Islet Gene View - a tool to facilitate islet research

Olof Asplund[1], Petter Storm[1,2], Emilia Ottosson-Laakso[3], Ulrika Krus[4], Olle Korsgren[5,6], Hindrik Mulder[7], Erik Renström[4], Ola Hansson[1,9], LUDC Consortium[8], Leif Groop[1,9a,b*], Rashmi B Prasad[1a,b*]

**a These authors contributed equally*
**b Senior authors*

[1] Genomics, Diabetes and Endocrinology, Department of Clinical Sciences, Lund University Diabetes Centre, 20502 Malmö, Sweden
[2] Center of Molecular Diagnostics, Division of Laboratory Medicine, Region Skåne, Lund, Sweden
[3] Clinical Genomics Facility, Department of Microbiology, Tumor and Cell Biology (MTC), C1, 17165 Solna, Sweden
[4] Islet Pathophysiology, Department of Clinical Sciences, Lund University Diabetes Centre, Lund University, Malmö, Sweden
[5] Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden
[6] Department of Clinical Chemistry and Transfusion Medicine, Institute of Biomedicine, University of Gothenburg, Gothenburg, Sweden
[7] Molecular Metabolism, Department of Clinical Sciences, Lund University Diabetes Centre, 20502 Malmö, Sweden
[8] Lund University Diabetes Centre (LUDC); https://www.ludc.lu.se
[9] Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki 00014, Finland

Corresponding authors:

Rashmi B Prasad
Department of Clinical Sciences, Clinical Research Centre, Lund University Diabetes Centre, Malmö, Sweden.
email: rashmi.prasad@med.lu.se

Leif Groop
Department of Clinical Sciences, Clinical Research Centre, Lund University Diabetes Centre, Malmö, Sweden.
Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki 00014, Finland
email: leif.groop@med.lu.se

Word count:
Abstract: 158
Main text: 4774
Number of figures: 3
Number of tables: 0
Number of supplementary figures: 9
Number of supplementary tables:4

Abstract

Changes in the hormone producing pancreatic islet cells are culprits in the development of diabetes, and these changes are often reflected by changes in expression of key genes. Analysis of such changes require advanced bioinformatics tools that can translate the process into simple visible information. To achieve these goals we have developed a tool called Islet Gene View (IGW), that aims to make information on gene expression in human pancreatic islets from organ donors easily accessible to the scientific community. Islet Gene View currently consists of information on islets from 188 donors where data on RNA expression of more than 15 000 genes can be related to phenotypic information (gender, age, BMI, HbA1c, insulin and glucagon secretion etc.). Also this information can be related to expression other genes and how gene expression is regulated by genetic variants(so called eQTLs, expression Quantitative Traits) The data will be accessible thorugh the easy-to-use Islet Gene View web application.

Significance Statement

We present Islet Gene View (IGW), a web resource that aims to make information on gene expression in human pancreatic islets from organ donors easily accessible to the scientific community. In IGW, we leveraged global RNA expression data from 191 donor-islets to understand their relationship with islet phenotypes such as gender, age, BMI, HbA1c, insulin and glucagon secretion as well as with islet hormones coding genes. GWAS based on European studies have shown that 403 genetic variants associate with T2D risk however the target genes and function in islets are largely unknown. We linked T2D risk variants to gene expression in islets and further functionally report if their expression is modulated in relation to T2D status and other related features.

Introduction

Gene expression analysis provides a link between genetics and cellular function and is crucial for the elucidation of pathophysiological mechanisms. Information on gene expression in different tissues has been instrumental to the scientific community. The Genotype-Tissue Expression (GTEx) project (2013) has provided a pioneering example on how to share information from deceased humans. Unfortunately, GTEx has limited information on human pancreatic islets of Langerhans, as sequencing has only been performed on whole pancreas from a limited number of deceased donors. Importantly, the pancreas needs to be removed while blood flow is still intact to retain functionality of the pancreatic cells, and therefore more information on human pancreatic islets can be derived from organ donors for transplantation purposes where blood flow still has been kept intact.

The Human Tissue Laboratory (HTL), which was created as part of a collaboration between Uppsala and Lund universities and funded by a strategic research grant, *i.e.* the Excellence of Diabetes Research in Sweden (EXODIAB), has generated a large repository of tissue samples (human pancreatic islets, fat, liver and muscle) from deceased organ donors. Here, we leveraged gene expression (RNA sequencing) and genome wide genetic variation data (GWAS) from human pancreatic islets to elucidate gene expression modulation in relation to T2D status, as well as analysis of effects of genetic variation on gene expression, *i.e.* eQTLs (Expression Quantitative Traits). Additionally, gene expression can be explored in relation to gender, BMI, and age as well as glucose-stimulated insulin secretion and expression of other pancreatic hormones. To provide rapid and robust overview of the data, as well as “look-ups”, for scientists in EXODIAB, a web-based tool, Islet Gene View, was created and will be made available to the scientific community.

Materials and methods

Sample acquisition

Human pancreatic islets (n = 188), fat (n = 12), liver (n = 12) and muscle (n = 12) were obtained through the EXODIAB network from the Nordic Transplantation Program

(http://www.nordicislets.org). The isolation of total RNA including miRNA was carried out using the miRNeasy (Qiagen) or the AllPrep DNA/RNA (Qiagen) mini kits. The quality of isolated RNA was controlled using a 2100 Bioanalyzer (Agilent Technologies) or a 2200 Tapestation (Agilent Technologies) and quantity was measured using NanoDrop 1000 (NanoDrop Technologies) or a Qubit 2.0 Fluorometer (Life Technologies). Clinical characteristics of the donors are shown in Supplementary Table 1.

Islet Phenotypes

Purity of islets was assessed by dithizione staining, and estimates of the contribution of exocrine and endocrine tissue was assessed as previously described (Friberg et al., 2011).

Phenotypic information:: Diagnosis of Type 2 diabetes (T2D) was either based on an existing clinical diagnosis of T2D (N = 33) or on an HbA1c above 6.5% (NGSP units; equal to 48 mmol/mol in IFCC) (N = 25). IGT was defined as HbA1c between 6 and 6.5% (N = 30).

BMI and gender information was obtained from donor records. Stimulatory index (SI) was used as a measure of glucose-stimulated insulin secretion. For this purpose islets were subjected to static perfusion of glucose, which was raised from 1.67 to 16.7 mmol/l; insulin was measured at both high and low glucose. The fold change in insulin levels between the two conditions was used as a measurement of glucose-stimulated insulin secretion.

Sample preparation for sequencing

One µg of total RNA of high quality (RIN>8) was used for sequencing with a TruSeq RNA sample preparation kit (Illumina). The size selection was made by Agencourt AMPure XP beads (Beckman Coulter) aiming at a fragment size above 300 bp. The resulting libraries were quality controlled on a 2100 Bioanalyzer and a 2200 Tapestation (Agilent Technologies) before combining 6 samples into one flow cell to be sequenced using a HiSeq 2000 sequencer (Illumina).

Data analysis

The raw data were base-called and de-multiplexed using CASAVA 1.8.2 (Illumina) before alignment to hg38 with STAR version 2.4.1. To count the number of reads aligned to specific transcripts, featureCounts (v 1.4.4)(Liao et al., 2014) was used, with GENCODE version 22 as gene, transcript and exon models.

Raw data were normalized using trimmed mean of M-values (TMM) and transformed into log_2 counts per million (log_2 CPM), using voom (Law et al., 2014) before linear modeling. Samples with less than 10 million reads in total were excluded from further analysis. In addition, genes with lower than 1 CPM in at least 10 samples were also excluded, leaving 15,017 genes for analysis in the 188 samples. After TMM-normalization, the data were adjusted for batch effects using ComBat (Combatting batch effects when combining batches of microarray data) from the sva (surrogate variable analysis) R package (Leek et al., 2012). The effective length of each gene in each sample was calculated using RSEM (Li and Dewey, 2011). The resulting CPM values were converted into FPKM values by dividing the transcripts by their effective lengths and multiplying with 1000.

A potential association between gene expression and phenotypes was analyzed by linear modeling. P-values were calculated using the eBayes function in limma (Ritchie et al., 2015), and P-values adjusted for multiple testing were calculated across all genes using Benjamini-Hochberg correction (Benjamini and Hochberg, 1995).

As expression of 53% of the genes in the dataset was correlated with purity (mostly due to admixture of exocrine tissue), we included purity as a covariate in the linear models for all association analyses, together with sex and age. Six individuals were found to not have a type 2 diabetes diagnosis but have HbA1c of above 6.5%. The six individuals were excluded from the differential expression analysis of type 2 diabetes versus controls to avoid potentially contaminating the control group with potential undiagnosed diabetes cases.

An empirical and conservative approach was used to calculate P-values for gene-gene correlations. First, the Spearman correlations between all gene-gene pairs were calculated. Secondly, the function *fitdistr* (fitting of distributions) in the MASS (Modern Applied Statistics with S) package (Venables and Ripley, 2002) was used to estimate the mean and standard deviation of this distribution, resulting in an average correlation of ~0.0142 with a standard error of ~0.2631. This distribution model represents all observed gene-gene correlations in the data set, using a normal distribution. For both parameters, the standard error of the fit was small (order of magnitude = 10^{-5}). Based on these calculations, the observed data were used for the calculation of P-values. For all target secretory genes, P-values of correlations were calculated using the *pnorm* (probability of normal distribution) function in R (R Core Team, 2016) using the estimated parameters.

Depending on negative or positive correlations, the lower or the upper tail of the distribution was used for testing.

As a way to investigate the biological relevance of the RNA-Seq and eQTL data, we investigated the relationship between previously known risk variants and gene expression in Geneview. For a list of 404 variants previously associated with either T2D risk (Mahajan et al., 2018) or with β cell function, we identified all variants which were nominally significant eQTLs in the islets. For each eQTL, the targeted gene was extracted. The list of genes affected by eQTLs was tested for enrichment of genes whose expression was correlated with secretory genes using bootstrapping. This was done by repeatedly selecting random sets of genes of the same size(10 000 iterations). The mean empirical P-value for each run was then calculated which was used as a null distribution to calculate the probability of the observed mean P-value in the observed set of genes.

Results

Differentially expressed genes between T2D and non-diabetic donors

In order to study global gene expression in human pancreatic islets, we performed RNA sequencing on 188 islets from organ donors using Illumina's HiSEQ platform. The average number of counts mapped to genes was 26.1 million reads (± 13.1 million) ranging from 10.0 million up to 79.4 million. Analysis of differentially expressed genes (DEG) between 33 diabetic and 155 non-diabetic subjects revealed that 2295 out of 1406 genes(15%) were differentially expressed between the two groups at $p < 0.05$; of them 206 genes (1.4%) reached genome-wide significance (FDR <= 0.05, Supplementary Table 2). The five most significant DEGs upregulated in T2Ds were Serpin Family E Member 2 (*SERPINE2*, Figure 1), Iodothyronine Deiodinase 2 (*DIO2*), Synapse Differentiation Inducing 1 (*SYNDIG1*), Potassium Voltage-Gated Channel Subfamily A Regulatory Beta Subunit 2 (*KCNAB2*) and RASD family member 2 (*RASD2*). Conversely the five most significant downregulated genes were *PLA1A* (Phospholipase A1 Member A), Glycine Receptor Alpha 1 (*GLRA1*), Hedgehog Acyltransferase Like (*HHATL*), Arginase 2 (*ARG2*) and RAS Guanyl Releasing Protein 1 (*RASGRP1*).

We then stratified our analysis based on Hba1c levels whereby Hba1c > 6.5 were considered T2D, Hba1c between 6% and 6.5% were considered to have IGT and HbA1c < 6 normal glucose tolerance (NGT). Out of a total of 15017 genes 12% were nominally (P<=0.05) significantly DEGs between IGT and NGT; however, these differences did not reach genome-wide significance (FDR <= 0.05). Similarly 11.5% of genes were nominally differentially expressed (P <= 0.05) and notably, only 0.4% (55 genes, Supplementary Table 3) were genome-wide DEGS between T2D and NGT

(normoglycemic) donors when diagnosed based on Hba1c strata alone. Gene expression association analysis using HbA1c as a continuous variable revealed expression of 5.1% of genes being nominally associated with HbA1c ; only 0.1% (21 genes, Supplementary Table 4) were genome-wide significant.

Of 1619 genes which were found to be nominally significantly associated with HbA1c levels in a previous study based on a subset of the same dataset (ref Fadista et. al), 36.5% were replicated in this larger dataset as being nominally differentially expressed in either the T2D stratum or the IGT stratum.

Co-expression of genes with expression of islet secretory genes

Pancreatic islets comprise distinct cell types defined by the proteins they secrete; alpha cells secrete glucagon, beta cells secrete insulin and islet amyloid polypeptide (*IAPP)*, delta cells secrete somatostatin, and F cells (also called PP cells) secrete pancreatic polypeptide Y. We therefore also performed a co-expression analysis of all genes with these 5 secretory genes including insulin (*INS*, Figure 1I), glucagon (*GCG*, Figure 1J), somatostatin (*SST*, Figure 1K), islet amyloid polypeptide (*IAPP*, Figure 1L), and pancreatic polypeptide Y (*PPY*, Figure 1M).

INS expression was correlated with the largest number of genes expressed in human pancreatic islets (Supplementary Table 5), followed by *GCG*, *SST*, *IAPP* and *PPY* respectively. Notably, 581 genes showed positive co-expression with *INS*, 343 with *GCG*, 98 with *SST* and 50 with *IAPP* (Figure 2A) whereas 632 genes were negatively co-expressed with *INS*, 298 with *GCG*, 92 with *SST* and 12 with *IAPP* (Figure 2B). 112 genes were are either positively or negatively co-expressed with two or more secretory genes (Figure 2C-D). Six genes showed opposite correlation with *INS* and *GCG, e.g. WIZ*, *NUDC*, *IMP4*, *ZNF787*, *KLHL21*, and *ISG20*. Top 10 correlated genes are shown in Supplementary table 6.

INS was most strongly positively correlated with the Interferon Regulatory Factor 2 Binding Protein 1 (IRF2BP1) gene (r = 0.85, p = 0.002) but negatively with the GC-Rich Promoter Binding Protein 1 Like 1 (*GPBP1L1)* gene (r = -0.72, P = 0.0053). Glucagon (*GCG*) showed strong positive correlation with the Tumor Suppressor Candidate 3 (*TUSC3*) gene and negative with Ankyrin Repeat And BTB Domain Containing 2 (*ABTB2*). The expression of the Retinol Binding Protein 4 (*RBP4*) gene showed the strongest positive correlation with expression of the *SST* gene. Further, *RBP4* is negatively associated with HbA1c (P = 0.0085) and BMI (P = 0.023) but positively correlated with *INS* expression (P = 0.048). The *WWC* gene showed the strongest negative correlation with SST. *WWC2* was also negatively associated with purity ($P = 1.8 \cdot 10^{-15}$) and correlated negatively with expression of *INS* (P = 0.021), suggesting that it might also have an exocrine origin.

Expression quantitative traits (eQTLs)

In order to explore the effect of genetic variation on gene expression in human pancreatic islets, we performed eQTL analysis using GWAS data imputed with the 1000 Genomes Phase I (March 2012) integrated variant set reference panel and RNA data from the 188 donors. The effect of SNPs proximal to a given gene on its expression was computed by cis-eQTL analysis using Matrix eQTL. For this analysis, only cis-eQTLs located within 2 Mb of the target genes were considered. 13 296(95%) of all genes had at least one nominally significant eQTL(P < 0.05), and 25% had at least one genome-wide significant eQTL (FDR <= 0.05) adjusted for all tested variants. The five strongest eQTLs were seen in the *ERAP2*, *ERV3-1*, *XRRA1*, *XKR9*, and *ASAH2* genes (Supplementary Figure 1). Among the 80 top DEGs across all phenotypes, 58 had at least one genome-wide significant eQTL (FDR <= 0.05). Among all studied genes, 3494 genes (23.2%) had at least one genome-wide significant eQTL and 13993 (93.2%) had at least one nominally significant eQTL.

Integration of gene expression in islets with T2D risk variants

Previous GWAS studies have reported 404 unique variants associated with T2D risk (Mahajan et al., 2018). These 404 SNPs resulted in a list of 383 target eGenes (genes whose expression is influenced by genetic variants), 41 of which were specifically related to general β-cell function. To further elucidate the role of these genes in T2D, we showed by enrichment analysis and bootstrapping these 383 eGenes, enrichment was significant for co-expression with *INS* but not with *GCG* or *SST* expression (figure 3A). For the list of β-cell related genes, enrichment was significant for INS as well as GCG. Genes which were significantly positively co-expressed with INS or SST(empirical P <= 0.05) tended to be upregulated in T2D (Figure 3B), and conversely negatively co-expressed genes were downregulated in T2D. *GCG* showed the opposite pattern; genes positively co-expressed with *GCG* were downregulated in T2D and vice versa.

Four genes were found to be both nominally significant eGenes in islets and downregulated in T2Ds – *CYB561*, *PKIG*, *PEX16*, *CKB* and *PTGDS*. Similarly 3 genes were eGenes in islets and downregulated in T2D – *GPRC5D-AS1*, *PARP14*, and *SNAP23*.

In order to functionally annotate all genes expressed in human pancreatic islets we created the Islet Gene View (IGW) web tool (see below).

Islet Gene View (IGW) - Features

Islet Gene View (IGW) will be accessible on the web page after a free registration process. IGW uses several common gene identifiers (*e.g.* gene symbol, Ensembl gene ID, and full gene name), and provides graphs of gene expression in relation to islet phenotypes and expression of other genes. Gene expression can be related to diabetes-related phenotypes such as age, BMI, and a

diagnosis of T2D, as well as glucose tolerance defined by HbA1c strata (normal glucose tolerance, NGT: HbA1c <6%, impaired glucose tolerance, IGT: HbA1c between 6-6.5%, and T2D: HbA1c > 6.5%). Of interest is also co-expression of other cell-specific genes such as insulin (*INS)*, glucagon (*GCG)* and islet amyloid polypeptide (*IAPP)*. The Ensembl gene ID is the primary key identifier for the expression database, and additional gene identifiers and other annotations were derived from Ensembl using the biomaRt API (Durinck et al., 2009). In addition to islet gene expression data, a 12-donor tissue panel of biopsies from fat, islets, liver and skeletal muscle are used to provide a comparison of gene expression in different tissues. Top 10 genes associated with islet phenotypes are shown in supplementary table 7.

As an example we here show one of the genes found to both be an eGene targeted by a T2D risk variant as well as being differentially upregulated in T2D vs NGT (*PTGDS*, *SNAP23* and *KCNQ1*).

PTGDS

In the integrative analysis of diabetes risk SNPs, *PTGDS* (Supplementary Figure 2) came up as the top hit as most strongly associated with T2D, while also being co-expressed with INS and targeted by a diabetes risk SNP eQTL. In Geneview, it is also nominally associated with BMI and HbA1c, and has a genome-wide significant eQTL signal from two linked loci (rs28592848 and rs28375538).

SNAP23

SNAP23 (Supplementary Figure 3) is another top hit from the integrative analysis of diabetes risk SNPs. In contrast to PTGDS this gene is nominally down-regulated in T2D (as well as in patients with HbA1c >=6.5%) and negatively co-expressed with INS. Additionally it was negatively co-expressed with the *SST* gene and negatively associated with purity possibly suggesting a higher expression in the exocrine pancreas.

To further demonstrate the utility of IGW, we generated IGW outputs for a few genes previously implicated in monogenic or polygenic diabetes.

KCNQ1

KCNQ1 expression was strongly negatively correlated with purity suggesting that it is also enriched in the exocrine pancreas (Supplementary Figure 4). It was significantly downregulated in patients with HbA1c >=6.5% but not in diagnosed T2D patients. Additionally it is negatively co-expressed with the *INS* gene (p = 0.048). Also, *KCNQ1* expression is also nominally influenced by rs8181588, a previously established T2D risk SNP located in an intron of *KCNQ1*.

MNX1

MNX1 (Supplementary Figure 5) was highly expressed in islets but not in fat, liver or skeletal

muscle. It was also positively correlated with purity and INS expression suggesting that it is a beta-specific gene.

PDX1

PDX1 was highly expressed in islets, and correlated positively correlated with islet purity and significantly associated with stimulatory index(P = 0.037) as well as purity (Supplementary Figure 6). In addition, it was nominally downregulated in T2D (P = 0.022) and showed strong correlation with HbA1c and BMI.

WFS1

WFS1 (Supplementary Figure 7) was positively correlated with purity and nominally upregulated in IGT but not in T2D patients. Additionally it was positively correlated with *INS* and *SST* expression.

All six genes were found to have at least one nominally significant eQTL, with lowest P-values for 6.3e-04 (*PDX1*), 9.8e-04 (*WFS1*), 1.2e-03(*KCNQ1*), 1.8e-03(*SNAP23*) and 3.2e-03 (*MNX1*).

The example of *CHL1,previously associated with T2D*

We also provide an IGW graph for the *CHL1* gene (Supplementary Figure 8), which previously has been shown to be associated with T2D (Belongie et al., 2017)..Furthermore in keeping with a previous study from our laboratory (Fadista et al. 2014) expression of the *CHL1* gene was downregulated in T2D islets from this larger data set (Supplementary Figure 2). Moreover the fasting insulin-associated SNP rs9841287 at the *CHL1* locus was found to be an eQTL for the *CHL1* and *CHL1-AS1* genes (P = 0.028, increasing allele = G). In support of the above findings, it was also negatively associated with high HbA1c.

Expression of the *SERPINE2* gene showed the greatest difference between T2D and NGT

The Serpin Family E Member 2 (*SERPINE2*) gene was selected due to the strong association of its expression with T2D. In addition to having high expression in islets, it is also expressed in adipose tissue, liver and muscle (figure 1A) . *SERPINE2* was also modestly correlated with purity (r2 = 0.064, p = 0.011, figure 1B).

The expression of the *SERPINE2* gene in the islets was greater than e.g. a well-known β-cell gene, the Potassium Voltage-gated Channel Subfamily J Member 11 gene (*KCNJ11,* rank 40.3%). The most highly expressed gene in the dataset was the glucagon (*GCG)* gene, followed by the Regenerating Family Member 1α (*REG1A)* gene and Mitochondrially Encoded Cytochrome C Oxidase I (*MT-CO1)* gene.*The SERPINE2* gene was most highly up-regulated in T2D compared with NGT ($P_{FDR} = 1.1 \cdot 10^{-5}$, $\log_2$ fold change = 0.58 or ~150% increase, Figure 1D). Expression of SERPINE2 showed of all genes the strongest correlation with HbA1c (r2 = 0.13, P_{fdr}= 9.5 x 10-7).

In (Figure 1F). The most downregulated gene in T2D vs NGT was the Phospholipase A1 Member A (*PLA1A*) gene ($\log_2$ fold change = -1.00, $P_{FDR} = 5.9 \cdot 10^{-5}$).

The Solute Carrier Family 2 Member 2 (*SLC2A2*) gene (Supplementary Figure 9), encoding the glucose transporter 2 (GLUT2), showed the strongest negative correlation with HbA1c. In support of this *SLC2A2* expression was also lower in islets from T2D vs NGT donors ($P_{FDR}= 3.2 \cdot 10^{-4}$, $\log_2$ fold change = -1.08)(Figure 1G).

Discussion

Islet GeneView (IGW) aims to provide information on gene expression from human islets in combination with diabetes-related phenotypes in islets from individuals with and without T2D. Thereby, Islet Gene View is complementary to other related databases such as the Islet eQTL explorer (Varshney et al., 2017), which connects genetic variation to islet gene expression and chromatin states in 112 samples, and GTEx (2013) which comprehensively characterizes the human transcriptome across many tissues, including some whole pancreas samples. GTEx, however, does not provide data on the islet transcriptome, which clearly is very different from that in the exocrine part of the pancreas.

Compared to previous publications on the same dataset (Fadista et al., 2014; Taneera et al., 2012), the current data set included188 islets from human organ donors, Compared to our previous report (Fadista et al., 2014), several refinements to the analytical pipeline have been made. We have introduced batch correction using ComBat. In addition, we have refined the methodology for calculating correlation P-values with higher specificity and stringency in detecting potential gene-gene correlations, which are now less dependent on batch effects. This reduces the influence of nonspecific inter-gene correlations resulting from the normalization procedure for gene expression. All gene-gene correlations have been pre-calculated in order to estimate the null distribution of the correlation values. This is computationally intensive, but only has to be done once.

In this new expanded dataset, we identified a total of 215 genes which showed genome-wide t(FDR<=0.05) and 6157 genes which showed nominal significance (P<=0.05) when comparing T2D with NGT). For the stratified HbA1c phenotype (T2D,IGT and NGT), 1789 genes were nominal DEGs between the glycemic clusters T2D,IGT and NGT compared with 1619 genes in the previous analysis (Fadista et al., 2014).

T2D has been ascribed to the inability of pancreatic beta cells to produce sufficient insulin to meet the demands due to increased insulin resistance, therefore human islets are a key tissue in T2D pathogenesis. Genome wide association studies have reported 403 genetic loci association with T2D risk whereas their function is largely unknown. To further leverage this data to elucidate function at previously reported T2D loci, we cross referenced the data with gene expression and eQTL data in IGW to identify genes whose expression is modulated by T2D risk loci and further elucidate their association with islet traits. We here identified multiple gene loci which are transcriptionally regulated by T2D risk loci, which were also differentially expressed in type 2 diabetes patients, and correlated with INS secretion (Figure 3). Two examples of these genes, PTGDS and SNAP23, are described in further detail below. In total, this demonstrates how IGW

can be used to identify complex relationships between genetics, transcriptomics and diabetic phenotypes.

The *PTGDS* gene (Supplementary Figure 3) encodes an enzyme which converts prostaglandin H2 to prostaglandin D2. High PGD2 levels in turn have been found to increase insulin sensitivity(Fujitani et al., 2010), and knock-out of PTGDS gives the reverse effect resulting in insulin resistance, nephropathy and atherosclerosis through effects on adipose tissue(Ragolia et al., 2005). While we found little data on the correlation between diabetes and *PTGDS* in islet cells in previous scientific literature, our results suggests that expression of the gene in islets is correlated both with diabetes status, BMI and INS expression, indicating that *PTGDS* function and prostaglandin levels may be connected to insulin secretion as well as peripheral insulin resistance.

The *SNAP23* (Supplementary Figure 4) protein product is involved in vesicle fusion and is essential for peripheral insulin sensitivity through *GLUT4* transportation(Kawanishi et al., 2000). The role is reversed in islets, where an endocrine islet-specific knockout of *SNAP23* in mice shows improved glucose tolerance and insulin secretion(Kunii et al., 2016). Our results correlate with these previous findings in islets in several ways - *SNAP23* is nominally significantly downregulated in diabetics, negatively correlated with INS expression, and interestingly enough, also negatively correlated with somatostatin expression which has not been previously reported.

We then examined genes previously implicated in monogenic as well as polygenic diabetes using IGW. Monogenic diabetes mutations target genes that are essential for sufficient insulin secretion, while polygenic diabetes genes are more diverse in their different functions.

KCNQ1 (Supplementary Figure 5) is a voltage-gated potassium channel that affects the repolarization of the islet cell membrane. Earlier studies suggest that *KCNQ1* overexpression negatively affects insulin secretion through premature repolarization(Kasuga, 2011). Our new data shows also a negative correlation with INS expression. The explanation for these somewhat disparate findings is unclear but could involve a compensatory response from the beta cell to slow down repolarization.

PDX1 (Supplementary Figure 6) encodes another transcription factor that is required for beta-cell maturation and maintenance (Spaeth et al., 2017). We here add new information showing that *PDX1* is downregulated in T2D supporting previous findings that its expression correlates with chronic hyperglycemia (Fujimoto and Polonsky, 2009) .

The *WFS1* gene (Supplementary Figure 7) produces the wolframin protein. Inactivation of *WFS1* further increases ER stress through wolframin resulting in progressive beta cell death (Fonseca et al., 2005). This is also shown in Geneview where WFS1 shows clear positive correlation with *INS* expression. Interestingly somatostatin is also significantly correlated with *WFS1*.

Somewhat surprisingly the stimulatory index (SI) did not show genome-wide significant correlation with expression of any gene. There might be several explanations for this. First, although the largest dataset of its kind, we still have limited power to obtain genome wide significance for a crude measure like SI. Despite this, the information may be useful as exploratory information.

IGW provides information on co-expression of genes with secretory genes from α-cells, β-cells, δ-cells, ε-cells and PP-cells. *IRF2BP1*, a co-repressor of Interferon Regulatory Factor 2 (*IRF2*) (Mashima et al., 2011), showed the strongest positive correlation with INS. *IRF2* is a transcription factor that suppresses the expression of interferon α and β, which might affect islet cell survival and proliferation. Interestingly, *IRF2* is negatively correlated with *INS* expression in this dataset

The GC-rich promoter binding protein 1 like 1 gene (*GPBP1L1*) showed the strongest negative correlation with INS expression. While the function of GPBP1L1 is unknown, it likely acts as a transcription factor, like its paralogous gene *GPBP1* (Hsu et al., 2003). This could indicate that GPBP1L1 expression is suppressed by INS expression or insulin secretion or is downregulated in INS-producing beta-cells.

Bidirectional expression of several genes with INS and GCG was observed (Figure 2). Interestingly, expression of 6 genes correlated positively with expression of *INS* and negatively with expression of *GCG*, *e.g. WIZ*, *NUDC*, *IMP4*, *ZNF787*, *KLHL21*, and *ISG20*. Of them, the Widely Interspaced Zinc Finger Motifs (*WIZ)* gene is part of the G9a/GLP complex, involved in H_3K_9 methylation of CTCF binding sites resulting in suppressed gene expression. (Isbel et al.). Therefore, the G9a/GLP complex might act as a switch that regulates the balance between insulin and glucagon secretion and has been shown to affect insulin signaling in the liver(Xue et al., 2018). *NUDC* and *KLHL21* play important roles in cell division and could possibly influence the mass of a particular cell population. *IMP4* affects pre-rRNA processing and thus production of ribosomes. *ISG20* is an exoribonuclease that is stimulated by interferons, and *ZNF787* is a gene with unknown function.

SLC2A2, encoding the glucose transporter *GLUT2* plays a key role in islet function (Thorens, 2015). Its expression was downregulated already in patients with IGT and even more so in T2D. Common variants in *SLC2A2* associate with T2D, fasting glucose and HbA1c as well as insulin secretion in previous GWAS (Manning et al., 2012; Scott et al., 2017). Down-regulation of *SLC2A2* could be epigenetically mediated as it has been shown to be methylated in islets from patients with T2D (Volkov et al., 2017). Common variants in *SLC2A2* also associate with the response to metformin (Zhou et al., 2016).

The SNP rs9841287 in the *CHL1* gene has previously been associated with fasting insulin concentrations, with the A allele (Frequency = 0.69 in 1000 Genomes) increasing insulin levels i ($P = 7.78 \cdot 10^{-9}$) (Manning et al., 2012; Scott et al., 2017)).The IGW data suggest that high expression

of *CHL1* in islets could have beneficial effects on beta-cell function.

An important aspect of Islet Gene View is that it can compare expression of different genes as simple histograms. The most highly expressed genes in islets were *GCG*, followed by *REG1A*, *MT-CO1* and *MT-ATP6*. While the comparability between genes is higher with RNA-Sequencing than with microarrays, the gene sequence might influence the sequencing rate of different transcripts. Information on purity provides allows for a partial separation of expression patterns in endocrine and exocrine tissue.

Taken together Islet Gene View is a tool to facilitate research on human pancreatic islets to and is therefore made accessible to the entire scientific community. However, we need to remember that it has limitations, it is based upon a relatively small sample size of islets from patients recruited through intensive care units. The data are observational, but cannot distinguish correlation from causation. However, the exploratory use of IGW could help designing more comprehensive functional follow-up studies.

Acknowledgements

Human pancreatic islets, muscle, fat and liver samples were provided by The Nordic Network for Clinical Islet Transplantation. The work is this paper has been financially supported by grants from the Swedish Research Council: strategic research environment grant (EXODIAB, 2009-1039) and project grant (2015-2558) to LG, networking grant (2015-06722) to RP; collaborative grant from the Swedish Foundation for Strategic Research to the Lund Unversity Diabetes Centre (IRC 15-0067); JDRF(award 31-2008-416); Diabetes Wellness grant to RP (720-858-16JDWG); collaborative grants with Regeneron and Eli Lilly to LG. We thank Mattias Borell, Maria Sterner, Malin Neptin and Malin Svensson for technical support. We also want to express our deepest gratitude to the deceased organ donors as well as to their relatives.

Author contributions

OA, PS and RBP analysed the data and designed the study. OA, RBP, OH and LG interpreted the results and wrote the manuscript, with additional input from ER, HM, OK and UK. EOL and UK performed laboratory experiments and measurements. RBP and LG supervised the project. LC contributed the data. All authors provided critical feedback and helped shape the research, analysis and manuscript.

Conflicts of Interest

No conflict of interest is reported.

References

Belongie, K.J., Ferrannini, E., Johnson, K., Andrade-Gordon, P., Hansen, M.K., and Petrie, J.R. (2017). Identification of novel biomarkers to monitor β-cell function and enable early detection of type 2 diabetes risk. PLOS ONE *12*, e0182932.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Methodol. *57*, 289–300.

Fadista, J., Vikman, P., Laakso, E.O., Mollet, I.G., Esguerra, J.L., Taneera, J., Storm, P., Osmark, P., Ladenvall, C., Prasad, R.B., et al. (2014). Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. Proc. Natl. Acad. Sci. U. S. A. *111*, 13924–13929.

Fonseca, S.G., Fukuma, M., Lipson, K.L., Nguyen, L.X., Allen, J.R., Oka, Y., and Urano, F. (2005). WFS1 Is a Novel Component of the Unfolded Protein Response and Maintains Homeostasis of the Endoplasmic Reticulum in Pancreatic β-Cells. J. Biol. Chem. *280*, 39609–39615.

Friberg, A.S., Brandhorst, H., Buchwald, P., Goto, M., Ricordi, C., Brandhorst, D., and Korsgren, O. (2011). Quantification of the islet product: presentation of a standardized current good manufacturing practices compliant system with minimal variability. Transplantation *91*, 677–683.

Fujimoto, K., and Polonsky, K.S. (2009). Pdx1 and other factors that regulate pancreatic β-cell survival. Diabetes Obes. Metab. *11*, 30.

Fujitani, Y., Aritake, K., Kanaoka, Y., Goto, T., Takahashi, N., Fujimori, K., and Kawada, T. (2010). Pronounced adipogenesis and increased insulin sensitivity caused by overproduction of prostaglandin D2 in vivo. FEBS J. *277*, 1410–1419.

Hsu, L.-C., Liu, S., Abedinpour, F., Beech, R.D., Lahti, J.M., Kidd, V.J., Greenspan, J.A., and Yeung, C.-Y. (2003). The murine G+C-rich promoter binding protein mGPBP is required for promoter-specific transcription. Mol. Cell. Biol. *23*, 8773–8785.

Isbel, L., Prokopuk, L., Wu, H., Daxinger, L., Oey, H., Spurling, A., Lawther, A.J., Hale, M.W., and Whitelaw, E. Wiz binds active promoters and CTCF-binding sites and is required for normal behaviour in the mouse. ELife *5*.

Kasuga, M. (2011). KCNQ1, a susceptibility gene for type 2 diabetes. J. Diabetes Investig. *2*, 413–414.

Kawanishi, M., Tamori, Y., Okazawa, H., Araki, S., Shinoda, H., and Kasuga, M. (2000). Role of SNAP23 in insulin-induced translocation of GLUT4 in 3T3-L1 adipocytes. Mediation of complex formation between syntaxin4 and VAMP2. J. Biol. Chem. *275*, 8240–8247.

Kunii, M., Ohara-Imaizumi, M., Takahashi, N., Kobayashi, M., Kawakami, R., Kondoh, Y., Shimizu, T., Simizu, S., Lin, B., Nunomura, K., et al. (2016). Opposing roles for

SNAP23 in secretion in exocrine and endocrine pancreatic cells. J Cell Biol *215*, 121–138.

Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. *15*, R29.

Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E., and Storey, J.D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinforma. Oxf. Engl. *28*, 882–883.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinforma. Oxf. Engl. *30*, 923–930.

Mahajan, A., Taliun, D., Thurner, M., Robertson, N.R., Torres, J.M., Rayner, N.W., Steinthorsdottir, V., Scott, R.A., Grarup, N., Cook, J.P., et al. (2018). Fine-mapping of an expanded set of type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. BioRxiv 245506.

Manning, A.K., Hivert, M.-F., Scott, R.A., Grimsby, J.L., Bouatia-Naji, N., Chen, H., Rybin, D., Liu, C.-T., Bielak, L.F., Prokopenko, I., et al. (2012). A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nat. Genet. *44*, 659–669.

Mashima, H., Sato, T., Horie, Y., Nakagawa, Y., Kojima, I., Ohteki, T., and Ohnishi, H. (2011). Interferon regulatory factor-2 regulates exocytosis mechanisms mediated by SNAREs in pancreatic acinar cells. Gastroenterology *141*, 1102-1113.e1-8.

R Core Team (2016). R: A Language and Environment for Statistical Computing (Vienna, Austria: R Foundation for Statistical Computing).

Ragolia, L., Palaia, T., Hall, C.E., Maesaka, J.K., Eguchi, N., and Urade, Y. (2005). Accelerated glucose intolerance, nephropathy, and atherosclerosis in prostaglandin D2 synthase knock-out mice. J. Biol. Chem. *280*, 29946–29955.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Scott, R.A., Scott, L.J., Mägi, R., Marullo, L., Gaulton, K.J., Kaakinen, M., Pervjakova, N., Pers, T.H., Johnson, A.D., Eicher, J.D., et al. (2017). An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. Diabetes *66*, 2888–2902.

Spaeth, J.M., Gupte, M., Perelis, M., Yang, Y.-P., Cyphert, H., Guo, S., Liu, J.-H., Guo, M., Bass, J., Magnuson, M.A., et al. (2017). Defining a Novel Role for the Pdx1 Transcription Factor in Islet β-Cell Maturation and Proliferation During Weaning. Diabetes *66*, 2830–2839.

Taneera, J., Lang, S., Sharma, A., Fadista, J., Zhou, Y., Ahlqvist, E., Jonsson, A., Lyssenko, V., Vikman, P., Hansson, O., et al. (2012). A Systems Genetics Approach Identifies Genes and Pathways for Type 2 Diabetes in Human Islets. Cell Metab. *16*, 122–134.

Thorens, B. (2015). GLUT2, glucose sensing and glucose homeostasis. Diabetologia *58*, 221–232.

Varshney, A., Scott, L.J., Welch, R.P., Erdos, M.R., Chines, P.S., Narisu, N., Albanus, R.D., Orchard, P., Wolford, B.N., Kursawe, R., et al. (2017). Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. Proc. Natl. Acad. Sci. *114*, 2301–2306.

Venables, W.N., and Ripley, B.D. (2002). Modern Applied Statistics with S (New York: Springer).

Volkov, P., Bacos, K., Ofori, J.K., Esguerra, J.L.S., Eliasson, L., Rönn, T., and Ling, C. (2017). Whole-Genome Bisulfite Sequencing of Human Pancreatic Islets Reveals Novel Differentially Methylated Regions in Type 2 Diabetes Pathogenesis. Diabetes *66*, 1074–1085.

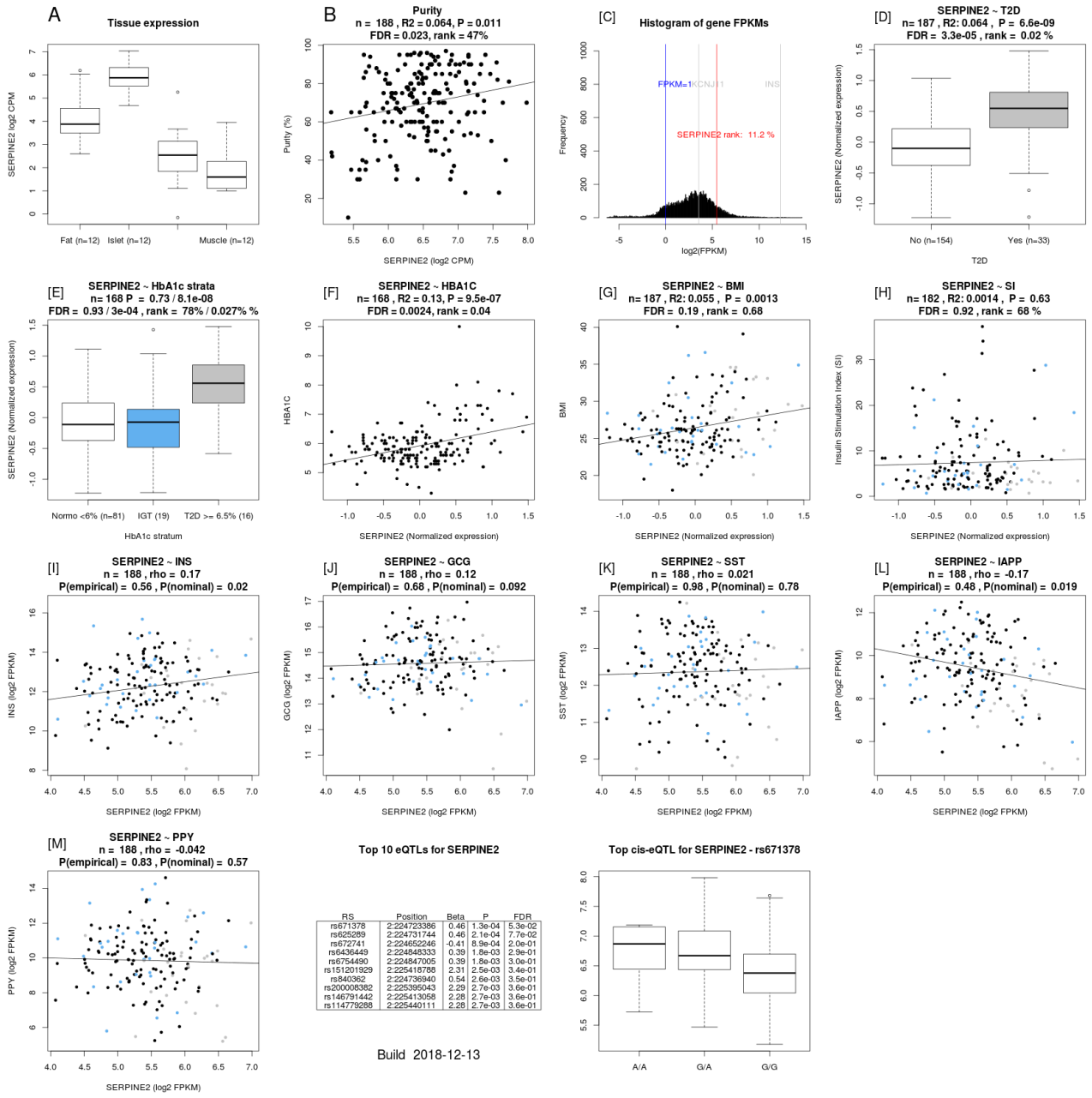
Xue, W., Huang, J., Chen, H., Zhang, Y., Zhu, X., Li, J., Zhang, W., Yuan, Y., Wang, Y., Zheng, L., et al. (2018). Histone methyltransferase G9a modulates hepatic insulin signaling via regulating HMGA1. Biochim. Biophys. Acta *1864*, 338–346.

Zhou, K., Yee, S.W., Seiser, E.L., van Leeuwen, N., Tavendale, R., Bennett, A.J., Groves, C.J., Coleman, R.L., van der Heijden, A.A., Beulens, J.W., et al. (2016). Variation in the glucose transporter gene SLC2A2 is associated with glycemic response to metformin. Nat. Genet. *48*, 1055–1059.

(2013). The Genotype-Tissue Expression (GTEx) project. Nat. Genet. *45*, 580–585., 580–585.

***Figure** 1: Example image from Islet Gene View of SERPINE2. A: expression of the gene in fat (F), islets (I), liver (L), and muscle (M), in the same pool of 12 individuals. B: Gene expression as a function of purity, defined as the percentage of endocrine tissue. C: Expression of the selected gene in relation to other genes*

in islets. D-H: Gene expression in relation to several diabetes-related phenotypes, i.e. T2D diagnosis (figure 1D), HbA1c stratum (1E), continuous HbA1c (1F), BMI (1G), and stimulatory index (1H). Test statistics are reported, namely: coefficient of determination (R2), nominal P-value, and percentage rank among all genes as calculated based on sorted P-values. I-M: Gene expression in relation to the secretory genes INS (figure 1I), GCG (1J), SST (1K), PPY (1L), and IAPP (1M). Spearman's ρ (r) and the P-value of the gene based on the empirical correlation distribution is reported. INS = Insulin, GCG = glucagon, SST = somatostatin, PPY = pancreatic polypeptide, IAPP = islet amyloid polypeptide.



RS	Position	Beta	P	FDR
rs671378	2:224723386	0.46	1.3e-04	5.3e-02
rs625289	2:224731744	0.46	2.1e-04	7.7e-02
rs672741	2:224652246	-0.41	8.9e-04	2.0e-01
rs6436449	2:224848333	0.39	1.8e-03	2.9e-01
rs6754490	2:224847005	0.39	1.8e-03	3.0e-01
rs151201929	2:225418788	2.31	2.5e-03	3.4e-01
rs840362	2:224736940	0.54	2.6e-03	3.5e-01
rs200008382	2:225395043	2.29	2.7e-03	3.6e-01
rs146791442	2:225413058	2.28	2.7e-03	3.6e-01
rs114779288	2:225440111	2.28	2.7e-03	3.6e-01

***Figure** 2: Venn diagram of genes coexpressed with secretory genes. A: Overlap of positively co-expressed genes (ρ ≥ 0, P ≤ 0.05) with insulin (INS), glucagon (GCG), somatostatin (SST) and islet amyloid polypeptide (IAPP). B: Overlap of negatively co-expressed genes (ρ ≤ 0, P ≤ 0.05). C: overlap of all co-expressed genes regardless of direction (P ≤ 0.05). D: Overlapping genes where co-expression is negative for one secretory gene and positive in the other (or the reverse).*

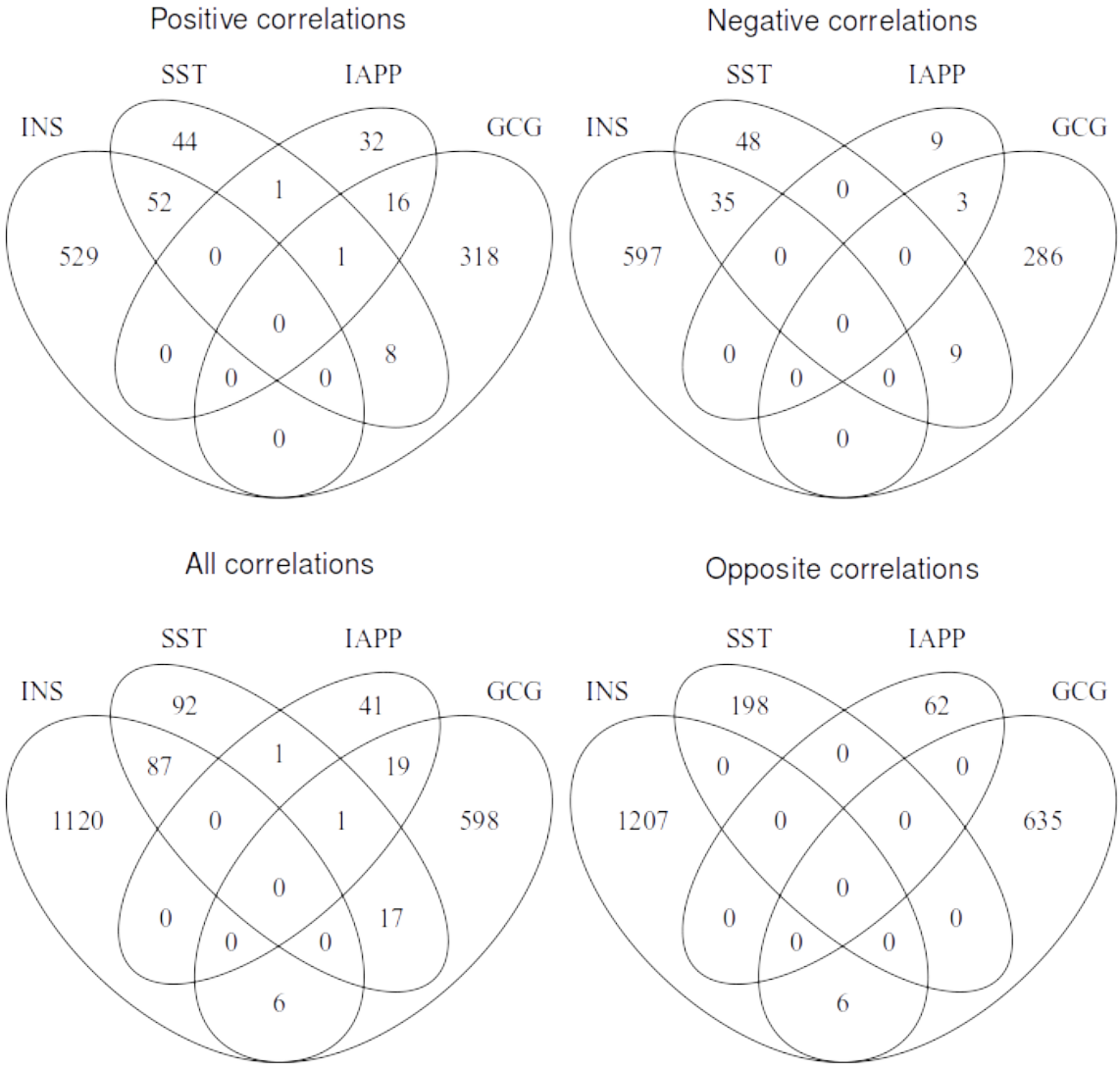
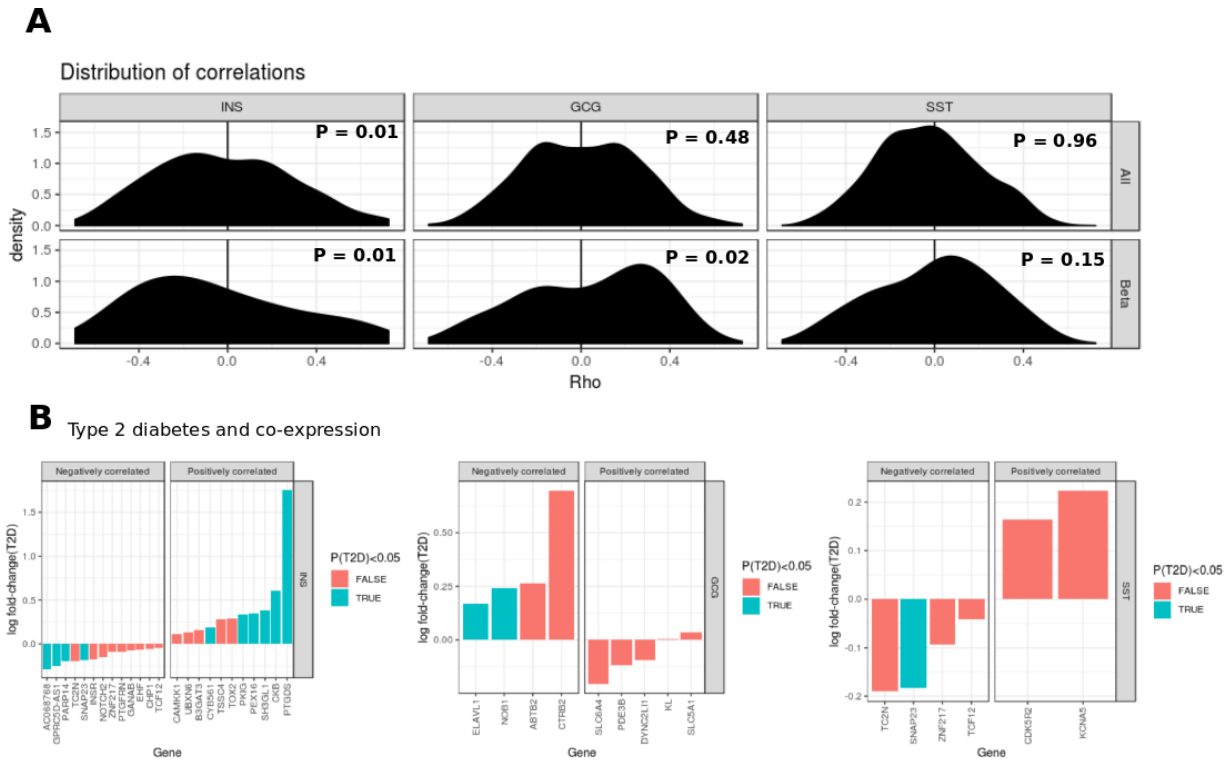
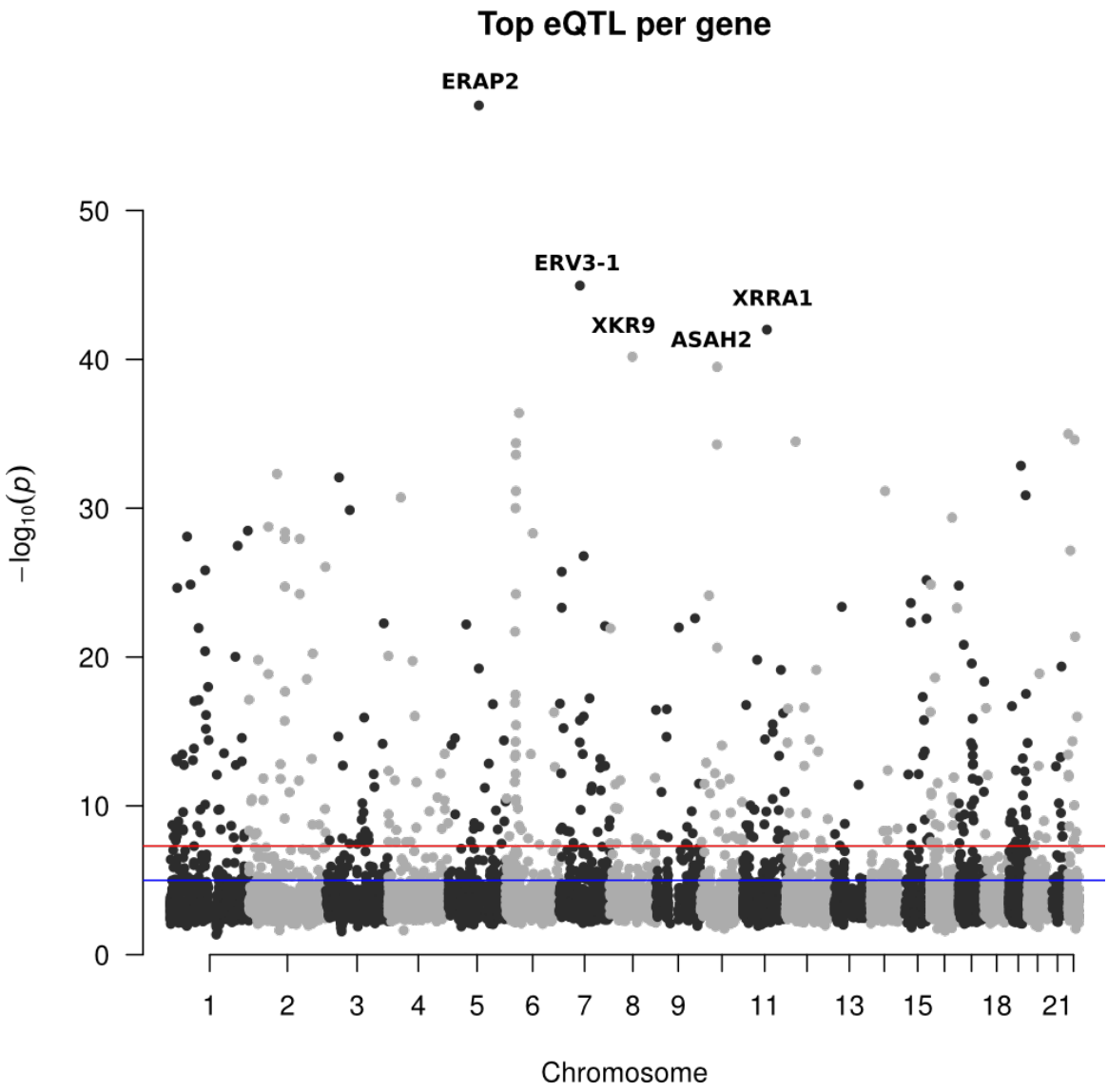


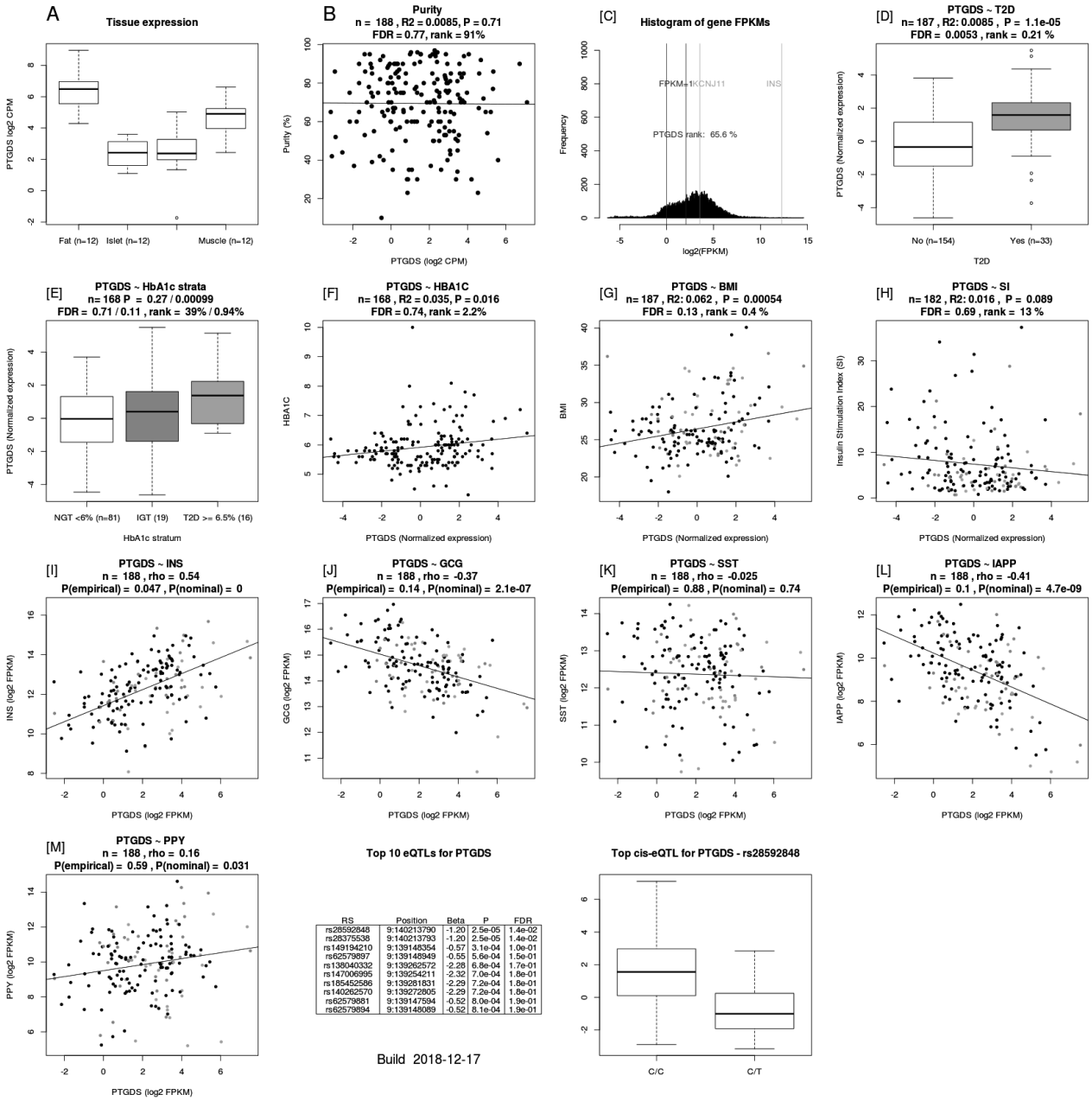
Figure 3: (A) Distribution of genes whose expression correlates with INS, GCG and SST for All genes (top) and beta cell function associated genes (bottom). (B) T2D eGenes and their association with INS, GCG and SST expression



Supplementary Figure 1: Manhattan plot of eQTLs with the lowest P-value per gene. The top 5 eQTLs are marked with their target gene.

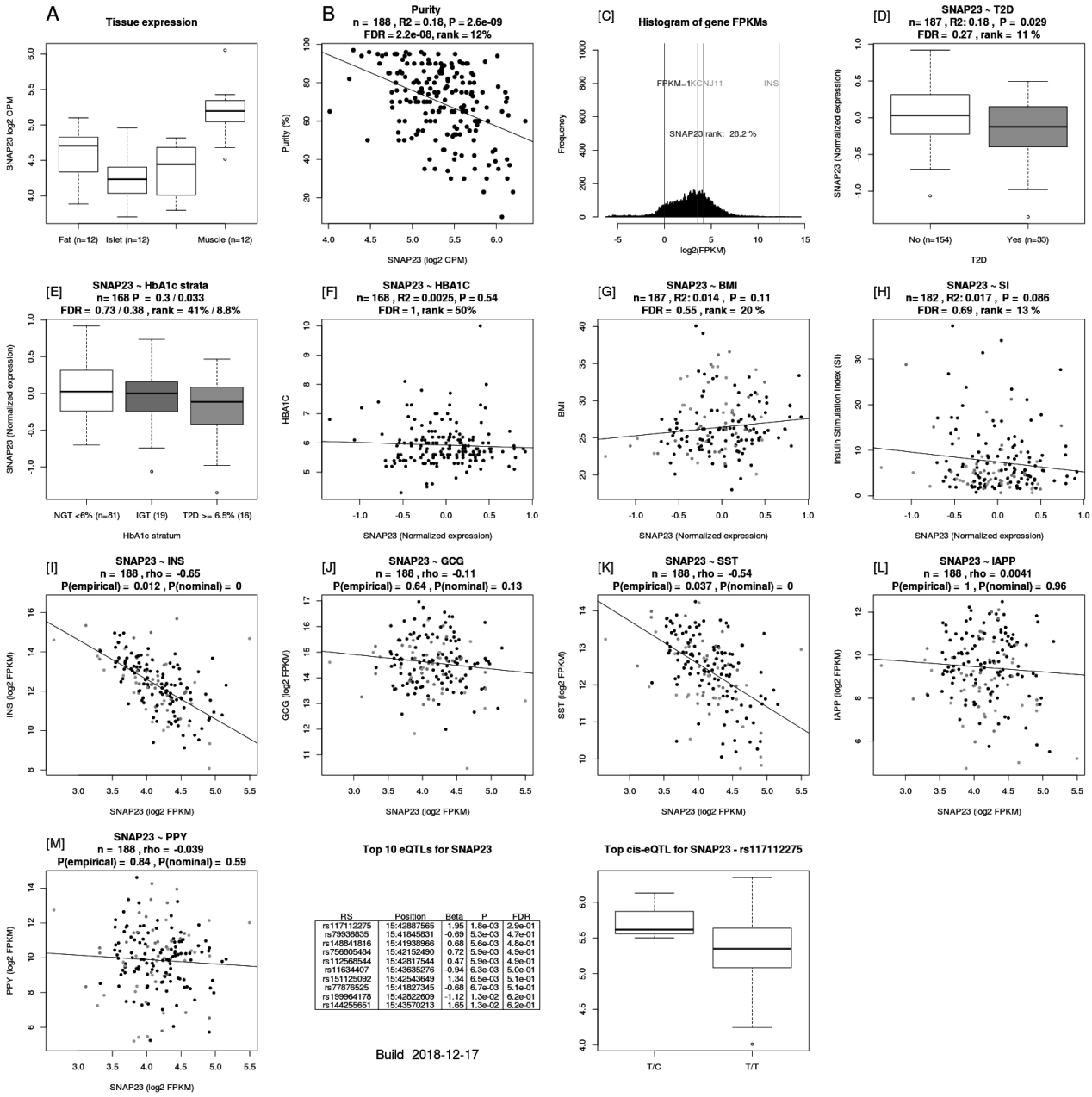


Supplementary Figure 2: Islet Gene View image of PTGDS.



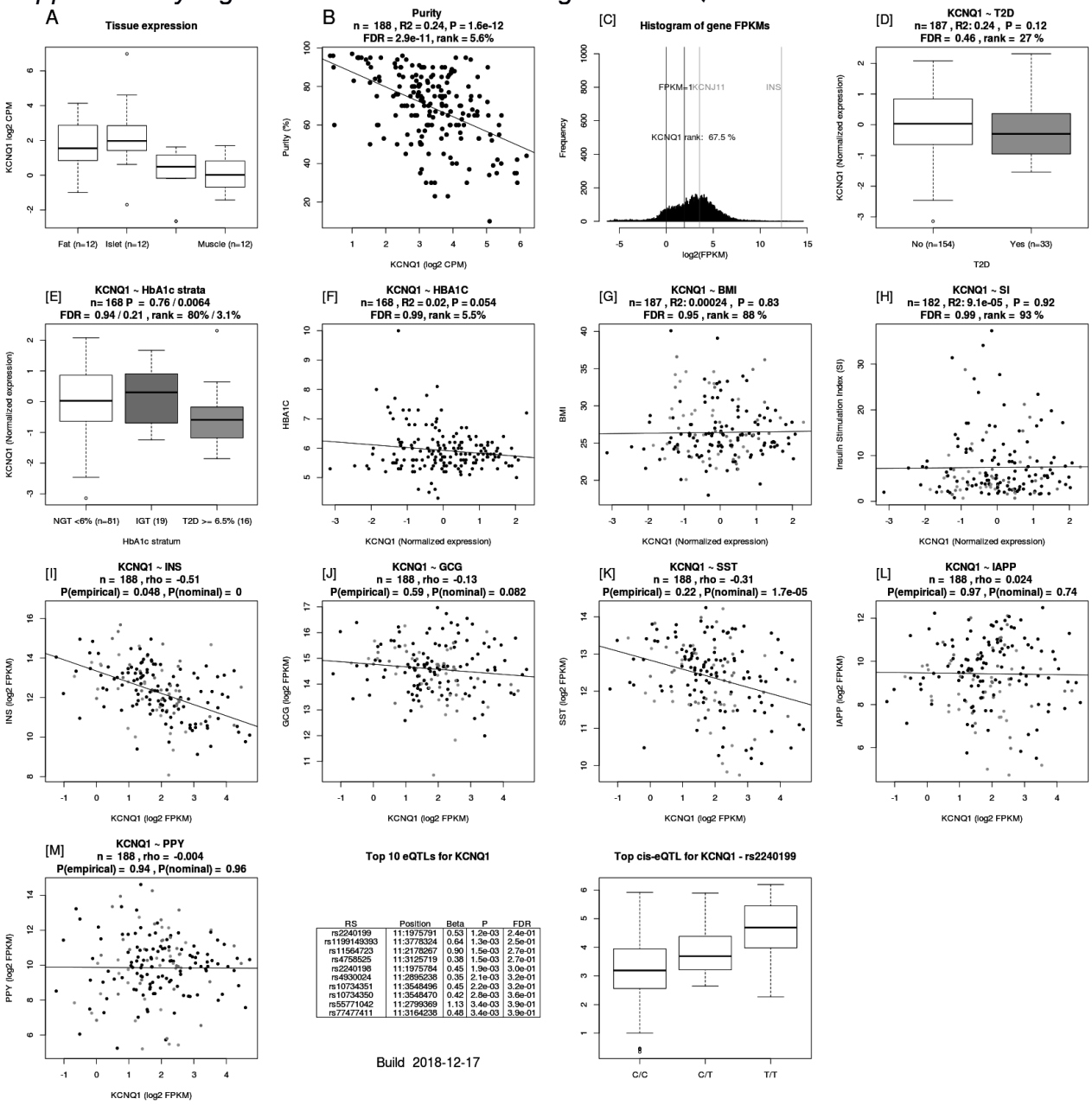
RS	Position	Beta	P	FDR
rs28592848	9:140213790	-1.20	2.5e-05	1.4e-02
rs28375538	9:140213793	-1.20	2.5e-05	1.4e-02
rs149194210	9:139148354	-0.57	3.1e-04	1.0e-01
rs62579897	9:139148949	-0.55	5.6e-04	1.5e-01
rs138040332	9:139262572	-2.28	6.8e-04	1.7e-01
rs147006995	9:139254211	-2.32	7.0e-04	1.8e-01
rs1854525586	9:139281831	-2.29	7.2e-04	1.8e-01
rs140262570	9:139272805	-2.29	7.2e-04	1.8e-01
rs62579881	9:139147594	-0.52	8.0e-04	1.9e-01
rs62579894	9:139148089	-0.52	8.1e-04	1.9e-01

Supplementary Figure 3: Islet Gene View image of SNAP23



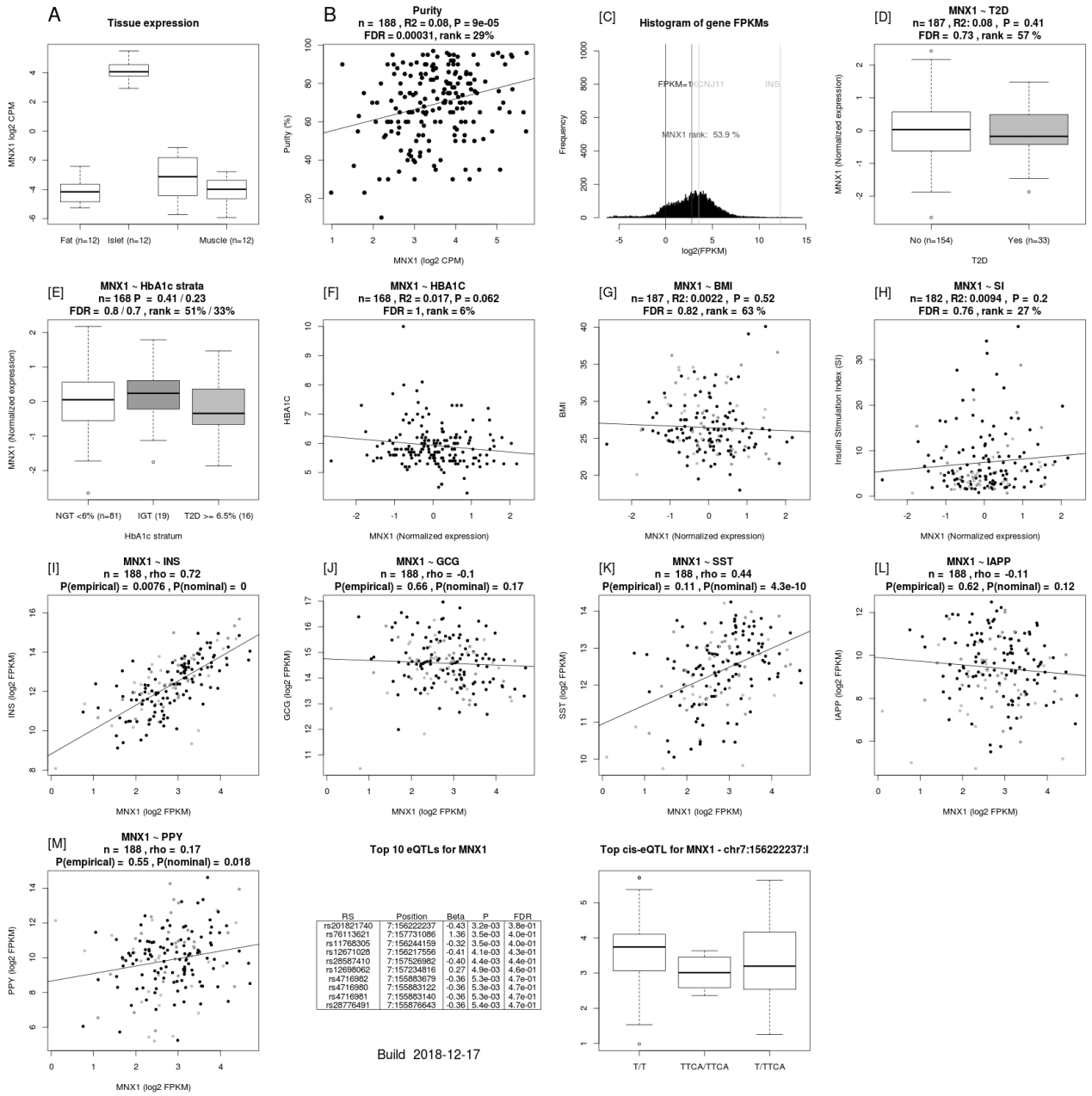
RS	Position	Beta	P	FDR
rs117112275	15:42887565	1.95	1.8e-03	2.9e-01
rs79936835	15:41845831	-0.69	5.3e-03	4.7e-01
rs148841816	15:41938966	0.68	5.6e-03	4.8e-01
rs756805484	15:42152490	0.72	5.9e-03	4.9e-01
rs112568544	15:42817544	0.47	5.9e-03	4.9e-01
rs111634407	15:43635276	-0.94	6.3e-03	5.0e-01
rs151125092	15:42543649	1.34	6.5e-03	5.1e-01
rs77876525	15:41827345	-0.68	6.7e-03	5.1e-01
rs199964178	15:42822609	-1.12	1.3e-02	6.2e-01
rs144255651	15:43570213	1.65	1.3e-02	6.2e-01

Supplementary Figure 4: Islet Gene View image of KCNQ1



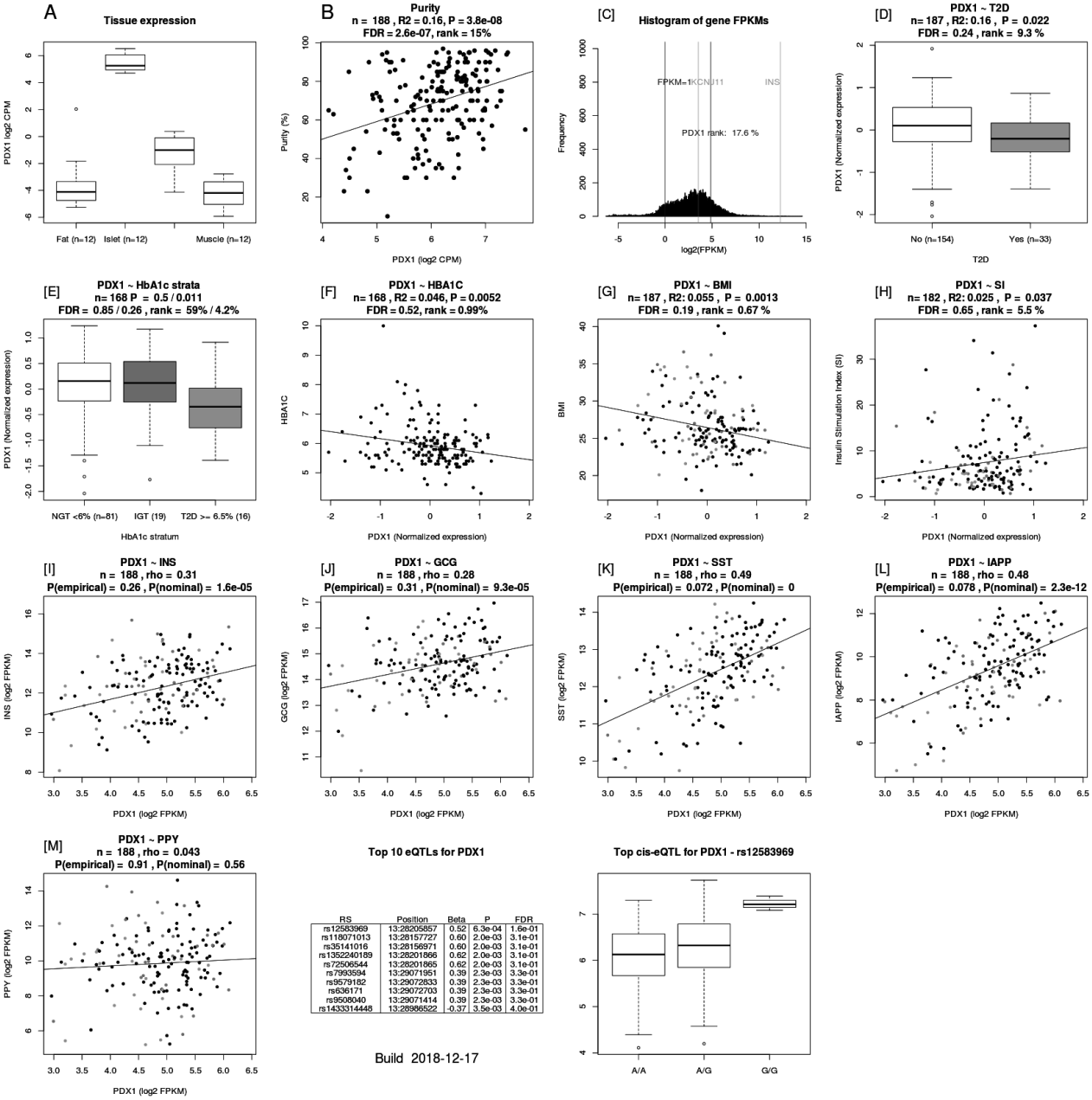
RS	Position	Beta	P	FDR
rs2240199	11:1975791	0.53	1.2e-03	2.4e-01
rs1199149393	11:3778324	0.64	1.3e-03	2.5e-01
rs11564723	11:2178267	0.90	1.5e-03	2.7e-01
rs4758525	11:3125719	0.38	1.5e-03	2.7e-01
rs2240198	11:1975784	0.45	1.9e-03	3.0e-01
rs4930024	11:2895238	0.35	2.1e-03	3.2e-01
rs10734351	11:3548496	0.45	2.2e-03	3.2e-01
rs10734350	11:3548470	0.42	2.8e-03	3.8e-01
rs55771042	11:2799369	1.13	3.4e-03	3.9e-01
rs77477411	11:3164238	0.48	3.4e-03	3.9e-01

Supplementary Figure 5: Islet Gene View image of MNX1



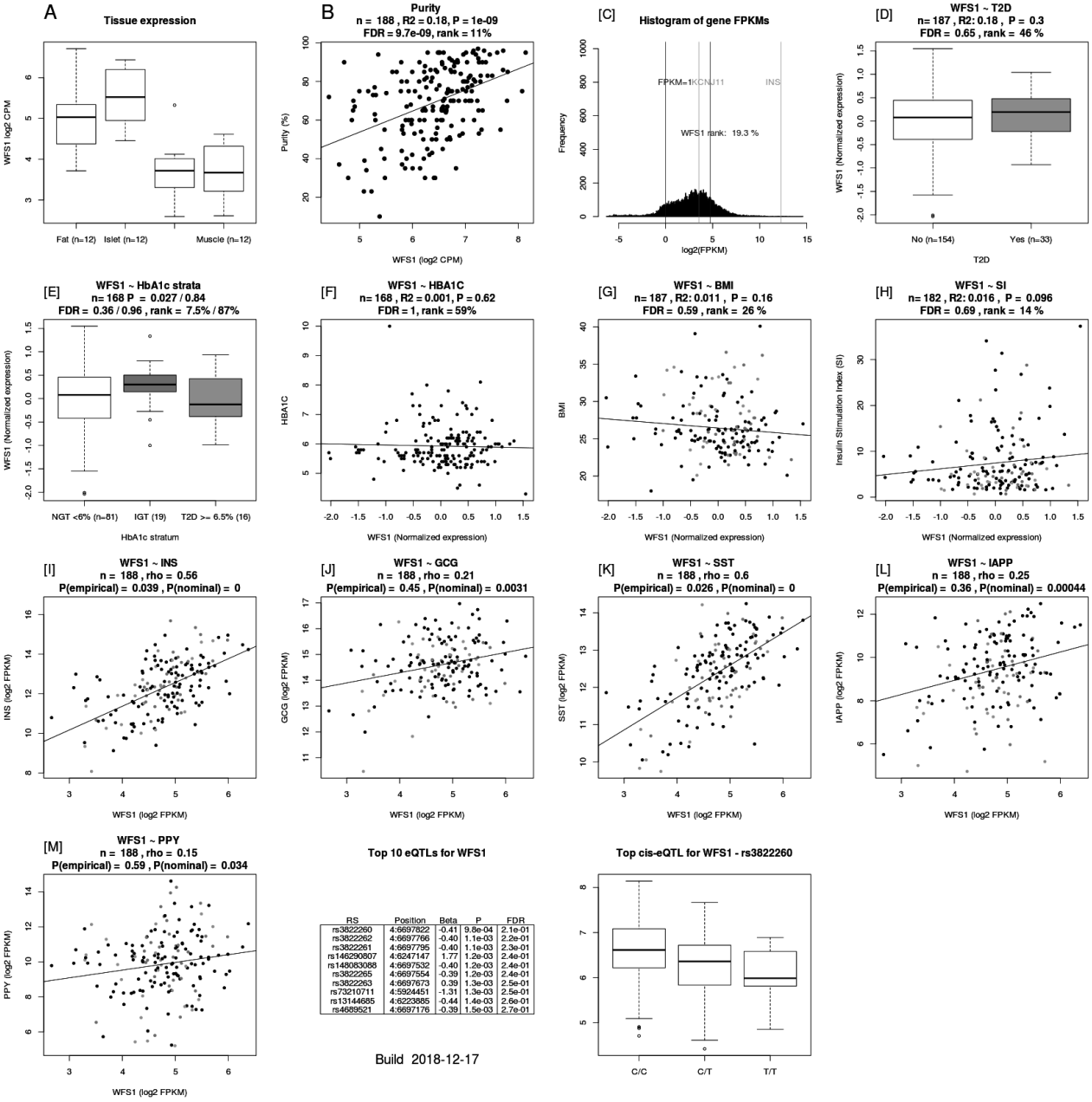
RS	Position	Beta	P	FDR
rs201821740	7:156222237	-0.43	3.2e-03	3.8e-01
rs76113621	7:157731086	1.36	3.5e-03	4.0e-01
rs11768305	7:156244159	-0.32	3.5e-03	4.0e-01
rs12671028	7:156217556	-0.41	4.1e-03	4.3e-01
rs28587410	7:157526982	-0.40	4.4e-03	4.4e-01
rs12698062	7:157234816	0.27	4.9e-03	4.6e-01
rs4716982	7:155883679	-0.36	5.3e-03	4.7e-01
rs4716980	7:155883122	-0.36	5.3e-03	4.7e-01
rs4716981	7:155883140	-0.36	5.3e-03	4.7e-01
rs28776491	7:155876643	-0.36	5.4e-03	4.7e-01

Supplementary Figure 6: Islet Gene View image of PDX1



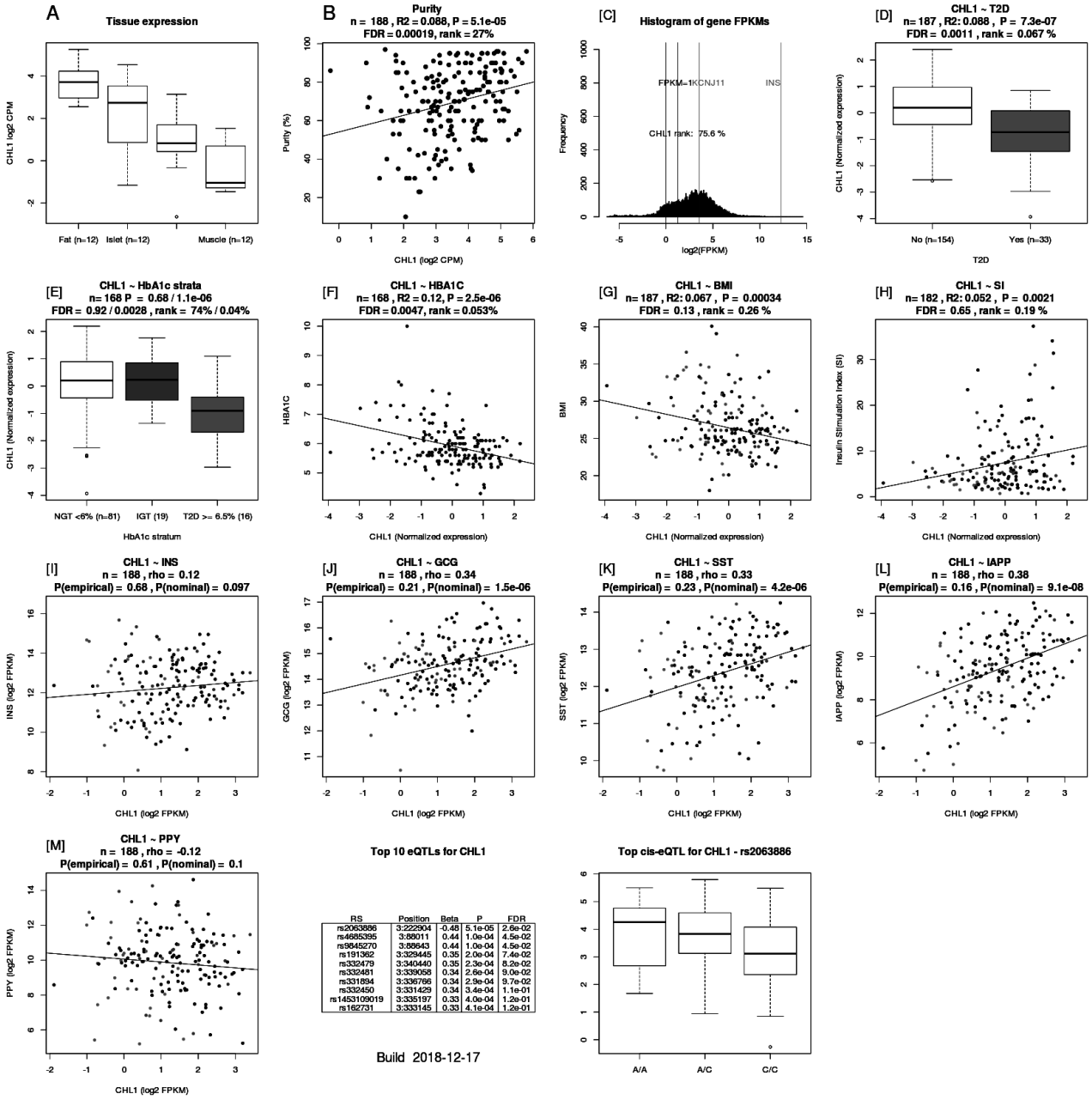
RS	Position	Beta	P	FDR
rs12583969	13:28205857	0.52	6.3e-04	1.6e-01
rs118071013	13:28157727	0.60	2.0e-03	3.1e-01
rs35141016	13:28156971	0.60	2.0e-03	3.1e-01
rs1352240189	13:28201866	0.62	2.0e-03	3.1e-01
rs72506544	13:28201865	0.62	2.0e-03	3.1e-01
rs7993594	13:29071951	0.39	2.3e-03	3.3e-01
rs9579182	13:29072833	0.39	2.3e-03	3.3e-01
rs836171	13:29072703	0.39	2.3e-03	3.3e-01
rs9508040	13:29071414	0.39	2.3e-03	3.3e-01
rs1433314448	13:28986522	-0.37	3.5e-03	4.0e-01

Supplementary Figure 7: Islet Gene View image of WFS1



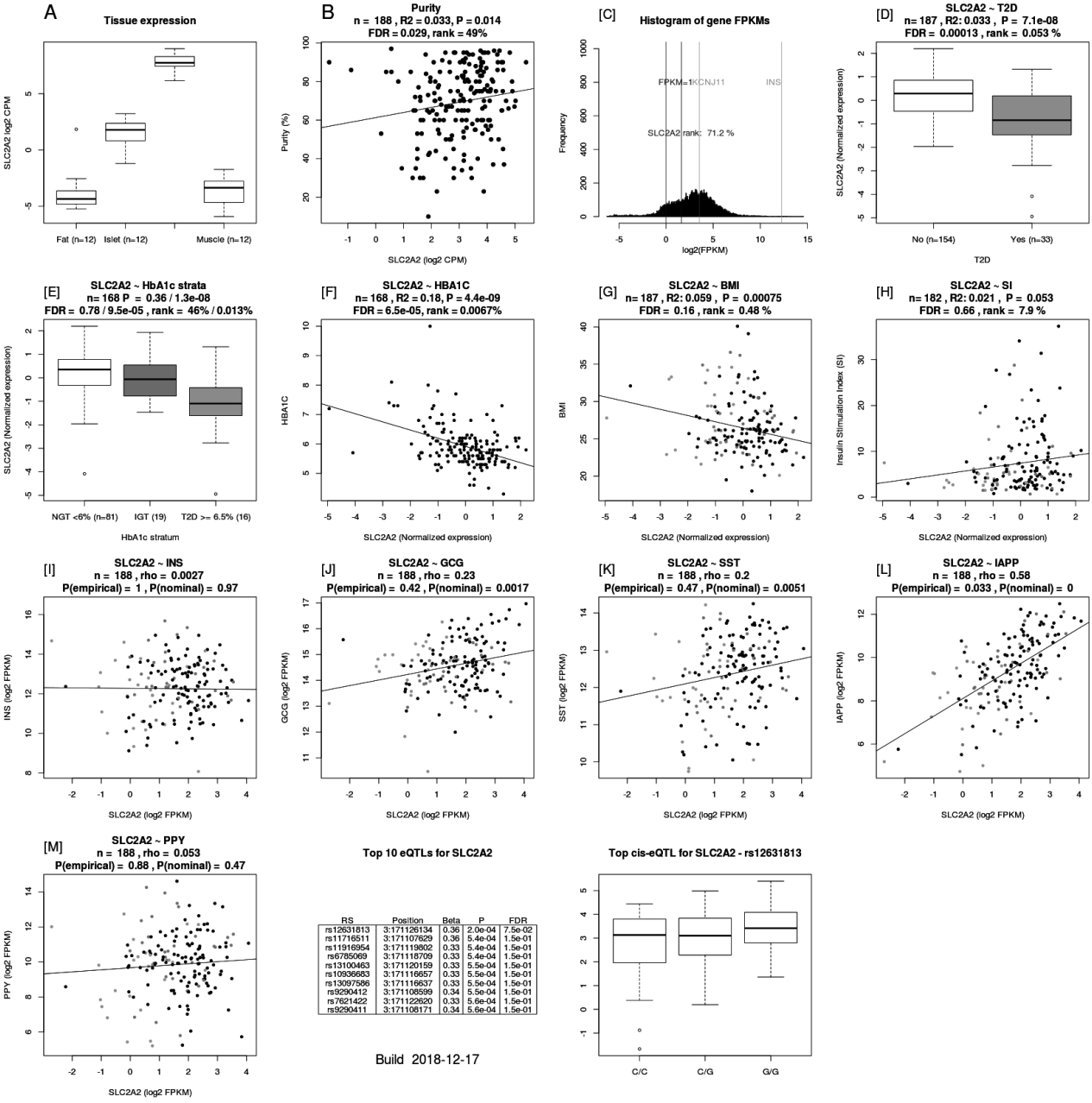
RS	Position	Beta	P	FDR
rs3822260	4:6697822	-0.41	9.8e-04	2.1e-01
rs3822262	4:6697766	-0.40	1.1e-03	2.2e-01
rs3822261	4:6697795	-0.40	1.1e-03	2.3e-01
rs146290807	4:6247147	1.77	1.2e-03	2.4e-01
rs148083088	4:6697532	-0.40	1.2e-03	2.4e-01
rs3822265	4:6697554	-0.39	1.2e-03	2.4e-01
rs3822263	4:6697673	0.39	1.3e-03	2.5e-01
rs73210711	4:5924451	-1.31	1.3e-03	2.5e-01
rs13144685	4:6223885	-0.44	1.4e-03	2.6e-01
rs4689521	4:6697176	-0.39	1.5e-03	2.7e-01

Supplementary Figure 8: Islet Gene View image of close homolog of L1 (CHL1), a gene with previous associations with diabetes-related traits such as HbA1c and insulin secretion.



RS	Position	Beta	P	FDR
rs2063886	3:222904	-0.48	5.1e-05	2.6e-02
rs4685395	3:88011	0.44	1.0e-04	4.5e-02
rs9545270	3:88643	0.44	1.0e-04	4.5e-02
rs191362	3:329445	0.35	2.0e-04	7.4e-02
rs332479	3:340440	0.35	2.3e-04	8.2e-02
rs332481	3:339058	0.34	2.6e-04	9.0e-02
rs331894	3:336766	0.34	2.9e-04	9.7e-02
rs332450	3:331429	0.34	3.4e-04	1.1e-01
rs1453109019	3:335197	0.33	4.0e-04	1.2e-01
rs162731	3:333145	0.33	4.1e-04	1.2e-01

Supplementary Figure 9: Islet Gene View image of SLC2A2.



RS	Position	Beta	P	FDR
rs12631813	3:171126134	0.36	2.0e-04	7.5e-02
rs11716511	3:171107629	0.36	5.4e-04	1.5e-01
rs11916954	3:171119802	0.33	5.4e-04	1.5e-01
rs6785069	3:171118709	0.33	5.4e-04	1.5e-01
rs13100463	3:171120159	0.33	5.5e-04	1.5e-01
rs10936683	3:171116657	0.33	5.5e-04	1.5e-01
rs13097586	3:171116637	0.33	5.5e-04	1.5e-01
rs9290412	3:171108599	0.34	5.5e-04	1.5e-01
rs7621422	3:171122620	0.33	5.6e-04	1.5e-01
rs9290411	3:171108171	0.34	5.6e-04	1.5e-01

Supplementary table 1: Clinical characteristics of the organ donors.

Variable	Number(data available)	Mean	SD	Min	Max
Age	187	59	10	24	81
BMI	188	26.4	3.8	17.6	40.1
HbA1c	169	5.9	0.7	4.3	10
Gender female/male	188	70/118			
Non-T2D / T2D	188	155/33			

Supplementary Table 2: Number of nominally and genome-wide significant genes ($p \leq 0.05$) associated with each diabetes-related phenotype. Values are expressed as percentages of the total list of 14 008 tested genes. Values in parentheses represent the absolute number of genes.

	Upregulated		**Downregulated**	
	Nominal	**Genome-wide**	**Nominal**	**Genome-wide**
$T2D_{diagnosed}$	7.82%(1175)	0.71%(108)	7.46%(1120)	0.65%(98)
HbA1c	2.49%(374)	0.05%(8)	2.62%(393)	0.09%(13)
Stimulatory index	2.66%(400)	0%(0)	4.89%(735)	0%(0)
BMI	6.21%(932)	0%(0)	4.42%(664)	0%(0)
HbA1c strata – IGT vs NGT	7.6%(1141)	0%(0)	4.32%(648)	0%(0)
Hba1c strata – $T2D_{HbA1c}$ vs NGT	5.83%(876)	0.15%(22)	5.66%(850)	0.22%(33)
Purity	27.38%(4111)	25.18%(3782)	29.81%(4476)	27.44%(4120)

Supplementary Table 3: Genes with genome-wide significant (FDR<=0.05) association with the >6.5% versus the <6% HbA1c strata.

Symbol	Ensembl	log fold-change	P-value	FDR
SERPINE2	ENSG00000135919	0.645	8.06E-08	3.03E-04
APOLD1	ENSG00000178878	0.936	2.18E-06	3.79E-03
SFRP1	ENSG00000104332	1.197	4.01E-06	5.47E-03
CALCA	ENSG00000110680	0.984	9.39E-06	1.08E-02
PCDHB15	ENSG00000113248	0.478	1.37E-05	1.27E-02
CLMP	ENSG00000166250	0.953	1.60E-05	1.27E-02
PAICS	ENSG00000128050	0.299	2.38E-05	1.67E-02
AGPAT5	ENSG00000155189	0.426	2.57E-05	1.67E-02
KCNAB2	ENSG00000069424	0.569	3.46E-05	1.93E-02
TMEM158	ENSG00000249992	0.912	5.76E-05	2.62E-02
APOD	ENSG00000189058	1.426	6.28E-05	2.71E-02
RASD2	ENSG00000100302	1.103	7.56E-05	3.07E-02
COLEC12	ENSG00000158270	0.702	7.98E-05	3.15E-02
FAIM2	ENSG00000135472	0.802	9.97E-05	3.61E-02
PDE7B	ENSG00000171408	0.511	1.04E-04	3.61E-02
REG4	ENSG00000134193	0.803	1.04E-04	3.61E-02
DKK3	ENSG00000050165	0.789	1.06E-04	3.61E-02
IL1RL1	ENSG00000115602	1.524	1.09E-04	3.61E-02
ARL4C	ENSG00000188042	0.437	1.10E-04	3.61E-02
MMP10	ENSG00000166670	1.252	1.12E-04	3.61E-02
	ENSG00000254615	0.59	1.13E-04	3.61E-02
VASH2	ENSG00000143494	0.667	1.14E-04	3.61E-02
PLA1A	ENSG00000144837	-1.051	7.27E-09	9.47E-05
SLC2A2	ENSG00000163581	-1.373	1.26E-08	9.47E-05
GLRA1	ENSG00000145888	-1.533	2.76E-08	1.38E-04
RASGRP1	ENSG00000172575	-0.711	6.99E-07	2.10E-03
CHL1	ENSG00000134121	-1.164	1.14E-06	2.77E-03
FFAR4	ENSG00000186188	-0.949	1.29E-06	2.77E-03
ERICH5	ENSG00000177459	-0.467	2.27E-06	3.79E-03
HHATL	ENSG00000010282	-1.099	2.55E-06	3.83E-03
C4orf19	ENSG00000154274	-0.488	5.64E-06	7.06E-03
PPP1R1A	ENSG00000135447	-0.848	1.05E-05	1.13E-02
CAPN13	ENSG00000162949	-0.829	1.42E-05	1.27E-02
DHFR	ENSG00000228716	-0.604	1.53E-05	1.27E-02
AASS	ENSG00000008311	-0.445	1.61E-05	1.27E-02
CTSV	ENSG00000136943	-0.648	1.75E-05	1.32E-02
PPFIBP2	ENSG00000166387	-0.52	2.47E-05	1.67E-02
ARG2	ENSG00000081181	-0.662	2.67E-05	1.67E-02
GRAMD3	ENSG00000155324	-0.288	3.11E-05	1.87E-02
UNC5D	ENSG00000156687	-0.879	3.38E-05	1.93E-02

FXYD2	ENSG00000137731	-0.815	3.67E-05	1.97E-02
GLP1R	ENSG00000112164	-0.791	4.83E-05	2.50E-02
PTPN3	ENSG00000070159	-0.386	5.13E-05	2.57E-02
ADH1C	ENSG00000248144	-1.626	5.52E-05	2.62E-02
ORC6	ENSG00000091651	-0.534	5.74E-05	2.62E-02
PARK2	ENSG00000185345	-0.443	6.31E-05	2.71E-02
PBLD	ENSG00000108187	-0.594	7.04E-05	2.94E-02
ARHGEF26	ENSG00000114790	-0.377	1.16E-04	3.61E-02
SOX6	ENSG00000110693	-0.41	1.18E-04	3.61E-02
FRK	ENSG00000111816	-0.368	1.35E-04	4.06E-02
TMEM37	ENSG00000171227	-0.474	1.55E-04	4.55E-02
GREM2	ENSG00000180875	-1.043	1.70E-04	4.91E-02
F11	ENSG00000088926	-0.684	1.80E-04	4.96E-02
	ENSG00000254226	-1.258	1.80E-04	4.96E-02
TMEM27	ENSG00000147003	-0.477	1.82E-04	4.96E-02

Supplementary Table 4: Genes with genome-wide (FDR <= 0.05) correlation with HbA1c.

Symbol	Ensembl	Beta	P-value	FDR
SERPINE2	ENSG00000135919	0.290	9.49E-07	2.37E-03
CPXM2	ENSG00000121898	0.537	1.75E-06	3.76E-03
SFRP1	ENSG00000104332	0.577	5.53E-06	9.22E-03
VASH2	ENSG00000143494	0.364	1.49E-05	1.88E-02
BHMT2	ENSG00000132840	0.284	1.63E-05	1.88E-02
FAIM2	ENSG00000135472	0.408	5.00E-05	4.34E-02
COLEC12	ENSG00000158270	0.350	5.88E-05	4.65E-02
CALCA	ENSG00000110680	0.436	6.53E-05	4.80E-02
SLC2A2	ENSG00000163581	-0.690	4.36E-09	6.55E-05
HHATL	ENSG00000010282	-0.610	6.90E-08	4.58E-04
PLA1A	ENSG00000144837	-0.480	9.15E-08	4.58E-04
GLRA1	ENSG00000145888	-0.709	1.79E-07	6.73E-04
PPP1R1A	ENSG00000135447	-0.476	3.49E-07	1.05E-03
CHL1	ENSG00000134121	-0.555	2.53E-06	4.74E-03
FFAR4	ENSG00000186188	-0.421	1.28E-05	1.88E-02
RASGRP1	ENSG00000172575	-0.305	1.55E-05	1.88E-02
UNC5D	ENSG00000156687	-0.438	2.48E-05	2.66E-02
CAPN13	ENSG00000162949	-0.390	3.07E-05	3.08E-02
GRAMD3	ENSG00000155324	-0.138	4.99E-05	4.34E-02
GLP1R	ENSG00000112164	-0.387	5.20E-05	4.34E-02
ERICH5	ENSG00000177459	-0.195	6.71E-05	4.80E-02

Supplementary Table 5: The number of genes positively or negatively co-expressed with 6 different genes encoding secreted pancreatic hormones.

	Number of correlated genes P ≤ 0.05	
Gene	**Positive**	**Negative**
INS	581	632
GCG	343	298
SST	106	92
PPY	1	0
IAPP	50	12
GHRL	0	0

Supplementary Table 6: Top negatively and positively correlated genes for secretory genes. Genes with nominal significant associations (p ≤ 0.05) are marked with bold text.

Secretory gene	Top correlated genes	Spearman's rho	P
INS	***IRF2BP1***	0.846	0.002
	GPBP1L1	-0.719	0.005
GCG	***TUSC3***	0.791	0.003
	ABTB2	-0.687	0.008
SST	***RBP4***	0.702	0.009
	WWC2	-0.621	0.016
PPY	***CARTPT***	0.636	0.018
IAPP	***SLCO1A2***	0.719	0.007
	FBXO32	-0.554	0.031
GHRL	*CNIH2*	0.438	0.107
	OSMR	-0.357	0.158

Supplementary Table 7: Top 10 genes positively and negatively associated with each studied phenotype, sorted by P-value.

Phenotype	Ensembl	Symbol	logFC	t	P-Value	FDR	B
T2D Positive	ENSG00000135919	SERPINE2	0.59	6.13	6.59E-09	3.30E-05	9.84
	ENSG00000211448	DIO2	1.58	5.14	7.96E-07	1.09E-03	5.50
	ENSG00000101463	SYNDIG1	0.81	5.05	1.17E-06	1.46E-03	5.15
	ENSG00000069424	KCNAB2	0.57	5.01	1.42E-06	1.65E-03	4.97
	ENSG00000100302	RASD2	1.12	4.89	2.45E-06	2.16E-03	4.48
	ENSG00000166250	CLMP	0.88	4.85	2.91E-06	2.43E-03	4.33
	ENSG00000104332	SFRP1	1.01	4.83	3.15E-06	2.49E-03	4.25
	ENSG00000188783	PRELP	1.20	4.80	3.58E-06	2.69E-03	4.14
	ENSG00000143494	VASH2	0.67	4.69	5.76E-06	3.82E-03	3.71
	ENSG00000135472	FAIM2	0.80	4.68	6.11E-06	3.82E-03	3.66
T2D Negative	ENSG00000144837	PLA1A	-1.00	-6.97	7.67E-11	1.15E-06	13.89
	ENSG00000145888	GLRA1	-1.48	-6.73	2.82E-10	2.12E-06	12.71
	ENSG00000010282	HHATL	-1.12	-5.96	1.49E-08	5.54E-05	9.10
	ENSG00000081181	ARG2	-0.75	-5.90	2.11E-08	5.54E-05	8.79
	ENSG00000172575	RASGRP1	-0.68	-5.89	2.21E-08	5.54E-05	8.74
	ENSG00000108984	MAP2K6	-0.50	-5.81	3.24E-08	6.95E-05	8.40
	ENSG00000163581	SLC2A2	-1.12	-5.65	7.09E-08	1.33E-04	7.68
	ENSG00000155324	GRAMD3	-0.32	-5.61	8.44E-08	1.41E-04	7.53
	ENSG00000134121	CHL1	-1.02	-5.16	7.27E-07	1.09E-03	5.58
	ENSG00000186188	FFAR4	-0.82	-4.99	1.56E-06	1.68E-03	4.89
HbA1c Positive	ENSG00000135919	SERPINE2	0.29	5.09	9.49E-07	2.37E-03	4.94
	ENSG00000121898	CPXM2	0.54	4.96	1.75E-06	3.76E-03	4.42
	ENSG00000104332	SFRP1	0.58	4.70	5.53E-06	9.22E-03	3.45
	ENSG00000143494	VASH2	0.36	4.46	1.49E-05	1.88E-02	2.63
	ENSG00000132840	BHMT2	0.28	4.44	1.63E-05	1.88E-02	2.55
	ENSG00000135472	FAIM2	0.41	4.16	5.00E-05	4.34E-02	1.61
	ENSG00000158270	COLEC12	0.35	4.12	5.88E-05	4.65E-02	1.48
	ENSG00000110680	CALCA	0.44	4.10	6.53E-05	4.80E-02	1.39
	ENSG00000050165	DKK3	0.39	4.03	8.64E-05	5.71E-02	1.16
	ENSG00000262655	SPON1	0.41	3.97	1.06E-04	6.36E-02	0.99
HbA1c Negative	ENSG00000163581	SLC2A2	-0.69	-6.20	4.36E-09	6.55E-05	9.47
	ENSG00000010282	HHATL	-0.61	-5.65	6.90E-08	4.58E-04	7.14
	ENSG00000144837	PLA1A	-0.48	-5.59	9.15E-08	4.58E-04	6.90
	ENSG00000145888	GLRA1	-0.71	-5.45	1.79E-07	6.73E-04	6.34
	ENSG00000135447	PPP1R1A	-0.48	-5.31	3.49E-07	1.05E-03	5.78
	ENSG00000134121	CHL1	-0.55	-4.88	2.53E-06	4.74E-03	4.11
	ENSG00000186188	FFAR4	-0.42	-4.50	1.28E-05	1.88E-02	2.75
	ENSG00000172575	RASGRP1	-0.31	-4.45	1.55E-05	1.88E-02	2.59

	ENSG00000156687	UNC5D	-0.44	-4.34	2.48E-05	2.66E-02	2.20
	ENSG00000162949	CAPN13	-0.39	-4.29	3.07E-05	3.08E-02	2.02
Stimulatory.index	ENSG00000145888	GLRA1	0.06	4.47	1.41E-05	2.11E-01	1.91
Positive	ENSG00000135447	PPP1R1A	0.04	3.73	2.56E-04	6.48E-01	-0.85
	ENSG00000039139	DNAH5	0.03	3.40	8.43E-04	6.48E-01	-1.97
	ENSG00000137331	IER3	0.06	3.35	9.77E-04	6.48E-01	-2.11
	ENSG00000170145	SIK2	0.01	3.33	1.06E-03	6.48E-01	-2.19
	ENSG00000151612	ZNF827	0.01	3.22	1.50E-03	6.48E-01	-2.51
	ENSG00000148143	ZNF462	0.02	3.17	1.78E-03	6.48E-01	-2.67
	ENSG00000054793	ATP9A	0.02	3.17	1.82E-03	6.48E-01	-2.69
	ENSG00000156564	LRFN2	0.03	3.15	1.93E-03	6.48E-01	-2.74
	ENSG00000130477	UNC13A	0.02	3.13	2.02E-03	6.48E-01	-2.78
Stimulatory.index	ENSG00000011638	TMEM159	-0.02	-3.42	7.84E-04	6.48E-01	-1.91
Negative	ENSG00000189043	NDUFA4	-0.02	-3.38	8.90E-04	6.48E-01	-2.02
	ENSG00000180098	TRNAU1AP	-0.01	-3.38	8.93E-04	6.48E-01	-2.03
	ENSG00000100823	APEX1	-0.01	-3.37	9.26E-04	6.48E-01	-2.06
	ENSG00000118680	MYL12B	-0.02	-3.35	9.77E-04	6.48E-01	-2.11
	ENSG00000080823	MOK	-0.02	-3.33	1.04E-03	6.48E-01	-2.17
	ENSG00000197360	ZNF98	-0.02	-3.32	1.08E-03	6.48E-01	-2.21
	ENSG00000165629	ATP5C1	-0.01	-3.29	1.22E-03	6.48E-01	-2.31
	ENSG00000123130	ACOT9	-0.02	-3.27	1.28E-03	6.48E-01	-2.36
	ENSG00000240541	TM4SF1-AS1	-0.03	-3.22	1.50E-03	6.48E-01	-2.51
BMI	ENSG00000115594	IL1R1	0.04	4.71	4.76E-06	7.10E-02	3.42
Positive	ENSG00000101938	CHRDL1	0.11	4.56	9.46E-06	7.10E-02	2.76
	ENSG00000228649	NA	0.08	4.43	1.61E-05	8.08E-02	2.25
	ENSG00000169756	LIMS1	0.03	4.24	3.59E-05	8.53E-02	1.49
	ENSG00000104332	SFRP1	0.09	4.22	3.85E-05	8.53E-02	1.42
	ENSG00000182326	C1S	0.06	4.21	3.98E-05	8.53E-02	1.39
	ENSG00000262655	SPON1	0.07	4.09	6.51E-05	9.10E-02	0.92
	ENSG00000169554	ZEB2	0.06	4.08	6.57E-05	9.10E-02	0.91
	ENSG00000158966	CACHD1	0.03	4.05	7.55E-05	9.10E-02	0.78
	ENSG00000169826	CSGALNACT2	0.03	4.04	7.81E-05	9.10E-02	0.75
BMI	ENSG00000162630	B3GALT2	-0.06	-4.21	3.93E-05	8.53E-02	1.40
Negative	ENSG00000172575	RASGRP1	-0.05	-4.12	5.72E-05	9.10E-02	1.05
	ENSG00000237289	CKMT1B	-0.03	-3.70	2.81E-04	1.28E-01	-0.45
	ENSG00000126733	DACH2	-0.04	-3.70	2.81E-04	1.28E-01	-0.45
	ENSG00000134121	CHL1	-0.07	-3.65	3.41E-04	1.28E-01	-0.64
	ENSG00000172264	MACROD2	-0.04	-3.64	3.54E-04	1.28E-01	-0.67
	ENSG00000138622	HCN4	-0.04	-3.53	5.17E-04	1.32E-01	-1.02
	ENSG00000168781	PPIP5K1	-0.03	-3.53	5.33E-04	1.32E-01	-1.05
	ENSG00000174482	LINGO2	-0.05	-3.51	5.60E-04	1.32E-01	-1.10
	ENSG00000148584	A1CF	-0.06	-3.48	6.34E-04	1.42E-01	-1.21