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[1,8], Leif Groop[1,8a,b*], Rashmi B Prasad[1a,b*] on behalf of the LUDC consortium.

*a These authors contributed equally
*b Senior authors

[1] Diabetes and Endocrinology Unit, 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] 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: 137
Main text: 3181
Number of figures: 3
Number of tables: 4
Number of supplementary figures: 0
Number of supplementary tables: 1

Abstract

The comprehensive analyses of databases and biobanks, using advanced bioinformatics tools, have made a tremendous amount of data available to the scientific community. These advances prompt the development of tools that allow quick and robust overview of data, particularly for non-specialists. Therefore, we have developed Islet Gene View (IGW), a tool that aims to make information on gene expression in human pancreatic islets from organ donors accessible to the scientific community. Islet Gene View currently consists of information on islets from 188 donors where islet RNA expression of more than 15 000 genes can be related to phenotypic information (gender, age, BMI, HbA1c, insulin and glucagon secretion etc.) and expression of other genes as well as regulation by genetic variation (GWAS). The data can be accessed by the easy-to-use Islet Gene View web application. This tool will soon be available online after a simple sign-up process.

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 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 pancreas obtained from organ donors for transplantation purposes where blood flow still has been kept intact or from partial pancreatectomies of pancreatic tumors.

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 (islets, fat, liver and muscle) from deceased human donors. RNA sequencing and Genome Wide Association Studies (GWAS) have been performed to allow analysis of effects of genetic variation on gene expression, *i.e.* eQTLs (Expression Quantitative Traits). Also, individual information such as gender, BMI, and age are provided together with information on 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. It will now 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 Table 1.

Islet Phenotypes

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

T2D diagnosis: Type 2 diabetes diagnosis was defined in two ways; one was based on clinical diagnosis (T2D$_{diagnosed}$, N = 33) and the other, based on the patient exhibiting an HbA1c above 6.5% (NGSP; = 48 mmol/mol; IFCC) (T2D$_{HbA1c}$, 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 was used as a measure of glucose-stimulated insulin secretion. To calculate this, the islets were subjected to static perifusion 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 states 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 and 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 calculated 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.

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.

Results

Features

Islet Gene View (IGW) is accessible at http://www.example.com after registration. 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 type 2 diabetes ($T2D_{diagnosed}$), 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).

The example of the *SERPINE2* gene

The Serpin Family E Member 2 (*SERPINE2*) gene was selected to illustrate the information that is provided by the Islet Gene View tool. *SERPINE2* shows highest expression in islets, but is also expressed in adipose tissue, liver and muscle (figure 1A). *SERPINE2* was modestly albeit

significantly positively correlated with purity ($r2 = 0.064$, $p = 0.011$, figure 1B). The rank (47% for *SERPINE2*) relates to the rank of the P-value for the correlation between expression of the *SERPINE2* gene with purity as compared to all other genes within the islet samples; this means that expression of *SERPINE2* is strongly influenced by purity, i.e. among the top 47%. For comparison, the Collapsin Response Mediator Protein 1 (*CRMP1*) gene was the top-ranked gene (0.013%) positively associated with purity, suggesting that it is almost exclusively expressed in the endocrine part of the pancreas. In contrast, the BAI1-associated protein 2 like 1 (*BAIAP2L1*) (0.0067%) gene showed the most significant negative correlation with purity, suggesting that it is enriched in the exocrine part of the pancreas.

The average expression of the *SERPINE2* in the islets in relation to all other genes, measured in Fragments Per Kilobase of transcript per Million mapped reads (FPKM) was in the top 11% of all genes expressed. For comparison, *SERPINE2* had a higher average expression than *e.g.* the Potassium Voltage-gated Channel Subfamily J Member 11 gene (*KCNJ11*) and was among the top 11% of genes with the highest average expression. The most highly expressed gene in the dataset was the glucagon (*GCG*) gene, followed by the Regenerating Family Member 1 α (*REG1A)* gene and Mitochondrially Encoded Cytochrome C Oxidase I (*MT-CO1)* gene.

Association between gene expression and diabetes-related phenotypes

SERPINE2 was the most significantly up-regulated gene in T2D patients compared to non-diabetic organ donors ($P_{FDR} = 1.1 \cdot 10^{-5}$, $\log_2$ fold change = 0.58 or ~150% increase, Figure 1D). In contrast, the most downregulated gene in T2D donors was the Phospholipase A1 Member A (*PLA1A*) gene ($\log_2$ fold change = -1.00, $P_{FDR} = 5.9 \cdot 10^{-5}$).

In support of the association with T2D, the *SERPINE2* gene was upregulated in patients with high HbA1c whereas *PLA1A* was the most strongly down-regulated gene in donors with the highest HbA1c (Figure 1E).

Expression of *SERPINE2* showed a significant positive association with HbA1c (rank = 0.04%, $r2 = 0.13$, $r= 9.5 \times 10^{-7}$). In fact, of all expressed genes, *SERPINE2* showed the strongest positive association with HbA1c (Figure 1F). In contrast, the Solute Carrier Family 2 Member 2 (*SLC2A2*) gene, encoding the glucose transporter 2 (GLUT2), showed the strongest negative association with HbA1c. In support of this *SLC2A2* expression was also lower in islets from T2D donors ($P_{FDR}= 3.2 \cdot 10^{-4}$, $\log_2$ fold change = -1.08). *SERPINE2* was nominally correlated with BMI with a rank of 0.68 % compared with all genes (Figure 1G). The gene showing the strongest positive association with BMI was the interleukin receptor 1 type 1 (*IL1R1*) gene whereas the β-1,3-Galactosyltransferase 2 (*B3GALT2*) gene showed the strongest negative correlation with BMI. However, none of these associations reached genome-wide significance suggesting that BMI does not have a strong influence on islet gene expression.

The expression of *SERPINE2* was not significantly associated with insulin secretion, in the form of stimulatory index (SI; Figure 1H). The gene showing the strongest positive association with stimulatory index was Glycine Receptor α 1 (*GLRA1*) whereas the Transmembrane Protein 159 (*TMEM159*) gene, also known as promethin, showed the strongest negative association with SI. However, none of these associations reached genome-wide significance.

Among the expressed genes, 350 were significantly differentially expressed between islets from non-diabetic donors compared to T2D donors with an established clinical diagnosis but only 55 when diagnosis of T2D were based upon HbA1c. The expression of 21 genes correlated with HbA1c, and 7902 genes were impacted by purity (Table 2). Top-ranking genes associated with islet-related phenotypes are shown in Supplementary Table 1.

Co-expression of genes with expression of islet secretory genes

The expression of the gene of interest is compared with 5 genes encoding proteins known to be secreted from distinct cell types within islets. These included 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). *IRF2BP1* showed the highest co-expression with *INS*, whereas *TUSC3*, *RBP4*, *CARTPT*, *SLCO1A2* and *CNIH2* showed the highest co-expression with *GCG*, *SST*, *PPY*, *IAPP* and *GHRL* respectively (Table 4). The expression of the *INS* gene is correlated with the largest number of genes (Table 3), followed by *GCG*, *SST*, *IAPP* and *PPY*. For *INS*, the most strongly positively correlated gene was Interferon Regulatory Factor 2 Binding Protein 1 (*IRF2BP1)(r = 0.85, p = 0.002)* whereas the most strongly negatively correlated gene was GC-Rich Promoter Binding Protein 1 Like 1 (*GPBP1L1).* For glucagon (*GCG),* the most strongly positively correlated gene was Tumor Suppressor Candidate 3 (*TUSC3)* and the most strongly negatively correlated gene 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. *RBP4* is negatively associated with HbA1c (P = 0.0085) and BMI (P = 0.023) but positively correlated with *INS* expression (P = 0.048). The gene showing the strongest negative correlation with SST was *WWC2*. *WWC2* was negatively associated with purity (P = $1.8 \cdot 10^{-15}$) and was negatively correlated with expression of *INS* (P = 0.021), suggesting that it might also have an exocrine origin. *SERPINE2* was not significantly co-expressed with any of the pancreatic hormone genes.

We also explored how many genes expressed in the islets show positive co-expression with secretory genes (Figure 3A), *i.e* INS (581 genes), GCG (343), SST (98), and IAPP (50). Similarly we explored the number of genes, which were negatively co-expressed with each secretory gene (Figure 3B), *i.e.* INS (632 genes), GCG (298), SST (92) and IAPP (12). 112 genes are either positively or negatively co-expressed in relation to two or more secretory genes (Figure 3C) and

six genes were co-expressed but in opposite directions in relation to different secretory genes (Figure 3D). The six genes correlated with both *INS* and *GCG,* however, in opposite direction, *e.g.* *WIZ*, *NUDC*, *IMP4*, *ZNF787*, *KLHL21*, and *ISG20*.

An example of a locus associated with T2D – *CHL1*

Expression of the *CHL1* gene was shown to be downregulated in islets from patients with T2D in a previous smaller study from our laboratory (Fadista et al., 2014). This was consistent in Islet Gene View with larger number of samples (Figure 2). Moreover the fasting insulin-associated SNP rs9841287 at the *CHL1* locus was found to be an expression quantitative trait locus (eQTL) for the *CHL1* and CHL1-AS1 genes (P = 0.028, increasing allele = G). Additionally, it was negatively associated with high HbA1c.

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 data included in our expanded dataset (Fadista et al., 2014; Taneera et al., 2012), the current data set includes 188 islets from human organ donors, thereby representing the largest in its kind to date. Compared to our previous report (Fadista et al., 2014), several refinements to the analytical pipeline have been made. We have introduced batch correction with 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.

An important aspect of Islet Gene View is that it can present comparisons between 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 high in RNA-Seq compared to gene expression microarrays, gene sequence might influence the sequencing rate of different transcripts. Information on purity provides information of expression in the endocrine vs the exocrine components of the pancreas. Islet Gene View also provides information on differential expression between T2D and non-diabetic organ donors. Here we used the *SERPINE2* gene as an example, as it was the gene showing the strongest differential expression between islets from T2D and normoglycemic individuals.

The stimulatory index did not show genome-wide significant associations with expression of any gene. There might be several explanations for this. First, although being the largest dataset of its kind, it still has limited power to provide genome-wide significance. Stimulatory index is the result of experimental intervention. Thus, the variability of repeated measurements is likely to be large in comparison to patient-oriented phenotypes such as BMI or type 2 diabetes status.

Islets are comprised of many cell types, *e.g.* α-cells, β-cells, δ-cells, ε-cells and PP-cells. Islet

Gene View provides information on co-expression of genes with the expression of the hormone mainly secreted by these cell types. Among the many genes co-expressed with *INS*, *IRF2BP1*, a co-repressor of Interferon Regulatory Factor 2 (*IRF2*) (Mashima et al., 2011), showed the strongest positive correlation. *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. This link between would provide an interesting candidate for functional follow-up.

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).*

Bidirectional expression of several genes with INS and GCG was observed (Figure 3). 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. 2016). 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 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 in a previous GWAS (P = 7.78 · 10^{-9}) (Manning et al., 2012; Scott et al., 2017). Many of the findings suggest high expression of CHL1 in islets could provide positive effects.

The IGW has its limitations. First, the power of the study to detect significant expression-phenotype associations is limited by its sample size. Second, being an observational study of patients recruited through intensive care units, the data in Islet Gene View shows correlations, but

10

cannot on its own be used to distinguish correlation from causation. However, Islet Gene View is a tool to facilitate research on human pancreatic islets to serve the scientific community.

Acknowledgements

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, LG and OH supervised the project. LC contributed the data. All authors provided critical feedback and helped shape the research, analysis and manuscript.

11

References

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.

Durinck, S., Spellman, P.T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. *4*, 1184–1191.

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.

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.

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. (2016) Wiz binds active promoters and CTCF-binding sites and is required for normal behaviour in the mouse. ELife *5*.

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.

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).

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.

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.

Tables

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			

Table 2: Number of nominally and genome-wide significant genes (p ≤ 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**
$T2D_{diagnosed}$	9.97%(1497)	1.31%(197)	8.78%(1318)	1.06%(159)
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)

14

Table 3: *The number of genes positively or negatively co-expressed with 6 different genes encoding secreted pancreatic hormones.*

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

Table 4: *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

15

Figures

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.
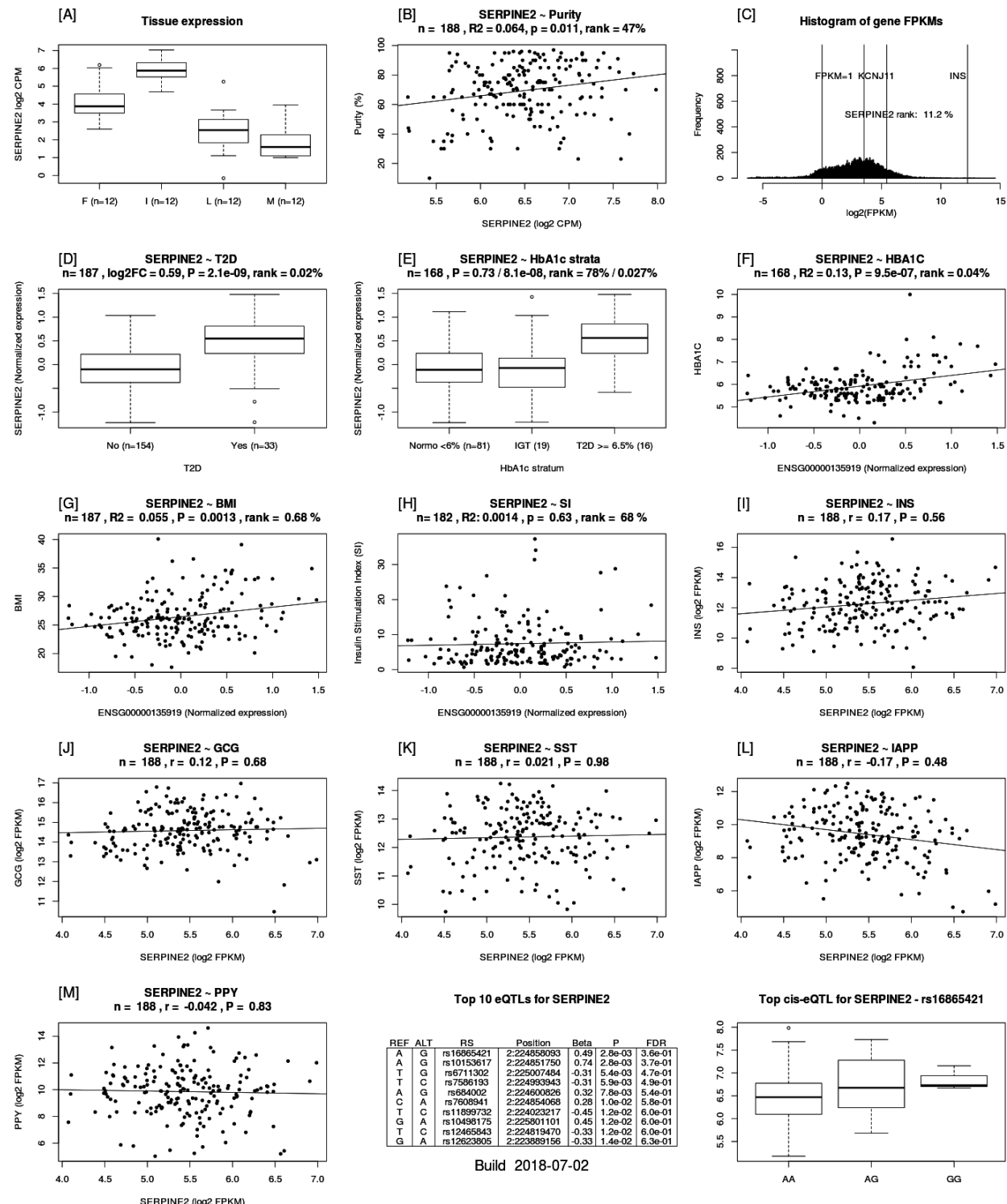
18

Figure 2: 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.