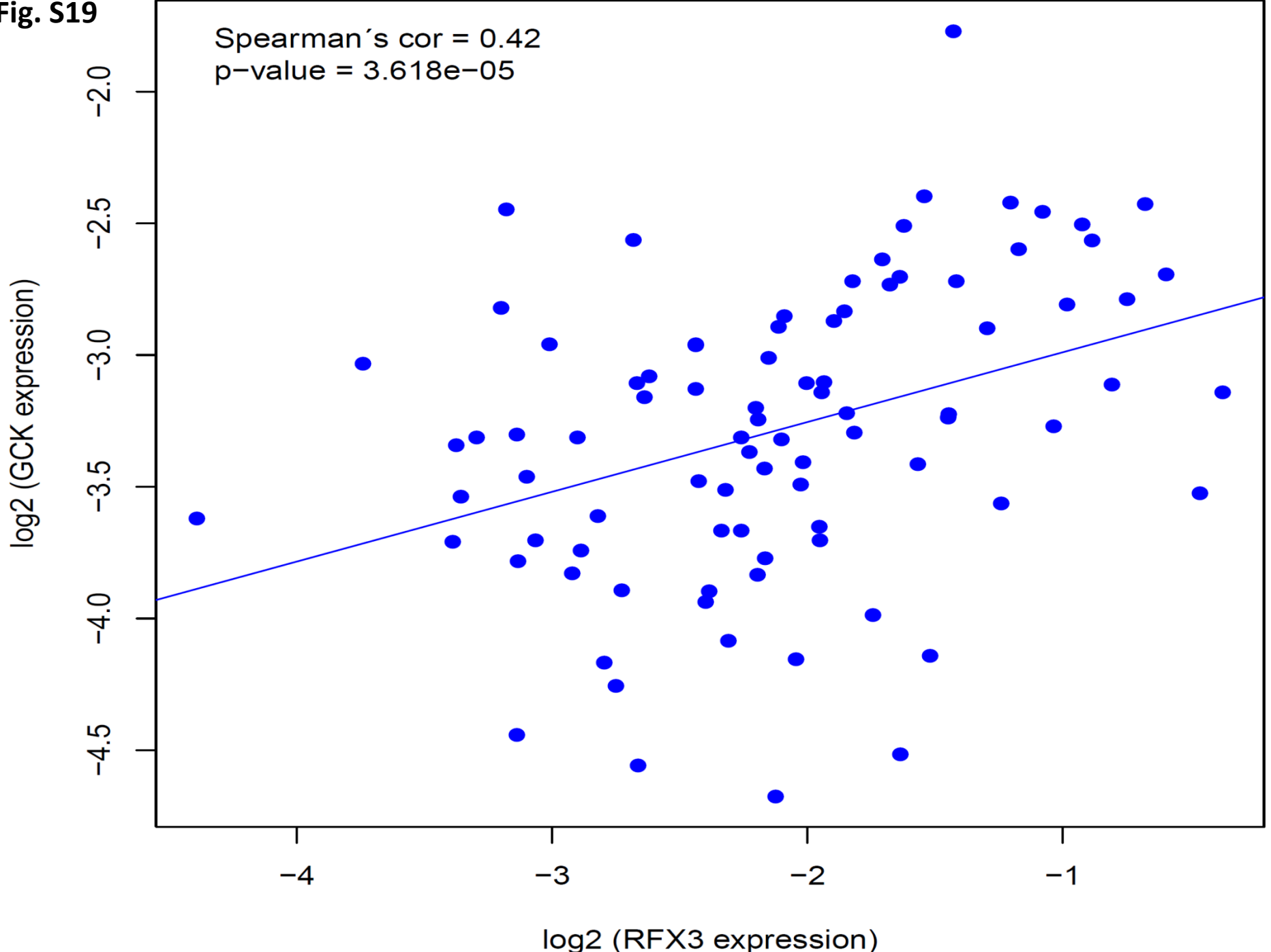
MAP6 A → G chr11 75 316 759

Fig. S19

Fig. S20

Fig. S21

A

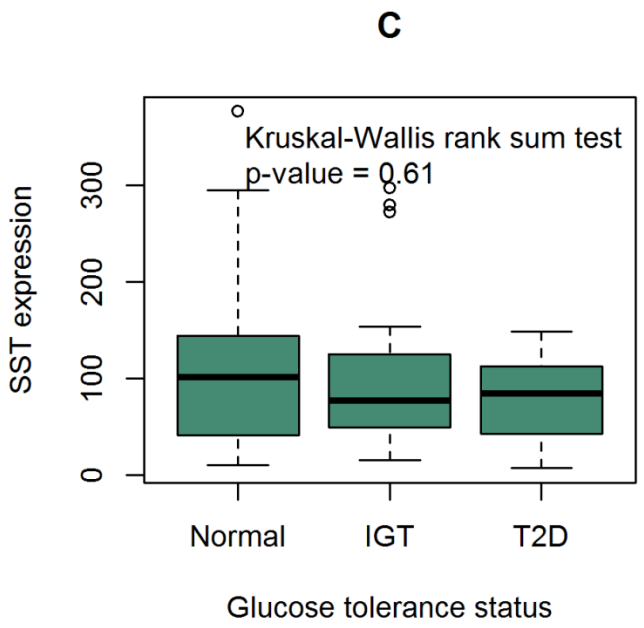GCG expression — Glucose tolerance status (Normal, IGT, T2D)

Kruskal-Wallis rank sum test
p-value = 0.07

B

MAFA expression — Glucose tolerance status (Normal, IGT, T2D)

Kruskal-Wallis rank sum test
p-value = 0.47

C

SST expression — Glucose tolerance status (Normal, IGT, T2D)

Kruskal-Wallis rank sum test
p-value = 0.61

D

AMY2A expression — Glucose tolerance status (Normal, IGT, T2D)

Kruskal-Wallis rank sum test
p-value = 0.30

Fig. S22

A

B