Multi-omics profiling of living human pancreatic islet donors reveals heterogeneous beta cell trajectories toward type 2 diabetes

Leonore Wigger[1*], Marko Barovic[2,3,4*], Andreas-David Brunner[5*], Flavia Marzetta[1], Eyke Schöniger[2,3,4], Florence Mehl[1], Nicole Kipke[2,3,4], Daniela Friedland[2,3,4], Frederic Burdet[1], Camille Kessler[1], Mathias Lesche[6], Bernard Thorens[7], Ezio Bonifacio[3,4,8], Cristina Legido-Quigley[9,10], Pierre Barbier Saint Hilaire[11], Philippe Delerive[12], Andreas Dahl[6], Christian Klose[13], Mathias J Gerl[13], Kai Simons[13], Daniela Aust[14,15], Jürgen Weitz[16], Marius Distler[16], Anke M Schulte[17], Matthias Mann[5#], Mark Ibberson[1#], Michele Solimena[2,3,4#]

[1]Vital-IT Group, SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland; [2]Department of Molecular Diabetology, University Hospital and Faculty of Medicine, TU Dresden; [3]Paul Langerhans Institute Dresden (PLID), Helmholtz Center Munich, University Hospital and Faculty of Medicine, TU Dresden; [4]German Center for Diabetes Research (DZD e.V.), Neuherberg, Germany; [5]Max Planck Institute of Biochemistry, Martinsried, Germany; [6]DRESDEN-*concept* Genome Center, c/o Center for Molecular and Cellular Bioengineering, Technische Universität Dresden, Germany; [7]Center for Integrative Genomics, University of Lausanne, Lausanne Switzerland; [8]Center for Regenerative Therapies Dresden, Faculty of Medicine and Center for Molecular and Cellular Bioengineering,Technische Universität Dresden, Dresden, Germany; [9]Steno Diabetes Center Copenhagen, Gentofte, Denmark; [10]King's College London, London, UK; 11DMPK Center, Technologie Servier, Orléans, France [12]Institut de Recherches Servier, Pôle d'Innovation Thérapeutique Métabolisme, Suresnes, France; [13]Lipotype GmbH, Dresden, Germany; [14]Department of Pathology, Medical Faculty, University Hospital Carl

Gustav Carus, Technische Universität Dresden, Dresden, Germany; [15]NCT Biobank Dresden, University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany; [16]Department of Visceral, Thoracic and Vascular Surgery, University Hospital Carl Gustav Carus, Medical Faculty, Technische Universität Dresden, Dresden, Germany; [17]Sanofi-Aventis Deutschland GmbH, Diabetes Research, Industriepark Höchst, 65926 Frankfurt am Main, Germany.

* These authors contributed equally

# These authors jointly supervised this work

Corresponding authors:

Matthias Mann - mmann@biochem.mpg.de

Mark Ibberson - Mark.Ibberson@sib.swiss

Michele Solimena - Michele.Solimena@uniklinikum-dresden.de

Abstract

Existing studies do not sufficiently describe the molecular changes of pancreatic islet beta cells leading to their deficient insulin secretion in type 2 diabetes (T2D). Here we address this deficiency with a comprehensive multi-omics analysis of metabolically profiled pancreatectomized living human donors stratified along the glycemic continuum from normoglycemia to T2D. Islet pools isolated from surgical samples by laser-capture microdissection had remarkably heterogeneous transcriptomic and proteomic profiles in diabetics, but not in non-diabetic controls. Transcriptomics analysis of this unique cohort revealed islet genes already differentially regulated in prediabetic individuals with impaired glucose tolerance. Our findings demonstrate a progressive but disharmonic remodeling of mature beta cells, challenging current hypotheses of linear trajectories toward precursor or trans-differentiation stages in T2D. Further, integration of islet transcriptomics and pre-operative blood plasma lipidomics data enabled us to define the relative importance of gene co-expression modules and lipids positively or negatively associated with HbA1c levels, pointing to potential prognostic markers.

Abstract

Most research on human pancreatic islets is conducted on samples obtained from normoglycemic or diseased brain dead donors and thus cannot accurately describe the molecular changes of pancreatic islet beta cells as they progress towards a state of deficient insulin secretion in type 2 diabetes (T2D). Here, we conduct a comprehensive multi-omics analysis of pancreatic islets obtained from metabolically profiled pancreatectomized living human donors stratified along the glycemic continuum, from normoglycemia to T2D. We find that islet pools isolated from surgical samples by laser-capture microdissection display remarkably more heterogeneous transcriptomic and proteomic profiles in patients with diabetes than in non-diabetic controls. The differential regulation of islet gene expression is already observed in prediabetic individuals with impaired glucose tolerance. Our findings demonstrate a progressive, but disharmonic, remodeling of mature beta cells, challenging current hypotheses of linear trajectories toward precursor or trans-differentiation stages in T2D. Furthermore, through integration of islet transcriptomics with pre-operative blood plasma lipidomics, we define the relative importance of gene co-expression modules and lipids that are positively or negatively associated with HbA1c levels, pointing to potential prognostic markers.

Introduction

Type 2 diabetes (T2D) mellitus defines a cluster of genetically complex pathological states characterized by persistent hyperglycemia, often leading to cardiovascular complications, kidney failure, retinopathy and neuropathies. Affecting more than 450 million people, with rising incidence rates over the past decades, this syndrome is a major threat for public health and society globally[1]. Common determinant and ultimate cause of T2D is the inability of pancreatic islet beta cells to secrete insulin in adequate amounts relative to insulin sensitivity, in the absence of evidence for their autoimmune destruction or a monogenetic deficit. Beta cell failure typically results from a lengthy process spanning many years. Remarkably, however, it can be rapidly reverted upon bariatric surgery or severe caloric restriction[2,3]. These observations argue against the occurrence of major beta cell apoptosis in T2D, especially since adult beta cells hardly replicate, while robust evidence of beta cell neogenesis after puberty is also lacking. Hence, the prevailing opinion is that persistent metabolic stress drives mature beta cells to phenotypically de-differentiate into progenitor cells or trans-differentiate into other islet endocrine cell types over time[4–6]. As the pathogenesis of beta cell dysfunction in T2D remains largely unclear, the diagnosis of this disease relies on accepted, surrogate parameters and cutoffs that have been primarily developed for clinical practice to optimize therapeutic interventions[7].

Insight into molecular alterations associated with impaired insulin secretion in T2D has been largely obtained from pancreatic islets isolated enzymatically from brain-dead or cadaveric subjects classified according to a categorical division into non-diabetic and diabetic, rather than on a continuum from euglycemia to steady hyperglycemia. This approach has multiple

shortcomings[8]. Briefly, islet researchers do not generally have access to extensive clinical and laboratory information about the donors prior to their admission to an intensive therapy unit[9]. Moreover, the islet state is perturbed by the metabolic stress associated with a terminal condition and the related pharmacological treatments[10,11]. Enzymatic isolation of islets and their in vitro culture can further change their molecular profile[12,13]. In the attempt to overcome, at least in part, these limitations, we established a complementary platform for the procurement of islets which relies on the collection and analysis of pancreatic specimens from metabolically profiled living donors undergoing pancreatectomy for a variety of disorders[8,14]. We showed that this approach is very reproducible and scalable and provides a novel view on transcriptomic and functional alterations in pancreatic islets of subjects with T2D[15–17]

The aim of the present study has been to profile in greater detail gene expression changes occurring along the progression from euglycemia to long-standing T2D in human islets *in situ* and to integrate this knowledge with clinical traits, circulating lipid levels and the islet proteome, hence enabling inferences about the mechanisms driving islet dysfunction and the identification of potential biomarkers for it.

Results

Living donors enable islet studies along progression to T2D

To gain insight into the history of islet cell deterioration along the progression from normal glycemic regulation to T2D, we collected surgical pancreatic tissue samples from 133 metabolically phenotyped pancreatectomized patients (PPP). Eighteen were non-diabetic (ND), 41 had impaired glucose tolerance (IGT), 35 Type 3c Diabetes (T3cD) and 39 T2D (Fig. 1A and Fig. 1B). These group assignments were based on glycemic values at fasting and at the 2 h time point of an oral glucose tolerance test (OGTT) using the thresholds defined in the

guidelines of the American Diabetes Association[7], or, when applicable, on a previously established diagnosis of T2D. In this cohort, 51.9% were males and the mean age was 65.36±11.54 years, with ND PPP being on average younger than the other three groups (Fig. 1C and Supplementary Table 1). The body mass index (BMI) was significantly lower in ND compared to IGT, T3cD and T2D PPP. The HbA1c value, as a parameter of longer-term glycemia, was 5.25±0.3 in ND, 5.75±0.42 in IGT, 6.29±0.95 in T3cD and 7.41±1.29 in T2D PPP (Fig. 1C and Supplementary Table 1). Moreover, based on histopathology, malignant tumors occurred in 50%, 60.97%, 74.29% and 69.23% of ND, IGT, T3cD, and T2D PPP, respectively (Supplementary Table 1).

Islet gene expression drifts with glycemic deterioration

Gene expression profiles of islets isolated by laser capture microdissection (LCM) from resected and snap-frozen pancreas samples of ND, IGT, T3cD and T2D PPP were assessed by RNA sequencing. After removal of genes with low expression levels, the overall islet transcriptome encompassed 19,119 genes, of which 14,699±693 were present (raw read counts >0) in ND PPP, 14,967±455 in IGT PPP, 14,939±493 in T3cD PPP and 14,997±428 in T2D PPP. Genes with a fold change (FC)>1.5 and a false discovery rate (FDR)≤0.05 were considered to be differentially expressed (DE) between the groups. Multiple group comparison by linear modeling was performed (Supplementary Table 2). Subsequent pairwise group comparisons of IGT vs. ND, T3cD vs. ND and T2D vs. ND revealed an exacerbation of gene dysregulation with deterioration of glycemic control (Fig. 2A). Notably, no DE islet genes were identified between IGT vs. ND PPP, while 161 and 650 DE genes were found between T3cD vs. ND PPP and T2D vs. ND PPP, respectively (Fig. 2A and Supplementary Table 2).

Restricting the transcriptomic analysis to libraries in which insulin (*INS*) was the most expressed gene resulted in the retention of islet datasets from 15 ND, 35 IGT, 21 T3cD and 24 T2D subjects, without substantially affecting the overall composition of the cohort in regard to diabetes status and major descriptive parameters (Supplementary Table 3). Deconvolution analysis indicated that in 78.3% of retained samples the proportion of beta cells was >50% (Extended data Fig. S1), supporting the choice of this strategy to discriminate samples especially enriched in beta cell transcripts. This analysis further pointed to the overall enrichment in beta cell content of the LCM isolated islets in comparison to another large study based on islets isolated by enzymatic digestion (median beta cell:non-beta endocrine cell ratio 3.98:1 and 1.4:1)[18]. This enrichment can conceivably be attributed to the selectivity of the LCM isolation procedure for beta cell rich areas due to their higher autofluorescence. Despite the expected reduction in statistical power due to ~ 30% smaller size of this "restricted" cohort (92 samples retained from 133), the number of significantly DE genes increased in the multiple group comparison, as well as in pairwise comparisons between islets of T2D vs. ND PPP by 51% to 984 (782 up, 202 down), and by 59% to 256 (209 up, 47 down) between islets of T3cD vs. ND PPP (Fig. 2A, Supplementary Table 4). Seven of the 984 DE genes are among the putative effectors of GWAS risk loci for T2D (https://t2d.hugeamp.org/), two upregulated (*SGSM2* and *BCL2*) and five downregulated (*RASGRP1, G6PC2, SLC2A2, ZMAT4* and *PLUT*)[19], while most of the remaining genes have not been previously reported to be altered in islets of subjects with T2D[14,20].

Among the DE genes in islets of T2D PPP, *INF2* and *AKR7L* were negatively correlated in a moderate fashion with duration of the disease measured in years (Spearman correlation coefficient -0.32 and -0.41 respectively), albeit they were both upregulated relative to islets of ND PPP. Most notably, this filtering step enabled, for the first time, the identification of 185 DE genes between islets of IGT vs. ND PPP. Most of these DE genes were upregulated (181/185),

and 98 also differentially regulated with the same directionality (97 up, 1 down) between islets of T2D vs. ND PPP. Intriguingly, and apparently at variance with previous findings[21], the proposed T2D risk genes *ARAP1* and its neighboring gene *STARD10* were both upregulated and among the 77 genes differentially regulated in islets of IGT PPP only. No islet cell type specific genes[22] were enriched in any of the differential expression analyses. Furthermore, no shift of islet cell type proportions with the progression of the disease was observed in the deconvolution analysis (Extended Data Fig. S1A).

Concerning samples with the highest transcript other than insulin, these were not noteworthily different from the other samples in any of the clinical parameters or anatomical part of the pancreas the tissue originated from. Nine of them had *PRSS1* (coding for trypsin) as the most enriched transcript, pointing to exocrine contamination and one was marked by *MALAT1* and was therefore excluded as suspect for cancerous cell contamination. The remaining samples were remarkable for expressing a non-beta-cell endocrine gene, specifically 13 samples with predominant alpha cell (*GCG* or *TTR*) and 18 samples with predominant gamma cell (*PPY*) characteristic genes. This is partially reflected by the results of the deconvolution analysis (Extended Data Figure S1). This specific group of 41 samples was not analyzed further since the number of subjects in each of the four glycemic groups was too small for statistical analysis.

For both the “restricted” and the full data set, heatmaps of gene expression levels in the four patient groups were prepared as a visual complement to the statistical analysis (Fig. 2B and Extended Data Fig. S2A). Despite the marked differences between the findings in the “restricted” and complete cohort, upregulation prevailed as the direction of gene dysregulation in both of them (Fig. 2A and Extended Data Fig. S2A). Based on these observations, pancreatic tissue sections of 5 ND and 5 T2D PPP with the “restricted” cohort were immunostained with antibodies specific for histone H3 and H4 lysine acetylation – an epigenetic modification

associated with greater access of transcription factors to promoter sites resulting in increased gene expression. Notably, qualitative assessment by immunostaining indicated a remarkable increase of the signals for acetylated H3 and H4 in the islets, and also in the surrounding exocrine cells of T2D PPP, compared to ND PPP (Fig. 2D).

Gene pathways are progressively perturbed from IGT to T2D

We further analyzed differentially expressed gene functions by gene set enrichment analysis using Gene Ontology terms and KEGG pathways (Fig. 2C, Extended Data Fig. S2B and Supplementary Tables 5 and 6). Results obtained from the different gene set collections cross-validated each other, since similar biological themes emerged. Islets of pre-diabetic and diabetic subjects displayed upregulation of islet genes that were functionally related to cell-extracellular matrix interaction, immune response and signaling pathways, while expression of genes related to RNA processing, protein translation and mitochondrial oxidative phosphorylation were downregulated. Importantly, the analysis performed on the “restricted” cohort, differently from the full dataset, also revealed that the strength of the enrichment increased with progression of the disease (Fig. 2C and Extended Data Fig. S2B). These data suggest that early dysregulation of gene pathways exacerbates with the decline of beta cell function.

WGCNA identifies islet gene modules correlated with HbA1c

To globally interpret transcriptomic data and identify sets of genes likely to be functionally related and co-regulated, we grouped genes based on similarities in their expression profiles into modules using a network-based approach[23]. In the cohort of 133 PPP, we identified 36 co-expressed gene modules, which were arbitrarily labeled M1 through M36. The expression profiles of the genes in each module were summarized by a module eigengene, or first principal component of the expression matrix. Module eigengenes were used to computationally relate

modules to one another and to genes or clinical variables. Correlation between module eigengenes and diabetes-related clinical traits revealed modules M9 and M14 as those with the highest positive and negative correlation with HbA1c, respectively (Fig. 3A and Supplementary Table 7). The former consisted of a set of genes that showed similar patterns of increased expression in most PPP with T2D (Fig. 3B), while the latter was mostly composed of genes with coordinated down-regulation in diseased subject samples (Fig. 3C).

We next correlated the expression values of each gene contained in a module to the eigengene of the module. The correlation coefficient from this calculation is denoted as the "module membership" of the gene and serves as a quantitative measure of how representative a gene is of the module it belongs to. Strong module memberships point to genes that are highly connected in the underlying gene-gene similarity network of the WGCNA. This analysis allowed us to identify highly connected genes or "hub" genes for HbA1c-related modules (Fig. 3D-E). These included genes that we had previously identified as differentially expressed in subjects with T2D, and which were correlated with HbA1c either positively, such as module M9 genes *ALDOB* (FC=8.45 with adj. $p<0.001$ in T2D vs. ND in the "restricted" cohort) and *FAIM2* (FC=7.11 with adj. $p<0.001$ in T2D vs. ND in the "restricted" cohort) or negatively, such as module M14 genes *SLC2A2* (FC=-2.77 with adj. $p<0.001$ in T2D vs. ND in the "restricted" cohort) and *TMEM37* (FC=-1.73 with adj. $p<0.001$ in T2D vs. ND in the "restricted" cohort). As in other studies in mouse models of diabetes[24], we found *ALDOB* to be also upregulated in islets from 13-week-old diabetic *db/db* mice compared to the heterozygous *db/+* littermate (Extended Data Fig. S3A) as well as in a mouse beta, but not alpha, cell line upon exposure to high glucose (Extended Data Fig. S3B). However, the overexpression of *ALDOB* in beta cells of T2D PPP could not be verified by immunofluorescence on tissue sections due to the cross-reactivity of the only available "specific" anti-ALDOB antibody with other aldolase isoforms (Extended Data Fig. S3C).

The islet proteome is more heterogeneous in T2D

To verify and extend the transcriptomic data at the functional level of proteins, we analyzed the mass spectrometry (MS)-based proteomic profiles of LCM pancreatic islets from five ND and five T2D PPP (Supplementary Table 8). We chose these samples primarily based on tissue availability and secondarily based on the levels of *ALDOB* found in RNA sequencing. Their islet transcriptomics profiles closely resembled the one of the complete cohort: top regulated genes in these 10 samples from T2D and ND PPP (Supplementary Table 2 and Extended Data Figure S4C) were also among the most significantly differentially expressed islet genes in the entire cohort. Using a very high sensitivity workflow on a novel trapped-ion mobility Time of Flight mass spectrometer and miniaturized sample preparation[25], we identified 2,237±499 islet proteins for ND PPP and 1,819±412 islet proteins for T2D PPP (Figure 4A). Quantitative reproducibility between biological replicates was high with Pearson correlations ranging from 0.83 to 0.95 (Extended Data Fig. S4A). Principal component analysis (PCA) clustered the data into two distinct groups matching the clinical stratification (Fig. 4B, see methods for detailed data processing steps). Interestingly, islets of ND PPP clustered closely, indicating a very similar proteome signature, while those of T2D PPP revealed substantial proteome heterogeneity among each other. Differential expression analysis confirmed that islets of T2D and ND PPP have very different proteomic profiles. The main differential drivers are well-characterized markers of pancreatic islet cells, including SLC2A2[26], and many proteins implicated in mitochondrial structure, translation, energy supply and amino acid or fatty metabolism such as YMEL1, MRPL12, BA3(C14orf159), ACADS and its paralogue ACADSB, which were highly depleted in islets of T2D PPP (Fig. 4C). Besides AKR7L, ACADS was the only other upregulated and differentially expressed gene in islets of both IGT and T2D PPP, while being also downregulated at the protein level. All differentially expressed mitochondrial

proteins are encoded by the nuclear genome (Fig. S4B). Intriguingly, the level of the sulfonylurea receptor ABCC8 subunit[27] was also strongly reduced in islets of T2D PPP. This downregulation might be an effect secondary to pharmacological treatment, as three among these patients had been treated with anti-diabetic SUR1 antagonists glibenclamide (DP197), glimepiride (DP118) or mitiglinide (DP087) (Extended Data Fig. S4C). Furthermore, we found that transcriptome and proteome levels of pancreatic islets from the same donors are very different (Extended Data Fig. S4D), as shown in another cellular system at single-cell level[28]. Nevertheless, we report the glycolytic enzyme ALDOB to be consistently upregulated (Proteome: 4-fold, Transcriptome: FPKM: 76.16±50.82 in T2D PPP vs. 4.63±0.95 in ND PPP), and the glucose transporter SLC2A2 to be downregulated (Proteome: 4-fold, Transcriptome: 4-fold) in islets of T2D vs. ND PPP samples on both modalities (Extended Data Fig. S4E, 4C). This is consistent with our transcriptomic data and that of previous studies[14,15] and our current WGCNA analyses. Other proteins robustly overexpressed in islets of T2D PPP included the alpha-L-fucosidase FUCA1 and the surface marker for hematopoietic stem cells THY1.

Next, we employed the proteomic ruler algorithm and annotations of subcellular localization to compare the protein mass distribution of major cellular compartments[29](Fig. 4D). Islets of T2D PPP lost an estimated protein mass of 6% in the Golgi apparatus, 24% in the endoplasmic reticulum, and 27% in the mitochondria compared to those of ND PPP, while the cytoskeleton protein mass was unchanged. Unsupervised hierarchical clustering of all 2,622 detected proteins, clustered the data according to clinical categories (Fig. 4E). One-dimensional gene ontology enrichment[30] revealed that two distinct clusters whose protein intensity levels associated with the terms ‘membrane attack complex’ ($p<2.18E-04$) and ‘Immunoglobulin C-domain’ ($p<2.68E-06$) were enriched by 2.27-fold and 2.36-fold in islets of T2D vs. ND PPP, respectively. Proteins with the gene ontology-term ‘differentiation’ ($p<3.09E-04$) and ‘mitochondrion’ ($p<2.19E-08$) were expressed 1.65 and 1.78-fold in islets of ND PPP.

Plasma phospho- and sphingo-lipid trends are opposite in T2D

Our study encompassed two independently generated lipidomics data sets. First, shotgun lipidomics was performed on peripheral blood plasma samples of the aforementioned cohort (4 ND, 21 IGT and IFG, 13 T3cD and 17 T2D) (Supplementary Tables 9 and 10). Second, sphingolipid profiling was performed on peripheral blood samples of subjects within the cohort subjected to transcriptomic analysis (11 ND, 32 IGT and IFG, 26 T3cD and 32 T2D) (Supplementary Tables 11 and 12). Prior to data analysis, lipidomics samples from PPP with very high bilirubin values (>100 µmol/l) were removed to avoid bias in lipidomics profiles. In each of the two data sets, all available samples from non-diabetic PPP (ND, as previously defined) and the subset of IGT PPP with HbA1c<6.0 were combined into one group, which is referred henceforth as ND for readability. The resulting sample sizes used in patient group comparisons were as follows: 17 ND, 13 T3cD and 17 T2D in the shotgun lipidomics data set; 32 ND, 21 T3cD and 27 T2D in the targeted lipidomics data set.

Statistical tests included covariates to adjust for age, sex and BMI, similar to the transcriptomics analysis. Pairwise comparisons of T2D vs ND and T3cD vs ND were performed. In shotgun lipidomics, 113 lipid species from 11 classes were included in the data analysis. When comparing T2D and T3cD to ND PPP, the majority of lipid classes displayed a remarkably homogeneous downward-trend of the individual lipid species they comprised (Fig 5A-B). Most prominently, plasma concentrations of lipids within the class of ether-linked phosphatidylcholines (PC O-), a large class with 30 measured species, were lower in T2D versus ND PPP. Fourteen lipids of this class were significantly decreased (adjusted $p<0.05$), with all of them showing at least a 1.4-fold change. A few lipid species from smaller phospholipid classes (one phosphatidylcholine (PC), one lysophosphatidylcholine (LPC) and one phosphatidylinositol (PI)), as well as two from the sphingomyelin class (SM), were also

significantly less abundant in T2D than in ND PPP (PC 18:0;0/18:2;0: FC=-1.43, adj. *p*=0.040; LPC 18:0;0: FC=-1.54, adj. *p*=0.037; PI 18:0;0/18:2;0: FC=-1.36, adj. *p*=0.045; SM 40:1;2:, FC=-1.33, adj. *p*=0.037; SM 34:1;2:, FC=-1.24, adj. *p*=0.04). (Fig. 5A-B and Supplementary Table 13).

Next, we performed targeted sphingolipidomics on 14 distinct lipid species for accurate plasma level estimation (ceramides, dihydroceramides and sphingoid bases) (Supplementary Table 16). Plasma levels of ceramides d18:1/18:0 and d18:1/20:0 were increased in T2D compared to ND PPP (Cer d18:1/18:0: FC=1.34, *p*=0.02; Cer d18:1/20:0: FC=1.22, *p*=0.01, without multiple testing correction). A similar trend towards elevation in T2D vs ND PPP was also observed in the two dihydroceramide species having the same chain lengths as these ceramides, although one of the two falls below the p-value threshold of 0.05 (DH Cer d18:0/18:0: FC=1.44, *p*=0.09; DH Cer d18:0/20:0: FC=1.35, *p*=0.02). Thus, in our data set, plasma concentrations of some ceramides and their precursor dihydroceramides appear to increase simultaneously in T2D. Cer d18:1/24:0, but not the corresponding dihydroceramide, was observed to move in the opposite direction being lower in T2D (FC=-1.28, *p*=0.017). Notably, ceramides were identified by LC-MS (see methods) and, independently, by shotgun FT-MS and both profiles matched. Regarding the LC-MS/MS analysis, the parent ion selected for dihydroceramides identification and quantification was the protonated ion (without water loss). In FT-MS, we observed no significant water loss from the ceramide standards. We therefore have reasons to believe that we detected dihydroceramides, but not deoxyceramides, which are isomeric of the water loss form of the dihydroceramides.[31]

Data integration identifies pathways for islet dysfunction

To identify a multivariate molecular profile that explains diabetes progression in the PPP cohort, we performed a large-scale integrative multi-omics analysis combining clinical data with islet transcriptomics and plasma lipidomics. Integration of transcriptomics and lipidomics data in the same model enables to weigh the relative importance of lipid and gene expression features in relationship to a chosen clinical trait. Hence, we explored the relationship between gene co-expression modules and plasma lipids by computing a consensus orthogonal partial least square (consensus OPLS)[32,33] model with HbA1c as the outcome. All three types of biological data, namely gene co-expression modules, lipids from the shotgun analysis and sphingolipids from the targeted analysis, contributed to the model (35%, 46.5% and 18.5%, respectively), suggesting that they help to explain HbA1c levels in a complementary way. Among them, different lipids and gene modules appear as the most relevant variables in the statistical modelling of HbA1c levels (Fig. 6A, 6B and Supplementary Table 14). Importantly, the model explained a large portion of data variance, highlighting a good fit with the experimental data (see Methods for more details).

Among all considered biological data, the co-expression modules M1, M4, M8, M9, M30, M35 and M36 were the top predictive variables for high HbA1c levels, along with the two ceramide species C20 and C18. TAGs were also contributing, although to a lesser extent (Fig 6A, right hand side). Conversely, low levels of HbA1c were strongly related to the co-expression modules M12 and M14 (Fig 6A, left hand side). However, the majority of the predominant predictors for low HbA1c were lipid species, most importantly the PC O- class. This class was also found to be lower in T2D compared to ND patient groups in differential abundance analysis, as shown in Fig 5A. A number of SM, PI and PC lipid species were next in the importance ranking related to low HbA1c, followed by the gene co-expression module M29. These results suggest that the profile of patients with increased HbA1c is characterized by multiple molecular components, some of which represent signals that were neither captured by differential abundance analyses

comparing diabetes status groups nor by correlating gene co-expression modules individually to HbA1c. Most importantly, consensus OPLS multi-omics analysis pointed towards additional gene co-expression modules that may play a role in glucose dysregulation.

Next, we used the results from the integrative data modelling to infer a network of key altered biological pathways in dysfunctional beta cells. To this end, we pooled gene modules positively associated with HbA1c levels (M1, M4, M8, M9, M30, M35 and M36) (Fig. 6A) and assessed their overlap to KEGG pathways by over-representation analysis. We found that the biological themes underlying these genes were very similar to the pathways upregulated in T2D and IGT PPP and include cell-matrix interaction, cell signaling and immune response (Fig. 6C and Supplementary Table 15). The same strategy was used to identify pathways associated with genes from modules with a negative prediction score for HbA1c (M12, M14 and M29) (Fig. 6A), revealing an enrichment for metabolic pathways (Fig. 6C and Supplementary Table 15). Of note, several islet genes differentially regulated in T2D PPP were driving the enrichment of these pathways. These include, for example, *ALDOB*, which stood out for its strong correlation to HbA1c levels (Fig. 3D and Fig. 6C). These genes, or the proteins encoded by them, should be regarded as putative candidate biomarkers for monitoring disease progression and therapeutic intervention.

Discussion

This study provides an extensive analysis on islets *in situ* and plasma samples from the largest cohort of in-depth metabolically profiled living donors. Multi-omics data were generated using state-of-the-art approaches and integrated in a fashion not previously used in studies on islet dysregulation in relation to hyperglycemia in humans. Our transcriptomic and proteomic data from islets *in situ* of ND subjects represent a valuable reference for future investigations. More

broadly, this dataset would be a worthwhile addition to the growing number of islet resources on type 1 and type 2 diabetes by different consortia, such as the Network for Pancreatic Organ Donors with Diabetes (nPOD)[34], Human Pancreas Analysis Program (HPAP)[35], or the Translational human pancreatic Islet Genotype tissue - Expression Resource (TIGER) (https://www.t2dsystems.eu/tiger-database). Furthermore, we could identify for the first time a set of islet genes altered in their expression already in subjects with impaired glucose tolerance. This, in turn, enabled us to acquire an unprecedented cross-sectional overview of the progression of islet gene dysregulation in parallel with the continuous elevation of HbA1c values, beyond conventional thresholds for clinical classification of patients.

Pathways involved in RNA biology and especially in mitochondrial function emerged to be most negatively perturbed - a conclusion which in the case of the latter was strongly corroborated by the proteomic analysis, which enabled the identification of known and unknown differentially expressed proteins in islets of T2D PPP. In this context, we emphasize the downregulation of mitochondrial ACADS and its paralogue ACADSB, which catalyze the beta oxidation of short-chain fatty acids, including sodium butyrate. This finding is intriguing in view of the ability of this metabolite to broadly upregulate gene expression through inhibition of histone deacetylases. Unlike in previous studies on isolated islets from brain-dead organ donors[14,19], but similar to previous studies by us in human[14] and mouse models of diabetes[36] the vast majority of differentially expressed genes in islets of T2D, but also IGT and T3cD PPP were upregulated. Among those genes, *ALDOB* stands out being the one with the strongest correlation with the islet gene module M9, which in turn has the strongest correlation with elevated HbA1c. Since *ALDOB* is a marker of beta cell precursors[37], its overexpression could be interpreted as a sign that in T2D, mature beta cells revert back to an immature stage of differentiation, or that a compartment equivalent to the lifelong niche of virgin beta cells identified in adult mice expands as a potential compensatory source of new beta cells[37]. However, no additional disallowed gene

of immature beta cells, markers of beta cell precursors or other islet cell types were differentially regulated, while key determinants of mature beta cells, such as *PDX1*, *MAFA*, *NKX6.1* or *UCN3* were unchanged, at least at the transcriptomic level. Retention of fractions of major islet cell types (alpha, beta and delta) within the islet in T2D, consistent with recent imaging studies in samples from pancreatectomized subjects (Cohrs et al)[17], was confirmed by deconvolution analysis. Our global unbiased proteomic analysis, which corroborated the upregulation of ALDOB, further showed that the expression profile of islet cells in T2D PPP is very divergent, opposite to its remarkable homogeneity in islet cells of ND subjects. Hence, the regression of beta cells toward a de-differentiated state following a linear trajectory recapitulating their developmental path to maturation or their transdifferentiation into other islet cell types seems less likely than a disharmonic relaxation of constraints on gene expression. Such processes, although possibly reversible, could perturb the coordinated operation of islet cells, including beta cells. In line with this, Lawlor *et al.* reported no evidence of beta cell dedifferentiation/transdifferentiation and alterations in fractions of islet cells in the context of T2D upon sequencing of single islet cells from a small cohort of ND and T2D organ donors, although this conclusion has been more recently challenged[38]. While we strived to selectively enrich the beta cell content of our omics data by laser capture microdissection of bulk islets based on the lipofuscin-associated autofluorescence of beta cells and by subsequent deconvolution of the data during their analysis, the unavoidable presence of other cell types in the samples introduces a degree of uncertainty. Thus, for the future it would be important to assess whether overexpression of ALDOB occurs indeed in beta cells and if it affects their glycolysis and metabolism, taking into account that its paralogue ALDOA, whose RNA and protein levels were unchanged, remains by far the predominant islet aldolase species. Attention may also be directed toward understanding whether impaired oxidative phosphorylation, as a likely outcome of the massively decreased expression of mitochondrial proteins, and thus energy balance homeostasis, accounts, at least in part, for the observed less restrained gene expression.

The transcriptome and proteome of islets from subjects with T2D displayed the occurrence of an immune response. At this time, however, we are not aware of factors which might readily account for the presence of such signals. Specifically, patients with antibodies against known autoantigens of T1D were excluded from the analysis. As in a previous report[15], histological examination of the specimens did not reveal insulitis or macrophage infiltration. Pancreatitis was more common among subjects with normoglycemia (22%) than with T2D (15%). The presence of cancer cells in our islet preparations is also unlikely. Specifically, a qualified pathologist routinely examined the surgical specimen to minimize the chance of contamination by neoplastic tissue before it was taken for downstream processing. Histological survey of the tissue did not reveal the presence of neoplastic cells in the islets. The transcriptomic analysis in a previous study indicated that exocrine contamination of LCM islets from PPP was comparable or less than in the case of enzymatically-isolated islets from organ donors (OD)[14]. Moreover, in the same study we found no evidence for an enrichment of tumor cell transcripts in LCM islets of PPP compared to islets of OD. Likewise, an enrichment analysis of pancreatic cancer specific genes in the differentially expressed islet gene sets reported here using hypergeometric test showed no enrichment for any of the four described pancreatic cancer subtypes (p = 0.87) as reported in Bailey et al 2016[39]. Rather, the gene expression clustering was driven by the islet isolation method and not by the origin of the tissue (OD vs. PPP). We still appreciate that in our cohort cancer prevalence was higher in the T2D (69%) than in the ND (50%) group. Thus, we cannot entirely rule out a metabolic pro-inflammatory impact of the cancer on islet gene and protein expression or function.

Our lipidomics analyses revealed lowered phospholipid species (14 PC O-, one PC, one LPC, one PI) and some elevated ceramides and dihydroceramides in T2D PPP. These findings match observations reported in other recent studies on larger cohorts. Huynh et al (2019)[40] presented

a comprehensive shotgun lipidomics study on the AusDiab cohort, including 640 samples and 636 lipid species. In this work, many PC, PC O-, LPC and PI had a significant negative association with blood glucose levels either after overnight fasting or at the 2-hour point of an OGTT, including nine species that were found negatively associated with T2D in our own study (PC 18:2;0/18:2;0, LPC 18:0;0, PI 18:0;0/18:2;0 and six PC O- species). In the same study, the ceramide Cer d18:1/18:0 and its precursor DH Cer d18:0/18:0 had both a significant positive association with fasting glucose, supporting our notion that this lipid pair might be linked to diabetes status. Furthermore, several prospective case-control studies reported significantly decreased PC, PC O- and LPC plasma concentrations[41] or elevated dihydroceramide levels[42,43] in progressors to T2D compared to non-diabetic controls. The congruency of these results points to these lipids as potential biomarkers of beta cell function in T2D.

Finally, we use a data fusion method[32,33] to generate a model of how different molecular features (islet gene co-expression, plasma shotgun lipidomics and targeted sphingolipidomics) contribute to HbA1c levels in a continuum from healthy individuals to those with overt T2D. This model allowed us to measure the *relative* importance of different molecular components in explaining HbA1c variability, providing unique insights into the molecular profiles of individuals as they lose glycemic control towards development of T2D. The rational for combining plasma lipidomics with islet gene expression data was that the levels of some plasma lipids may affect pancreatic islets and/or reflect changes occurring within them and thus be useful as biomarkers to assess beta cell dysfunction in prediabetes and T2D. To our knowledge this is the first time such an approach has been used in this field and we suggest that, by modelling multiple levels of information at the same time in deeply phenotyped populations such as the one presented here, we can gain a holistic view of the system and draw conclusions regarding key pathways, targets and biomarkers in metabolic and other diseases.

Data availability

RNA Sequencing data was deposited in the NCBI Gene Expression Omnibus with GEO accession number GSE164416. Human genome reference assembly GRCh38 is publicly available.

The proteomics raw datasets and the MaxQuant output files generated and analyzed throughout this study were deposited at the ProteomeXchange Consortium via the PRIDE partner repository with the project accession number PXD022561 (https://www.ebi.ac.uk/pride/archive/). Lipidomics data was deposited in the Zenodo database (zenodo.org, doi:10.5281/zenodo.4716063).

Acknowledgements

We wish to thank Leif Groop, Emma Ahlqvist, Stephan Speier, Triantafyllos Chavakis, Raphael Scharfmann and Andrej Schevchenko for discussion; Sylvain Pradervand for advice on RNA-Seq processing and QC; Corinne Rocher for Bioinformatics analysis support; Katja Pfriem for administrative assistance. This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 115881 (RHAPSODY). This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation program and EFPIA. This work is further supported by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 16.0097-2. Work in the Solimena lab is also supported with funds from the German Ministry of Education and Research to the German Center for Diabetes Research (DZD). The opinions expressed and arguments employed herein do not necessarily reflect the official views of these funding bodies.

Author contributions

J.W. and M.D., patient recruitment and surgery, provision of clinical data; E.S., N.K. and D.F., sample collection and processing, data entry; D.A., pathology; M.B., N.K. and E.S., patient database management and selection; A-D.B. and M.M., proteomics; M.L., A.D., RNA sequencing, C.L.Q., P.B.S.H, P.D., C. K., M. G., K.S., lipidomics and sphingolipidomics; L.W., M.B., A-D.B., F.Ma., F.Me., F.B. and Ca.K., analysis and integration of multi-omics data; E.B., autoantibody test; A.S., data in mouse tissue and cell lines; M.B., immunofluorescence stainings and antibody validation; B.T., D.A., J.W., A.S., M.M., M.I. and M.S., conceptual insights and provision of funds; L.W., M.B., A-D.B., F.Ma., F.Me., A.S., M.I., M.M. and M.S., writing of the manuscript. All authors read, revised and approved the final version of the manuscript.

Competing interests

KS is CEO of Lipotype GmbH. KS and CK are shareholders of Lipotype GmbH. MJG is an employee of Lipotype GmbH. PBSH and PD are employees of Servier. AS is an employee of Sanofi-Aventis Deutschland GmbH. The other authors declare no conflict of interest.

References

1. Saeedi, P. *et al.* Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. *Diabetes Res. Clin. Pract.* **157**, (2019).
2. Mizera, M. *et al.* Type 2 Diabetes Remission 5 Years After Laparoscopic Sleeve Gastrectomy: Multicenter Cohort Study. *Obes. Surg.* 1–7 (2020). doi:10.1007/s11695-020-05088-w
3. Lim, E. L. *et al.* Reversal of type 2 diabetes: Normalisation of beta cell function in association with decreased pancreas and liver triacylglycerol. *Diabetologia* **54**, 2506–2514 (2011).

4. Talchai, C., Xuan, S., Lin, H. V., Sussel, L. & Accili, D. Pancreatic β cell dedifferentiation as a mechanism of diabetic β cell failure. *Cell* **150**, 1223–1234 (2012).
5. Wang, Z., York, N. W., Nichols, C. G. & Remedi, M. S. Pancreatic β cell dedifferentiation in diabetes and redifferentiation following insulin therapy. *Cell Metab.* **19**, 872–882 (2014).
6. Cinti, F. *et al.* Evidence of β-Cell Dedifferentiation in Human Type 2 Diabetes. *J. Clin. Endocrinol. Metab.* **101**, 1044–1054 (2016).
7. American Diabetes Association. Classification and diagnosis of diabetes: Standards of Medical Care in Diabetes-2020. *Diabetes Care* **43**, S14–S31 (2020).
8. Barovic, M. *et al.* Metabolically phenotyped pancreatectomized patients as living donors for the study of islets in health and diabetes. *Molecular Metabolism* (2019). doi:10.1016/j.molmet.2019.06.006
9. Poitout, V. *et al.* A call for improved reporting of human islet characteristics in research articles. *Diabetes* **68**, 209–211 (2019).
10. Ebrahimi, A. *et al.* Evidence of stress in β cells obtained with laser capture microdissection from pancreases of brain dead donors. *Islets* **9**, 19–29 (2017).
11. Toyama, H., Takada, M., Suzuki, Y. & Kuroda, Y. Activation of macrophage-associated molecules after brain death in islets. *Cell Transplant.* **12**, 27–32 (2003).
12. Negi, S. *et al.* Analysis of Beta-Cell gene expression reveals inflammatory signaling and evidence of dedifferentiation following human islet isolation and culture. *PLoS One* **7**, 1–11 (2012).
13. Weir, G. C. Glucolipotoxicity, β-cells, and diabetes: The emperor has no clothes. *Diabetes* **69**, 273–278 (2020).
14. Solimena, M. *et al.* Systems biology of the IMIDIA biobank from organ donors and pancreatectomised patients defines a novel transcriptomic signature of islets from individuals with type 2 diabetes. *Diabetologia* **61**, 641–657 (2018).

15. Gerst, F. *et al.* The Expression of Aldolase B in Islets is Negatively Associated with Insulin Secretion in Humans. *J. Clin. Endocrinol. Metab.* **103**, 4373–4383 (2018).
16. Khamis, A. *et al.* Laser capture microdissection of human pancreatic islets reveals novel eQTLs associated with type 2 diabetes. *Mol. Metab.* **24**, 98–107 (2019).
17. Cohrs, C. M. *et al.* Dysfunction of Persisting β Cells Is a Key Feature of Early Type 2 Diabetes Pathogenesis. *Cell Rep.* **31**, (2020).
18. Viñuela, A. *et al.* Genetic variant effects on gene expression in human pancreatic islets and their implications for T2D. *Nat. Commun.* **11**, (2020).
19. Mahajan, A. *et al.* Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513 (2018).
20. Taneera, J. *et al.* Identification of novel genes for glucose metabolism based upon expression pattern in human islets and effect on insulin secretion and glycemia. *Hum. Mol. Genet.* **24**, 1945–1955 (2014).
21. Carrat, G. R. *et al.* Decreased STARD10 Expression Is Associated with Defective Insulin Secretion in Humans and Mice. *Am. J. Hum. Genet.* **100**, 238–256 (2017).
22. Xin, Y. *et al.* RNA Sequencing of Single Human Islet Cells Reveals Type 2 Diabetes Genes. *Cell Metab.* **24**, 608–615 (2016).
23. Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
24. Haythorne, E. *et al.* Diabetes causes marked inhibition of mitochondrial metabolism in pancreatic β-cells. *Nat. Commun.* **10**, (2019).
25. Meier, F. *et al.* Online parallel accumulation–serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer. *Mol. Cell. Proteomics* **17**, 2534–2545 (2018).
26. Thorens, B. GLUT2, glucose sensing and glucose homeostasis. *Diabetologia* **58**, 221–232 (2015).

27. Pipatpolkai, T., Usher, S., Stansfeld, P. J. & Ashcroft, F. M. New insights into KATP channel gene mutations and neonatal diabetes mellitus. *Nature Reviews Endocrinology* **16**, 378–393 (2020).

28. Brunner, A. D. *et al.* Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *bioRxiv* 2020.12.22.423933 (2020). doi:10.1101/2020.12.22.423933

29. Wiśniewski, J. R., Hein, M. Y., Cox, J. & Mann, M. A 'proteomic ruler' for protein copy number and concentration estimation without spike-in standards. *Mol. Cell. Proteomics* **13**, 3497–3506 (2014).

30. Cox, J. & Mann, M. 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. *BMC Bioinformatics* **13 Suppl 16**, (2012).

31. Zitomer, N. C. *et al.* Ceramide synthase inhibition by fumonisin B1 causes accumulation of 1-deoxysphinganine. A novel category of bioactive 1-deoxysphingoid bases and 1-deoxydihydroceramides biosynthesized by mammalian cell lines and animals. *J. Biol. Chem.* **284**, 4786–4795 (2009).

32. Boccard, J. & Rutledge, D. N. A consensus orthogonal partial least squares discriminant analysis (OPLS-DA) strategy for multiblock Omics data fusion. *Anal. Chim. Acta* **769**, 30–39 (2013).

33. Boccard, J. & Rutledge, D. N. Iterative weighting of multiblock data in the orthogonal partial least squares framework. *Anal. Chim. Acta* **813**, 25–34 (2014).

34. Campbell-Thompson, M. *et al.* Network for Pancreatic Organ Donors with Diabetes (nPOD): Developing a tissue biobank for type 1 diabetes. *Diabetes. Metab. Res. Rev.* **28**, 608–617 (2012).

35. Kaestner, K. H., Powers, A. C., Naji, A. & Atkinson, M. A. NIH initiative to improve understanding of the pancreas, islet, and autoimmunity in type 1 diabetes: The Human

Pancreas Analysis Program (HPAP). *Diabetes* **68**, 1394–1402 (2019).

36. Ebrahimi, A. G. *et al.* Beta cell identity changes with mild hyperglycemia: Implications for function, growth, and vulnerability. *Mol. Metab.* **35**, (2020).

37. van der Meulen, T. *et al.* Virgin Beta Cells Persist throughout Life at a Neogenic Niche within Pancreatic Islets. *Cell Metab.* **25**, 911-926.e6 (2017).

38. Avrahami, D. *et al.* Single-cell transcriptomics of human islet ontogeny defines the molecular basis of β-cell dedifferentiation in T2D. *Mol. Metab.* **42**, (2020).

39. Bailey, P. *et al.* Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* **531**, 47–52 (2016).

40. Huynh, K. *et al.* High-Throughput Plasma Lipidomics: Detailed Mapping of the Associations with Cardiometabolic Risk Factors. *Cell Chem. Biol.* **26**, 71-84.e4 (2019).

41. Suvitaival, T. *et al.* Lipidome as a predictive tool in progression to type 2 diabetes in Finnish men. *Metabolism.* **78**, 1–12 (2018).

42. Wigger, L. *et al.* Plasma Dihydroceramides Are Diabetes Susceptibility Biomarker Candidates in Mice and Humans. *Cell Rep.* **18**, 2269–2279 (2017).

43. Mamtani, M. *et al.* Lipidomic risk score independently and cost-effectively predicts risk of future type 2 diabetes: Results from diverse cohorts. *Lipids Health Dis.* **15**, (2016).

Figure Legends

Figure 1: Overview of the experimental procedures and cohort characteristics. A) Experimental procedures overview. Clinical data and peripheral blood were collected preoperatively, and the snap-frozen surgical pancreatic tissue used for LCM of the islets of Langerhans. Blood samples were analyzed for lipidomics, while LCM islets for transcriptomics and proteomics. Omics datasets were individually evaluated in relationship to glycemic status and further integrated with each other using Consensus Orthogonal Partial Least Squares

(OPLS) analysis. B) Waffle plot showing the structure of the cohort in terms of glycemic/diabetes categories based on American Diabetes Association criteria. Absolute numbers for each category are given in the legend boxes. C) Boxplots of four major clinical parameters relevant for diabetes diagnosis and management. Statistically significant differences from ND PPP were determined using the two-sided t-test (*p<0.05; **p<0.01). Boxplot spans from 25th until 75th percentile, with centerline at median, whiskers extend to the most extreme data point which is no more than 1.5 times the length of the box away from the box. Number of observations in each comparison and category: age and BMI - 18 for ND, 41 for IGT(p=0.006 and p=3.792-4), 35 for T3cD (p=0.001 and p=0.003) and 39 for T2D (p=0.003 and p=0.005); fasting glycemia - 16 for ND, 38 for IGT (p=2.936-6), 34 for T3cD (p=1.249-7) and 33 for T2D (p=2.692-7); glycemia at 2h point of OGTT: 15 for ND, 38 for IGT (p=1.486-6) and 23 for T3cD (p=3.111-11).LCM: Laser Capture Microdissection; ND: Non-diabetic; IGT: Impaired Glucose Tolerance; T3cD: Type 3c Diabetes; T2D: Type 2 Diabetes.

Figure 2: Transcriptional changes between non-diabetic, pre-diabetic and diabetic patients. A) Number of DE genes identified by comparing glycemic groups of PPP in the entire (all samples) or “restricted” cohort (*INS* filtered), using linear model with age, sex and BMI as covariates. B) Gene expression profile of DE genes in the “restricted” cohort. Columns represent patients grouped according to their glycemic status and ordered based on increasing HbA1c levels. Rows, representing DE genes (variance stabilizing transformation normalized counts), were clustered based on Euclidean distance. The colored side bar indicates in which comparisons a gene was identified as differentially expressed. C) Gene Set Enrichment Analysis of DE genes between IGT, T3cD or T2D and ND PPP in the “restricted” cohort. GO terms and KEGG pathways are colored according to the normalized enrichment score. Corresponding p-values are also indicated (*p<0.05, **p<0.01). D) Immunofluorescence for insulin (green), acetylated histones H3 (left) and H4 (right) (magenta) in representative samples

of formalin fixed paraffin embedded pancreatic tissues from 5 ND and 5 T2D PPP. Scale bars correspond to 20µm. DE: differentially expressed; ND: Non-diabetic; IGT: Impaired Glucose Tolerance; T3cD: Type 3c Diabetes; T2D: Type 2 Diabetes.

Figure 3: Identification of co-expressed gene modules related to diabetes traits. A) Correlation between module eigengenes and clinical traits including age, BMI, HbA1c, fasting glucose, glucose at 2-hours after OGTT, HOMA2-B and HOMA2-IR. Each cell contains the corresponding Spearman correlation coefficient and Student *p* value (in parenthesis). Cells are colored according to their correlation to clinical traits. Modules are ordered based on their correlation to HbA1c. B-C) Gene expression profiles of gene modules M9 (B) and M14 (C). Columns, representing PPP, were grouped according to their glycemic status and ordered based on increasing HbA1c levels. Rows, representing genes (variance stabilizing transformation normalized counts), were clustered based on Euclidean distance. D-E) Scatter plot of module membership vs. gene significance for HbA1c in modules M9 and M14. Genes with the highest module membership and gene significance ("hub genes") are labeled. ND: Non-diabetic; IGT: Impaired Glucose Tolerance; T3cD: Type 3c Diabetes; T2D: Type 2 Diabetes.

Figure 4: Proteomics Analysis. A) Number of identified proteins from pooled human pancreatic islet cells isolated by LCM from PPP classified as non-diabetic (ND, N=5) or with T2D (N=5). Boxplot spans from 25th until 75th percentile with centerline at median. Whiskers extend to the most extreme data points in either direction. B) Principal Component Analysis (PCA) of all grouped pancreatic islet measurements (ND=blue, T2D=orange). C) Volcano plot comparing *p* values and $\log_2$-fold changes between islets of ND and T2D PPP. Multiple hypothesis testing is controlled via Benjamini Hochberg correction at 5% False discovery rate. D) Percentage distribution of total protein islet mass and its contribution per organelle between ND and T2D PPP. The ND/T2D islet protein mass ratio in different organelles was normalized by the nucleus protein mass. E) Hierarchical clustering of all islet proteins identified in the T2D

and ND PPP clusters. Log_2-transformed intensity values were normalized by z-scoring before the clustering followed by one-dimensional gene ontology enrichment for cellular compartment and keywords for each of the clusters. Distribution of systematically enriched clusters is shown as the geometric mean at 95% confidence interval for each respective term in non-diabetic (ND, N = 5) and type 2 diabetics (T2D, N = 5) with centerline at the geometric mean with 95% confidence interval.

Figure 5: Lipidomics differential analysis. A-B) Shotgun lipidomics covering a variety of lipid classes: Ceramides (Cer), Diacylglycerols (DAG), Lysophosphatidylcholines (LPC), Lysophosphatidylethanolamines (LPE), Phosphatidylcholines (PC), Ether-linked Phosphatidylcholines (PC O-), Phosphatidylethanolamines (PE), Ether-linked Phosphatidylethanolamines (PE O-), Phosphatidylinositols (PI), Sphingomyelins (SM), Triacylglyerols (TAG). Volcano plots represent comparisons of plasma lipid levels between ND and T2D PPP. The X-axis shows direction and magnitude of the change; the Y-axis represents the statistical significance of the change. Each point is a lipid species, colored by lipid class to highlight class-specific trends. C) Targeted lipidomics on dihydroceramides (DH Cer), ceramides (Cer) and Sphingoid bases (SB). Each heatmap column represents the comparisons of plasma levels between ND and T2D PPP. Heatmap colors represent direction and magnitude of the change. Log_2 Fold Change: ratio of mean lipid concentration in the two groups, log_2 transformed. Statistical model used for all panels: linear regression with age, sex and BMI as covariates (p: *p* value); adjustment of *p* values across all lipid species by the Benjamini-Hochberg method (adj. *p*: adjusted *p* value). T2D: Type 2 Diabetes; T3cD: Type 3 Diabetes; ND & PD: non-diabetic and pre-diabetic (with impaired fasting glucose and/or impaired glucose tolerance) with HbA1c<6.0.

Figure 6: Multiblock data modeling of HbA1c. A) Bar plot showing the variable importance in the multiblock consensus OPLS model. The Y-axis represents the importance scores for the

753 predictors multiplied by the sign of the loadings on the predictive latent variable. Variables with
754 importance in projection > 1.2 were selected. B) Statistical significance of the model through
755 permutation test. C) Network representation of functional pathways enriched in modules with
756 best prediction scores for HbA1c. Pathways are represented as gray nodes. Genes are
757 represented as nodes sized based on their correlation to HbA1c and colored based on their
758 differential expression in T2D vs. ND PPP. Only genes with significant differential expression
759 (adj. p<0.05) in the “restricted” cohort are shown. VIP Variable Importance in Projection, DE:
760 Differentially expressed; ND: Non-diabetic, T2D: Type 2 Diabetes.

761

Material and methods

Cohort

Our cohort comprised 133 adult patients undergoing pancreatic surgery for a variety of indications (benign and malignant neoplasms, chronic pancreatitis, pancreatic cysts etc.) from the University Hospital Carl Gustav Carus Dresden who after informed consent participated in this study over a period of 5 years. The study was conducted with the ethical approval of the Ethical Committee of the Technische Universität Dresden. Based on the thresholds set by the American Diabetes Association[4] (ADA) for fasting glucose, HbA1c and 2-hour glycemia of an oral glucose tolerance test (OGTT) in the days immediately before surgery 18 of these patients were classified as non-diabetic (ND), 41 with impaired glucose tolerance (IGT), including 3 with impaired fasting glucose (IFG) only, 35 with Type 3c Diabetes (T3cD) and 39 with Type 2 Diabetes (T2D). A diagnosis of T3cD was made whenever the occurrence of diabetes was not recognized for longer than 1 year prior to the onset of the symptoms leading to surgery and the subject was negative for the presence of circulating autoantibodies against pancreatic islets, which were assessed as previously described[11]. In all analyses IFG and IGT subjects were merged in one group hereafter labeled as IGT PPP. Medical and family history and relevant clinical biochemistry data available from the routine medical processing of the patients were retrieved from the hospital database and referring physicians. Patients who underwent neoadjuvant chemotherapy as well as those with endocrine neoplasms of the pancreas were excluded from this study.

Human pancreatic tissue and peripheral blood processing

Surgical tissue specimens were examined by a certified pathologist immediately after resection as per regular clinical procedures. Fragments of healthy pancreatic tissue from the resection margins were excised, snap frozen in liquid nitrogen and stored at -80^{o}C either natively or embedded in TissueTek OCT compound. Estimated warm and cold ischaemia time was on

average 2 hours. Peripheral blood samples were stored at -80°C in aliquots of full blood, plasma and serum.

Transcriptomics

Islet procurement and RNA isolation

Pancreatic tissue was sectioned in a cryostat and mounted on UV pre-treated Zeiss MembraneSlide 1.0 PEN slides. Laser capture microdissection (LCM) was done with a Zeiss Palm MicroBeam system using autofluorescence to identify islets, as previously described[44].

RNA was isolated from approximately 20x6µm3 of islet tissue using the Arcturus PicoPure RNA Isolation Kit. Only preparations with RNA Integrity Number ≥5 were used for RNA sequencing. The entire handling of the tissue samples was done in a strictly RNAse free environment.

Library preparation, RNA Sequencing and alignment

Sequencing libraries were prepared from bulk RNA using the Illumina SmartSeq protocol. Single ended 76bp sequencing was done with an Illumina HiSeq 2500 or Illumina HiSeq 500 at the Next Generation Sequencing Core Facility of the CMCB Dresden, with the target depth of 35 million fragments per library. From FASTQ files, purity-filtered reads were trimmed with Cutadapt to remove adapters and low-quality sequences (v. 1.8)[45]. Reads matching to ribosomal RNA sequences were removed with fastq_screen (v. 0.11.1)[46]. Remaining reads were further filtered for low complexity with reaper (v. 15-065)[47]. Reads were aligned against Homo sapiens GRCh38.92 genome using STAR (v. 2.5.3a)[48]. The number of read counts per gene locus was summarized with htseq-count (v. 0.9.1)[49] using Homo sapiens GRCh38.92 gene annotation. Quality of the RNA-seq data alignment was assessed using RSeQC (v. 2.3.7)[50].

RNA Sequencing quality control, processing and differential expression analysis

RNA Sequencing datasets were screened for exocrine contamination in an initial quality control (QC) step. Analysis of the absolute number of detected expressed genes, gene body coverage

and cumulative gene diversity assessment flagged a number of libraries to be of insufficient quality for downstream analysis. Libraries were filtered for minimal expression by removal of genes with less than 5 mean raw reads. Reads were normalized for library size and transformed for variance stabilizing using tools from the DESeq2 Bioconductor package[51]. Further analysis revealed 41 libraries in which transcripts other than insulin (*INS*) displayed the highest normalized number of reads. Differential expression analysis across the clinical categories (ND, IGT, T3cD, T2D) was performed using limma function with voom approach from the limma Bioconductor package[52,53] on both the full dataset of 133 libraries which passed the QC analysis as well as on the "restricted" dataset of 92 libraries featuring *INS* as the highest expressed gene based on the linear model with age, sex and BMI as covariates. All analysis pertaining transcriptomic data was done on R platform (version 3.6.3).

Gene set enrichment analysis of differentially expressed genes

Functional enrichment analyses of differentially expressed genes in IGT, T2D or T3cD compared to ND patients were performed by weighted gene set enrichment analysis (GSEA) on unfiltered gene lists ranked by decreasing differential expression test statistics. Gene Ontology (GO) term and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway collections were restricted to gene sets with a minimum and maximum sizes of 100 and 500, respectively. The enrichment scores were normalized by gene set size and their statistical significance was assessed by permutation tests (n=1,000). GO enrichment analyses were carried out using the gseGO function from the R package clusterProfiler (version 3.10.1)[54]. GO terms enriched in at least one comparison were identified using *p* value and normalized enrichment score thresholds < 0.01 and > 2.5, respectively. Redundancy of enriched GO terms was removed using the clusterProfiler simplify function (selecting the most representative term by *p* value) and enrichment maps generated using the emapplot function from the R package enrichplot (version 1.2.0). KEGG pathway enrichment analyses were performed using the clusterProfiler gseKEGG

function. Results were filtered based on a *p* value threshold < 0.01 and a normalized enrichment score threshold > 2. To simplify results visualization and interpretation, redundant KEGG pathways were also collapsed into fewer biological themes using the enrichment map visualizations.

Weighted Gene Correlation Network Analysis

Gene Co-expression Network Construction

The gene co-expression network was created following the weighted gene correlation network analysis (WGCNA) protocol as implemented in the WGCNA package in R (version 1.68)[23], as previously described[14]. WGCNA was performed on batch-corrected, normalized and variance stabilizing transformed expression data from the full cohort of 133 subjects. The co-expression network was constructed by calculating an adjacency matrix using Pearson correlation, pairwise complete observations and unsigned method. The soft-threshold parameter was optimized with the function pickSoftThreshold and the best threshold ($\alpha = 7$) selected by visual inspection. The adjacency matrix was then computed into a topological overlap matrix (TOM), converted to distances, and clustered by hierarchical clustering using average linkage clustering. Modules were identified by dynamic tree cut using the hybrid method and parameters minClusterSize=20 and deepSplit=2. Similar modules were merged using a module eigengene distance of 0.15 as the threshold.

Identification of co-expressed gene modules

We correlated the module eigengenes to clinical traits using Spearman correlation (pairwise complete observations) and calculated the corresponding *p* values using the cor and corPvalueStudent functions from the WGCNA package, respectively. Module-trait correlations were represented as heatmap using the labeledHeatmap function from the WGCNA package. The modules displaying the most positive or negative correlation to HbA1c were further

analysed. Normalized and variance stabilizing transformed gene counts for selected modules were plotted as heatmap using the heatmap.2 function from the R gplots package (version 3.0.1.2). Rows (representing genes) were scaled and hierarchically clustered by Euclidean distances. Columns, representing patients, were custom ordered as described in the legend of figure 3. Module hub genes, such as highly connected genes within a module that could have a strong influence on a phenotypic trait, were identified as those with highest correlations with the particular trait and highest correlations with the module eigengene.

Significance of gene co-expression modules

We tested the significance of the co-expression modules by comparing their intramodular connectivity (connectivity between nodes within the same module, as computed by the WGCNA intramodularConnectivity function) to the background as follows. For each selected module of size N, we calculated a Z-score as in equation 1:

$$Z=(k-\mu)/\sigma \quad (1)$$

where k is the intramodular connectivity and μ and σ are the mean and standard deviation of the intramodular connectivity from 1,000 randomly sampled modules of size N respectively. Empirical p values were then calculated as the fraction of random intramodular connectivity values ≥ to the observed intramodular connectivity. For the modules with the highest variable importance in projection score in the HbA1c multiblock model, all of the random intramodular connectivity values were below the observed intramodular connectivity, suggesting that these modules were more compact than modules assembled by randomly sampling the same number of genes from the expression data (Supplementary Table 7).

Functional profiles of gene modules most predictive for HbA1c

The clusterProfiler enrichKEGG function was used to test for the over representation of selected co-expressed gene modules in KEGG pathways using hypergeometric distribution. A *p* value threshold < 0.01 was used to identify enriched terms. Enrichment map visualizations were used

to overcome gene set redundancy. Results were displayed as networks of enriched pathways and overlapping genes using cytoscape (version 3.5.1).

Deconvolution analysis

In all samples a cell proportions matrix was produced using the R package DeconRNASeq (v.1.26.0) on RPKM-transformed data. The signature file provided to DeconRNASeq comes from Xin et al. (2016)[22], Supplementary Table S2A, obtained using single-cell data. It was adapted to the human genome version 38 by excluding 15 obsolete genes.

Lipidomics

Sample availability and sample overlap with transcriptomics data

Pre-operative plasma lipidomics samples were obtained from a subset of the PPP cohort. Shotgun lipidomics analysis was performed on plasma from 55 PPP. These included 53 subjects who also had their islet transcriptomics profile included in this study plus two PPP who were not part of the transcriptomics analysis because the RNA-Seq data failed to pass the quality control. Moreover, targeted sphingolipid analysis was performed on plasma from 101 PPP. These included 98 PPP whose transcriptomics data was also included in this study plus three PPP whose RNA-Seq data was excluded for quality reasons. The number of samples in the two types of lipidomics analysis was smaller than in islet transcriptomic analysis because of the limited availability of plasma samples. The 55 PPP with shotgun lipidomics data were a subset of the 101 PPP with targeted sphingolipid data, with the difference in sample numbers being determined by plasma sample availability as well.

Shotgun lipidomics measurements

A streamlined mass-spectrometry (MS) -based platform for shotgun lipidomics developed by Lipotype GmbH (Dresden, Germany) was used for lipidomic profiling of patient plasma samples. Lipid extraction, internal standard addition and infusion into the mass spectrometer were performed as previously described[55]. The internal standard mixture contained: cholesterol D6 (chol), cholesterol ester 20:0 (CE), ceramide 18:1;2/17:0 (Cer), diacylglycerol 17:0/17:0 (DAG), phosphatidylcholine 17:0/17:0 (PC), phosphatidylethanolamine 17:0/17:0 (PE), lysophosphatidylcholine 12:0, (LPC) lysophosphatidylethanolamine 17:1 (LPE), triacylglycerol 17:0/17:0/17:0 (TAG) and sphingomyelin 18:1;2/12:0 (SM).

Samples were analyzed by direct infusion in a QExactive mass spectrometer (Thermo Scientific) in a single acquisition. Tandem mass-spectrometry (MS/MS) was triggered by an inclusion list encompassing corresponding MS mass ranges scanned in 1 Da increments. MS and MS/MS data were combined to monitor CE, DAG and TAG ions as ammonium adducts; PC, PC O-, as acetate adducts; and PE, PE O- and PI as deprotonated anions. MS only was used to monitor LPE as deprotonated anion; Cer, SM and LPC as acetate adducts and cholesterol as ammonium adduct.

Data post-processing and normalization were performed using an in-house developed data management system. Only lipid identifications with a signal-to-noise ratio >5 and a signal intensity 5-fold higher than in corresponding blank samples were considered for further analysis. The median coefficient of lipid subspecies variation (RSD), as accessed by the repeated analysis of reference samples, was 7.5%.

Targeted sphingolipid measurements

Ceramides (C16:0 cer, C18:0 cer, C18:1 cer, C20:0 cer, C22:0 cer, C24:0 cer and C24:1 cer), Dihydroceramides (C16:0 DHcer, C18:0 DHcer, C18:1 DHcer, C20:0 DHcer, C22:0 DHcer, C24:0 DHcer,C24:1 DHcer) and precursors (Sphingosine, Sphinganine, 1-Deoxysphinganine,1-

Methyldeoxysphinganine, SB) were quantified in plasma by liquid chromatography tandem mass spectrometry (LC-MS/MS). In addition to samples, seven-point calibration curves and 3 levels of quality controls were made from pure standards in BSA 5%. Finally, reference plasma spiked with analytes at two different levels were prepared as additional QC samples.

Lipid chromatographic separation was performed on a UPLC I-Class system (Waters) equipped with an Acquity BEH C18, 100 x 2.1 mm, 1.7 µm column (Waters) heated at 60°C. Mobile phase were A: 0.1 % (V/V) formic acid in water and B: 0.1 % (V/V) formic acid in acetonitrile/ isopropanol (60/40). Flow rate was set à 0.5 ml/min and a gradient was applied as follows: 0min: 45% A, 2min: 45% A, 3min: 15% A, 13min: 0% A, 14min: 45% A, 16 min: 45%A Mass. Mass analysis was performed on an API 6500 system (Sciex) operating with an electrospray source in positive mode. General parameters were set as follows: curtain gas: N2 (35 PSI), Ion source gas 1: Air (50 PSI), Ion source gas 2: Air (50 PSI), ion source voltage: 5500 V, temperature: 300°C, collision gas: N2 (7). Scheduled multiple reaction monitoring (MRM) mode was used with a target scan time of 0.5s and an MRM detection window of 60s.

Data was acquired using Analyst 1.6.2 (Sciex) and data processing was performed with MultiQuant 3.0 (Sciex). Peak area of analyte and internal standard were determined by the MultiQuant 3.0 (Sciex) integration system. Analyte concentrations were determined using the internal standard method. The standard curves were generated from the peak area ratios of analyte/internal standard using linear regression analysis with 1/x2 weighting (except for C24 cer: quadratic regression analysis). Quantifications of analytes were accepted based on quality control samples. A tolerance of 25% and 30% was applied for accuracy and precision of QC samples and spiked plasma samples, respectively. All concentrations were reported in ng/mL. Internal standards used are listed in the Supplementary Table 16.

Analysis of shotgun lipidomics and targeted sphingolipid data

The statistical analyses of the shotgun lipidomics and targeted sphingolipid data sets were kept separate. Identical analysis steps were applied to the two data sets. Both sets had missing data values. Lipid species with ≥25% missing values across all available plasma samples were removed from the data set. This filtering resulted in 113 lipid species that were kept in the shotgun data set (523 were removed) and 14 in the targeted data set (4 were removed). For the lipids that remained in the data sets, missing values were imputed using a random forest approach, applying the function missForest from the R package missForest, with default parameters. In a next step, samples were filtered based on subject characteristics: individuals with bilirubin levels ≥100 µmol/l were removed before all analysis; moreover, individuals categorized as IGT with an HbA1c≥6% were excluded from the group comparisons in differential analysis, but they were retained in other analyses involving lipidomics data. In differential analysis, due to the limited number of available ND samples, the ND and the included IGT samples were combined into a single group for comparison with other sample groups, as described in the result section.

For differential analysis, linear models were applied, using the function lm from the R stats package. For each comparison between two sample groups, a linear model that included diabetes status as the main explanatory variable and age, sex and BMI as covariates was fitted to the data from the two groups. P values for diabetes status were adjusted across all included lipid species with the Benjamini-Hochberg method, separately for each comparison. In addition, ANCOVA results from the three groups T2D, T3cD and ND (as defined above) with the same covariates were computed, with *p*-value adjustment across all lipid species as well.

Integrative analysis of transcriptomics and lipidomics

Multiblock modeling

Consensus Orthogonal Partial Least Squares (OPLS) model was computed with the MATLAB 9 environment with combinations of toolboxes and in-house functions that are available at https://gitlab.unige.ch/Julien.Boccard/consensusopls. Modified RV-coefficients were computed with the publicly available MATLAB m-file[56]. KOPLS-DA was assessed with routines implemented in the KOPLS open source package[57]. Consensus OPLS modeling was performed on shotgun lipidomics, targeted sphingolipids and transcriptomics data tables, which were all autoscaled prior to the analysis. The Consensus OPLS model distinguishes variation of data that is correlated to Y response and those which is orthogonal to Y response. This eases the biological interpretation of results and enables the link between variation of variables and variation of the outcome while removing information coming from other sources of variation.

The model resulted in 3 components: 1 predictive latent variable and 2 orthogonal latent variables. The quality of the model was assessed by R^2 and Q^2 values, which define the portion of data variance explained by the model and the predictive ability of the model, respectively. The predictive component carried 11% of the total explained variance of global data (R^2X) and explained 51.7% of variation of HbA1c (R^2Y). This indicates that the model was able to explain a large part of variation of the response variable based on the different data matrices. The Q^2 value was computed by a K-fold cross validation (K=7), which led to a goodness of prediction of $Q^2 = 0.26$.

To ensure the validity of the model, a series of 1,000 permutation tests were carried out by mixing randomly the original Y response (HbA1c patient values). The true model Q2 value was clearly distinguished and statistically different from the random models distribution ($p<0.001$, mean=−0.1778, standard deviation (SD)=0.150, n=1,000). The variable relevance to explain the HbA1c variation was evaluated using the variable importance in projection (VIP) parameter,

which reflects the importance of variables both with respect to the response and to the projection quality. The most relevant features were selected using a VIP threshold > 1.2.

Proteomics

Sample Preparation

Pooled pancreatic islet cells with an approximate surface area of 80,000 μm^2 were collected via Laser Capture Microdissection (LCM) onto adhesive cap tubes. Isolates were reconstituted in a 20 µl lysis buffer (PreOmics, Germany) and transferred into PCR tubes[58]. Samples were boiled at 95°C for 1min to denature proteins and reduce and alkylate cysteines without shaking in a thermocycler (Eppendorf GmbH) followed by sonication at maximum power (Bioruptor, Diagenode, Belgium) for 10 cycles of 30sec sonication and 30sec cooldown each. Sample liquid was briefly spun down and boiled again for 10min without shaking. 20µl of 100mM TrisHCl pH 8.5 (1:1 v/v) and 20ng Trypsin/LysC were added to each sample, followed by overnight digestion at 30°C without shaking. The next day, 40µl 99% Isopropanol 5% Trifluoroacetic acid (TFA) (1:1 v/v) was added to the solution and mixed by sonication. Samples were then subjected to stage-tip cleanup via styrenedivinylbenzene reversed-phase sulfonate (SDB-RPS). The sample liquid was loaded on one 14-gauge stage-tip plug. Peptides were cleaned up with 2x200µl 99% Isopropanol 5% TFA and 2x200µl 99% ddH2O 5% TFA in an in-house made Stage-tip centrifuge at 2,000xg, followed by elution in 40µl 80% Acetonitrile, 5% Ammonia and dried at 45°C in a SpeedVac centrifuge (Eppendorf, Concentrator plus) according to the 'in-StageTip' protocol (PreOmics, Germany). Peptides were resuspended in 0.1% TFA, 2% ACN, 97.9% ddH2O.

Liquid chromatography and mass spectrometry (LC-MS)

LC-MS was performed with an EASY nanoLC 1200 (Thermo Fisher Scientific) coupled online to a trapped ion mobility spectrometry quadrupole time-of-flight mass spectrometer (timsTOF Pro,

Bruker Daltonik GmbH, Germany) via nano-electrospray ion source (Captive spray, Bruker Daltonik GmbH). Peptides were loaded on a 50cm in-house packed HPLC-column (75µm inner diameter packed with 1.9µm ReproSil-Pur C18-AQ silica beads, Dr. Maisch GmbH, Germany). Sample analytes were separated using a linear 120min gradient from 5-30% buffer B in 95min followed by an increase to 60% for 5min, and by a 5min wash at 95% buffer B at 300nl/min (Buffer A: 0.1% Formic Acid, 99.9% ddH2O; Buffer B: 0.1% Formic Acid, 80% CAN, 19.9% ddH2O). The column temperature was kept at 60°C by an in-house manufactured oven.

Mass spectrometry analysis was performed in a data-dependent PASEF mode with 1 MS1 survey TIMS-MS and 10 PASEF MS/MS scans per acquisition cycle. Ion accumulation and ramp time in the dual TIMS analyzer was set to 100ms each and we analyzed the ion mobility range from $1/K_0$ = 1.6 Vs cm^{-2} to 0.6 Vs cm^{-2}. Precursor ions for MS/MS analysis were isolated with 2Th windows for m/z<700 and 3Th for m/z>700 in a total m/z range of 100-1,700 by synchronizing quadrupole switching events with the precursor elution profile from the TIMS device. The collision energy was lowered linearly as a function of increasing mobility starting from 59 eV at $1/K_0$=1.6 VS cm^{-2} to 20 eV at $1/K_0$=0.6 Vs cm^{-2}. Singly charged precursor ions were excluded with a polygon filter (otof control, Bruker Daltonik GmbH). Precursors for MS/MS were picked at an intensity threshold of 2.500 a.u. and resequenced until reaching a 'target value' of 20,000 a.u taking into account a dynamic exclusion of 40sec elution[25].

Before MS analysis, the LC-MS setup was subjected to a rigorous quality control procedure. These criteria included protein- and peptide-identifications as well as general technical specifications like chromatography performance. If those thresholds were met (>5.500 protein groups, >38.000 peptides from 200 ng tryptic HeLa digest, chromatographic peak FWHM of <=9 sec and peak base-to-base width <=17 sec on a 120 min liquid chromatography gradient; Quantitative reproducibility across two subsequent QC runs with a Pearson correlation of >0.97 and coefficients of variation of <=10%) the project measurements were initiated. Furthermore,

we subject our instruments to a rigorous weekly maintenance procedure (Maintenance of the Liquid chromatography platform and re-calibration of the mass spectrometer) to ensure highest overall performance and reproducibility.

Proteomics raw file processing

Raw files were searched against the human Uniprot databases (UP000005640_9606.fa, UP000005640_9606_additional.fa) MaxQuant (Version 1.6.7), which extracts features from four-dimensional isotope patterns and associated MS/MS spectra[59]. False-discovery rates were controlled at 1% both on peptide spectral match (PSM) and protein level. Peptides with a minimum length of seven amino acids were considered for the search including N-terminal acetylation and methionine oxidation as variable modifications and cysteine carbamidomethylation as fixed modification, while limiting the maximum peptide mass to 4,600 Da. Enzyme specificity was set to trypsin cleaving c-terminal to arginine and lysine. A maximum of two missed cleavages were allowed. Maximum precursor and fragment ion mass tolerance were searched as default for TIMS-DDA data, while the main search peptide tolerance was set to 20ppm. The median absolute mass deviation for the data was 0.68ppm. Peptide identifications by MS/MS were transferred by matching four-dimensional isotope patterns between the runs with a 0.7-min retention-time match window and a 0.05 $1/K_0$ ion mobility window[60]. Label-free quantification was performed with the MaxLFQ algorithm and a minimum ratio count of 1[61].

Bioinformatic analysis

Bioinformatics analysis was performed in Perseus (version 1.6.7.0 and 1.5.5.0) and GraphPad Prism (version 8.2.1)[62]. Reverse database, contaminant, and only by site modification identifications were removed from the dataset. Data were grouped by analytical replicates and filtered to at least 70% data completeness in one group. Missing values were imputed from a

data table specific normal distribution estimate with a downshift of 1.8 and a width of 0.3 standard deviations after log_2-transformation of the data. To represent the data reproducibility and variability, a principal component analysis was performed on the median data of analytical replicate measurements of each individual. Clinically classified T2D and ND individuals were tested for differences in their mean by a two-sided Student's t-test with S0=0.1 and a Benjamini-Hochberg correction for multiple hypothesis testing at an FDR of 0.05 preserving grouping of each individuals analytical replicate measurements, and presented as volcano plot. We then normalized the data by row-wise z-scoring followed by hierarchical clustering using Euclidean as the distance parameter for column- and row-wise clustering. 1D gene ontology enrichments of clustered and systematically changed proteins were performed with regards to their cellular compartment and keywords assignment[30]. Log_2 transformed LFQ data were used for the calculation of intensity shifts of the enriched keyword or cellular compartment term for each of the displayed clusters. Total protein copy number estimation of the median LFQ intensities for patients clinically classified as non-diabetic and diabetic were calculated using the Perseus plugin 'Proteomic ruler'[29]. Median LFQ intensity values for all T2D and ND were calculated. We annotated protein groups for the leading protein ID with the human Uniprot fasta file (UP000005640_9606.fa) and estimated the protein copy number with the following settings: Averaging mode. 'All columns separately', Molecular masses: 'Average molecular mass', Detectability correction: 'Number of theoretical peptides', Scaling mode: 'Histone proteomic ruler', Ploidy: '2', Total cellular protein concentration: '200g/l'. Proteins were annotated with regards to their cellular compartment by gene ontology. We calculated the median protein copy number for the samples from T2D and ND PPP separately and multiplied it by its protein mass. To calculate the subcellular protein mass contribution, we calculated the protein mass proportion for the GOCC terms 'Nucleus', 'Mitochondrion', 'Cytoskeleton', 'Golgi apparatus', and 'Endoplasmic reticulum'. For calculating the organellar change between T2D and ND PPP, protein mass contributions of each organelle were normalized by its respective 'Nuclear part'

contribution. Chromosomal annotation of significantly changed proteins between T2D and ND PPP was identified via Ensembl ID. For transcriptome to proteome correlation, the gene intersection of both data sets was scaled to 1E6 units, followed by log10-transformation.

Antibody validation

Rabbit polyclonal anti-ALDOB antibody (Proteintech, Cat.No. 18065-1-AP) was tested for specificity by western blotting of protein extracts of *ALDOB*$^{-/-}$ MIN6 cells generated with a CRISPR/Cas9 system, as described[63]. Primary antibodies against ALDOB, ALDOA (Abnova, Cat. No. H00000226-M01) and gamma tubulin as loading control (Sigma Aldrich, Cat.No. T-6557) were diluted in 5% non-fat milk 1:2000, 1:1000 and 1:5000, respectively. The knock-out of *ALDOB* was verified by Sanger sequencing of the target locus.

Isolated mouse islet and cell line experiments

Mouse (C57BL/6J, db/db (BKS.Cg-Dock7^{m} +/+ Lepr^{db}/J) and db/+ (Charles River Laboratories), 3 animals/strain, male, age 13 weeks) islets were cultured for 1 day post isolation. Islet beta MIN6c4 (MIN6 clone 4, from Osaka University under Material License Agreement) and alpha αTC1-clone 6 (ATCC, CRL-2934) cell lines were harvested for RNA extraction using Qiagen RNeasy Mini Kit according to the manufacturer's instructions. After quality control, RNA samples were sequenced using the Illumina HiSeq 2000 platform and processed as previously described[51,64,65]. All animal experiments were done in accordance with the ethical approval of the Sanofi-Aventis Animal Welfare Office, Frankfurt/Main, Germany. The animals were housed at 20-24°C, by 45-65 % humidity setting in an artificial day / night (12hrs) rhythm.

Immunofluorescence microscopy

Immunofluorescence staining was done on formalin-fixed paraffin embedded 5µm thick sections of human pancreatic tissue. Acetylated histone H3 and H4 were detected in separate sections

using rabbit polyclonal antibodies (Merck Millipore Cat.No. 06-598 and 06-599, respectively, dilution 1:100). A mouse monoclonal anti-insulin antibody (Thermo Fisher Scientific Cat.No. 53-9769-82, dilution 1:200) was used for co-staining, to identify the beta cell areas. Images were acquired using a Nikon C2+ confocal microscope with a 60x oil immersion objective, with acquisition parameters normalized to a negative control sample.

Materials and methods references

44. Sturm, D. *et al.* Improved Protocol For Laser Microdissection Of Human Pancreatic Islets From Surgical Specimens. *J. Vis. Exp.* 2–7 (2013). doi:10.3791/50231

45. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal; Vol 17, No 1 Next Gener. Seq. Data Anal. - 10.14806/ej.17.1.200* (2011).

46. Wingett, S. W. & Andrews, S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Research* **7**, 1338 (2018).

47. Davis, M. P. A., van Dongen, S., Abreu-Goodger, C., Bartonicek, N. & Enright, A. J. Kraken: A set of tools for quality control and analysis of high-throughput sequence data. *Methods* **63**, 41–49 (2013).

48. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

49. Anders, S., Pyl, P. T. & Huber, W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

50. Wang, L., Wang, S. & Li, W. RSeQC: Quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185 (2012).

51. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, (2014).

52. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing

and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

53. Smyth, G. K. *et al.* RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Research* **5**, (2018).

54. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: An R package for comparing biological themes among gene clusters. *Omi. A J. Integr. Biol.* **16**, 284–287 (2012).

55. Surma, M. A. *et al.* An automated shotgun lipidomics platform for high throughput, comprehensive, and quantitative analysis of blood plasma intact lipids. *Eur. J. Lipid Sci. Technol.* **117**, 1540–1549 (2015).

56. Smilde, A. K., Kiers, H. A. L., Bijlsma, S., Rubingh, C. M. & Van Erk, M. J. Matrix correlations for high-dimensional data: The modified RV-coefficient. *Bioinformatics* **25**, 401–405 (2009).

57. Bylesjö, M., Rantalainen, M., Nicholson, J. K., Holmes, E. & Trygg, J. K-OPLS package: Kernel-based orthogonal projections to latent structures for prediction and interpretation in feature space. *BMC Bioinformatics* **9**, (2008).

58. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* **11**, 319–324 (2014).

59. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).

60. Prianichnikov, N. *et al.* Maxquant software for ion mobility enhanced shotgun proteomics. *Mol. Cell. Proteomics* **19**, 1058–1069 (2020).

61. Cox, J. *et al.* Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **13**, 2513–2526 (2014).

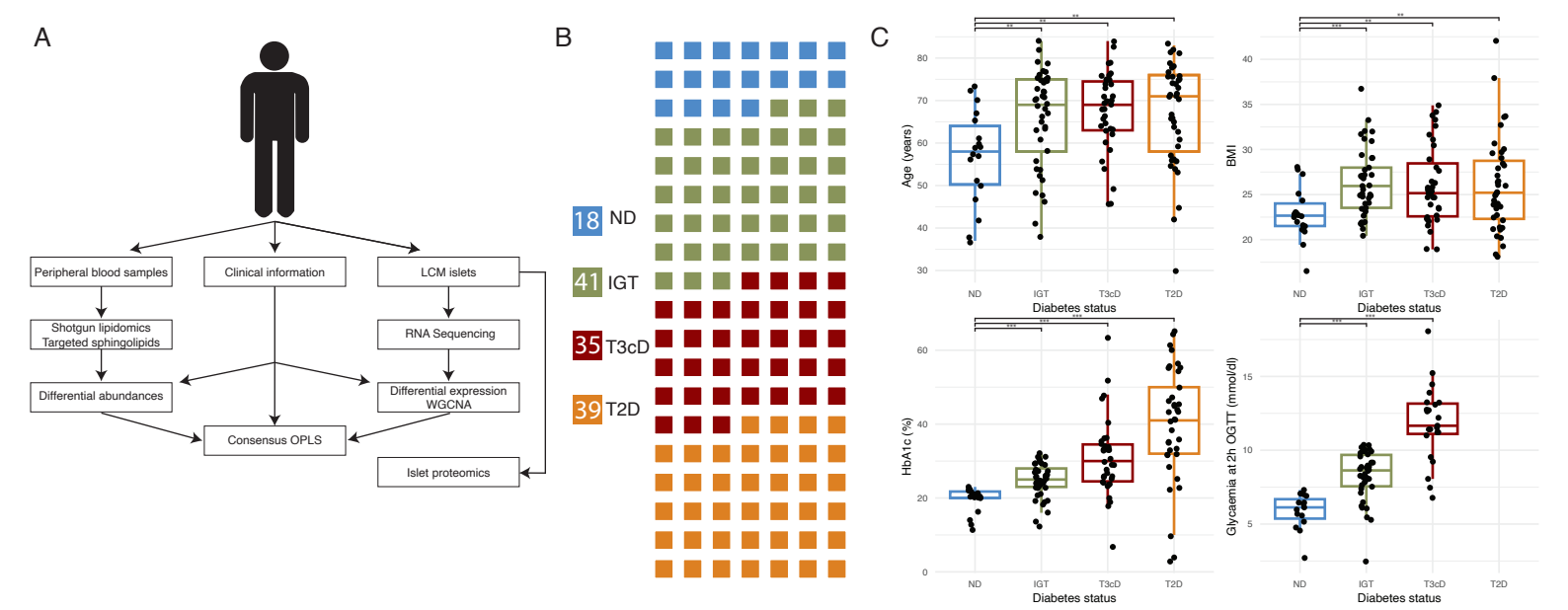
62. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of

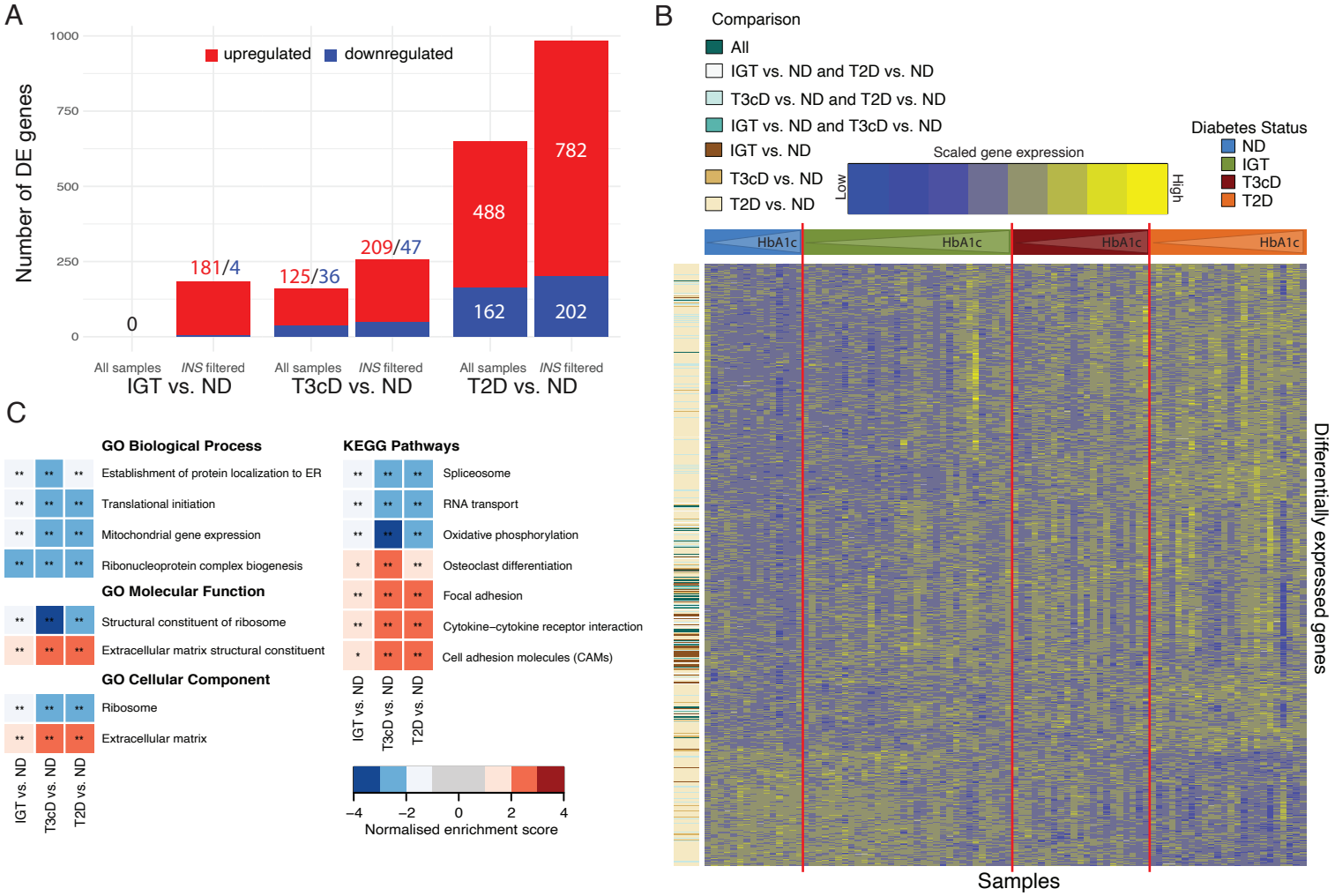
(prote)omics data. *Nature Methods* **13**, 731–740 (2016).

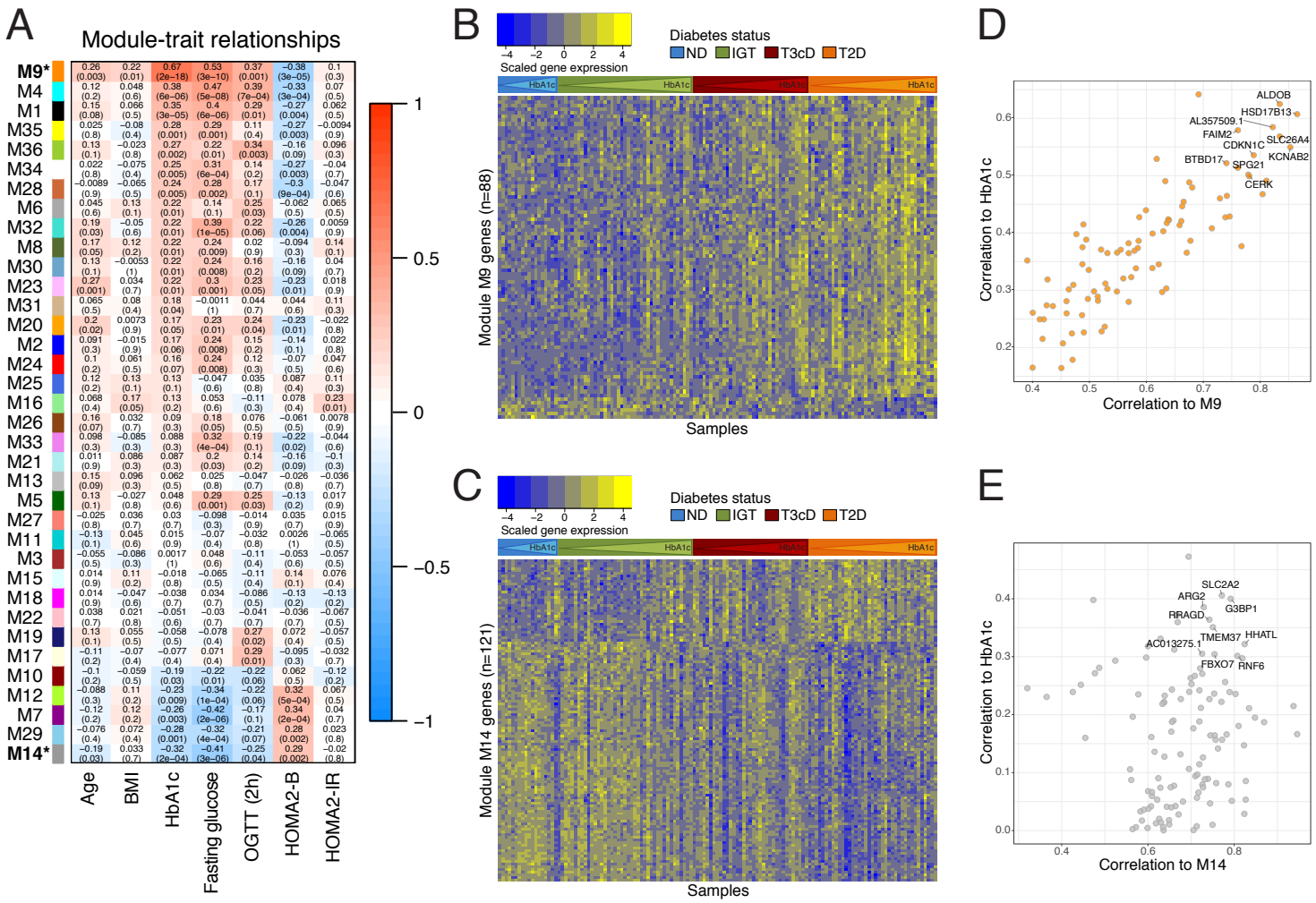
63. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).

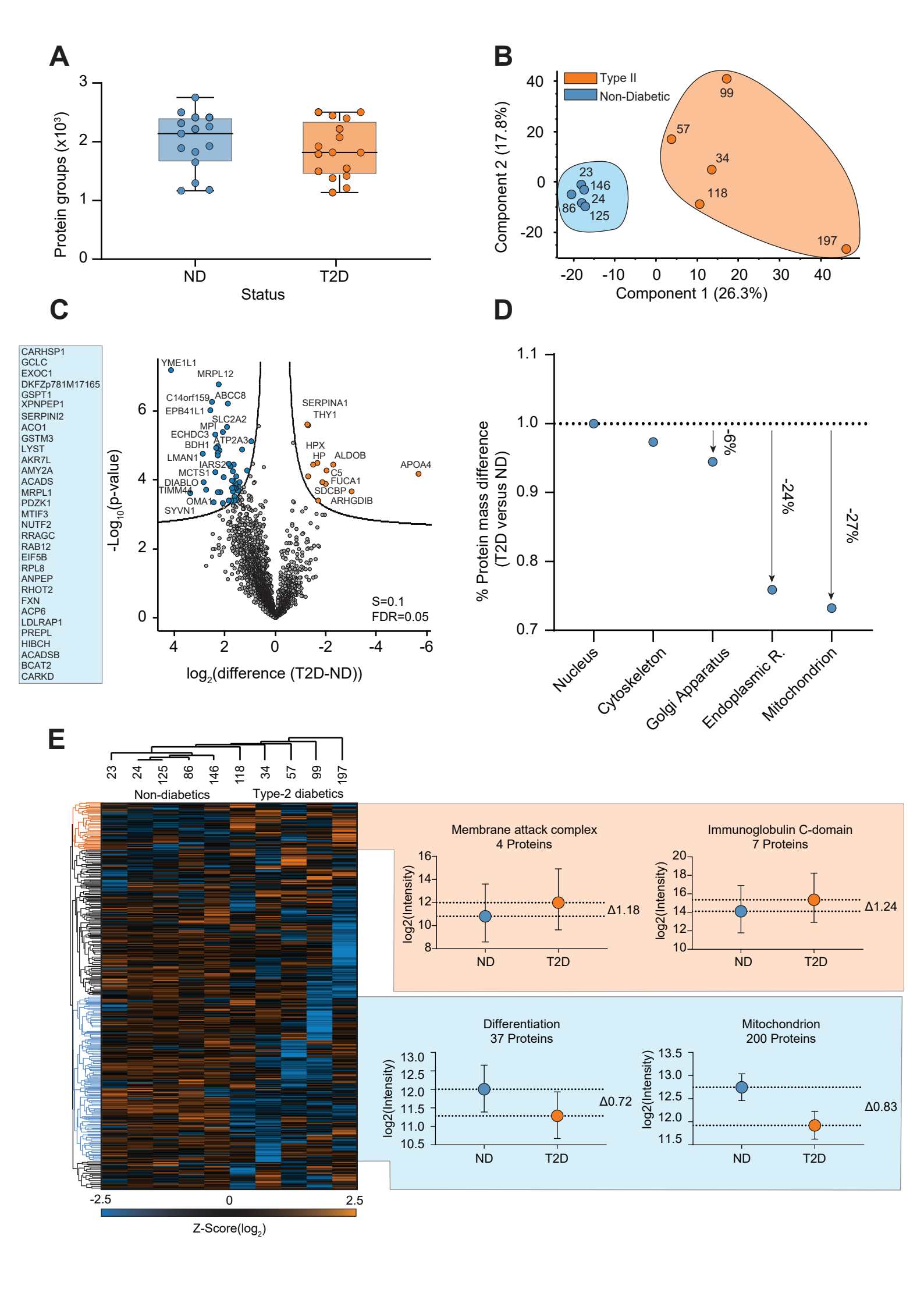
64. Hu, J., Ge, H., Newman, M. & Liu, K. OSA: A fast and accurate alignment tool for RNA-Seq. *Bioinformatics* **28**, 1933–1934 (2012).

65. Li, B., Ruotti, V., Stewart, R. M., Thomson, J. A. & Dewey, C. N. RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics* **26**, 493–500 (2009).

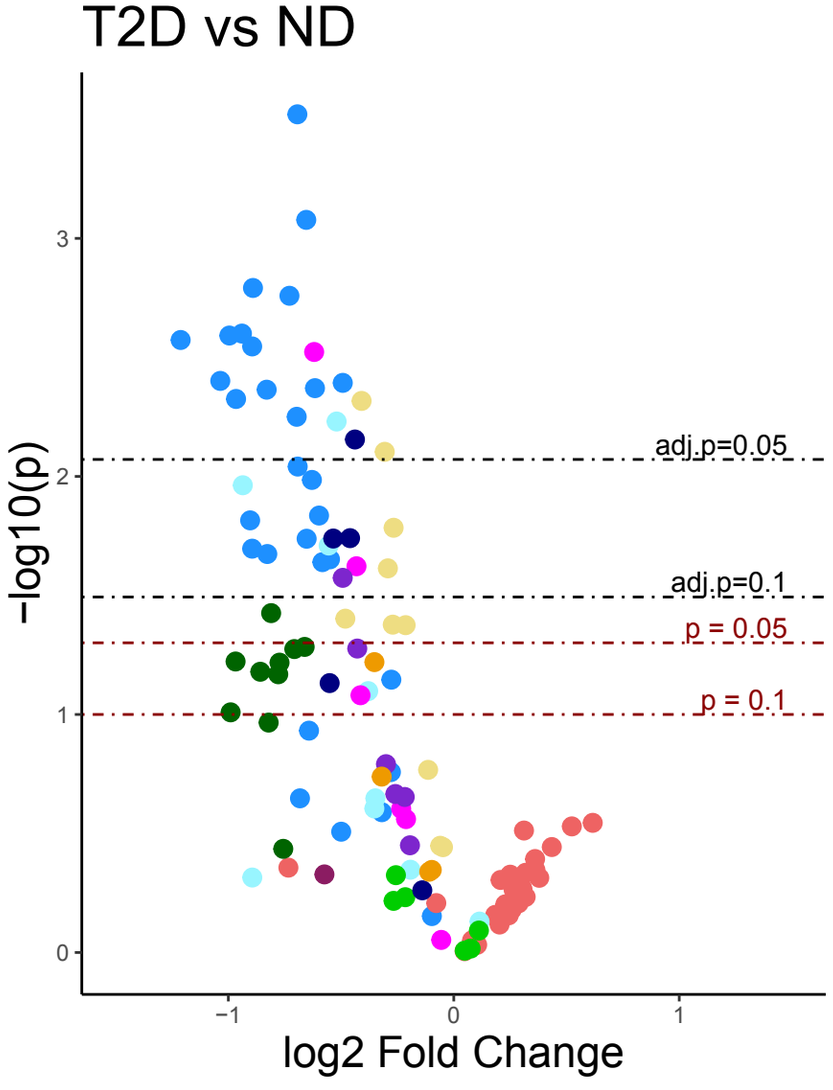

A
Peripheral blood samples
Clinical information
LCM islets
Shotgun lipidomics
Targeted sphingolipids
RNA Sequencing
Differential abundances
Differential expression
WGCNA
Consensus OPLS
Islet proteomics
B
18 ND
41 IGT
35 T3cD
39 T2D
C
Age (years)
BMI
HbA1c (%)
Glycaemia at 2h OGTT (mmol/l)
ND
IGT
T3cD
T2D
Diabetes status


A
upregulated
downregulated
Number of DE genes
1000
750
500
250
0
0
181/4
125/36
209/47
488
782
162
202
All samples
INS filtered
IGT vs. ND
T3cD vs. ND
T2D vs. ND
B
Comparison
All
IGT vs. ND and T2D vs. ND
T3cD vs. ND and T2D vs. ND
IGT vs. ND and T3cD vs. ND
IGT vs. ND
T3cD vs. ND
T2D vs. ND
Scaled gene expression
Low
High
Diabetes Status
ND
IGT
T3cD
T2D
HbA1c
Differentially expressed genes
Samples
C
GO Biological Process
Establishment of protein localization to ER
Translational initiation
Mitochondrial gene expression
Ribonucleoprotein complex biogenesis
GO Molecular Function
Structural constituent of ribosome
Extracellular matrix structural constituent
GO Cellular Component
Ribosome
Extracellular matrix
KEGG Pathways
Spliceosome
RNA transport
Oxidative phosphorylation
Osteoclast differentiation
Focal adhesion
Cytokine–cytokine receptor interaction
Cell adhesion molecules (CAMs)
IGT vs. ND
T3cD vs. ND
T2D vs. ND
−4
−2
0
2
4
Normalised enrichment score
D
INS
Acetyl-H3
Merge
INS
Acetyl-H4
Merge
ND
T2D

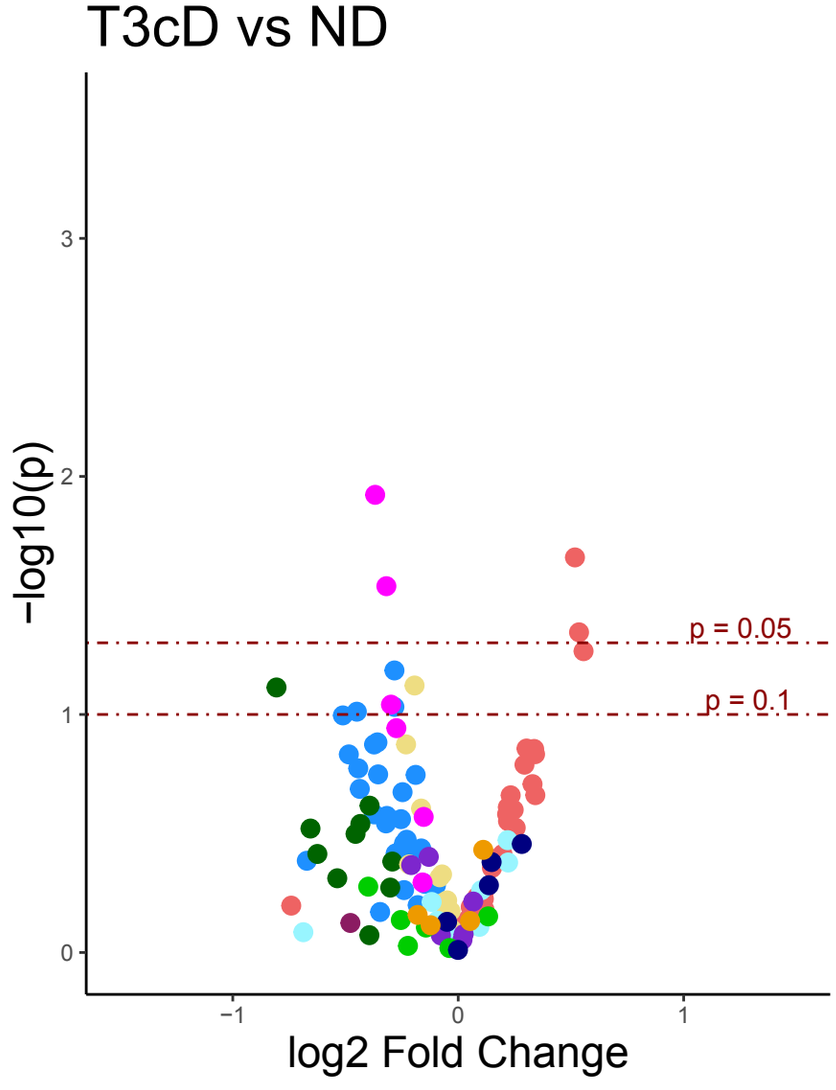

A
Module-trait relationships
M9*
M4
M1
M35
M36
M34
M28
M6
M32
M8
M30
M23
M31
M20
M2
M24
M25
M16
M26
M33
M21
M13
M5
M27
M11
M3
M15
M18
M22
M19
M17
M10
M12
M7
M29
M14*
Age
BMI
HbA1c
Fasting glucose
OGTT (2h)
HOMA2-B
HOMA2-IR
1
0.5
0
−0.5
−1
B
Scaled gene expression
Diabetes status
ND IGT T3cD T2D
HbA1c
Module M9 genes (n=88)
Samples
C
Module M14 genes (n=121)
D
Correlation to HbA1c
Correlation to M9
ALDOB
HSD17B13
AL357509.1
FAIM2
SLC26A4
CDKN1C
KCNAB2
BTBD17
SPG21
CERK
E
Correlation to M14
SLC2A2
ARG2
G3BP1
RRAGD
TMEM37
HHATL
AC013275.1
FBXO7
RNF6

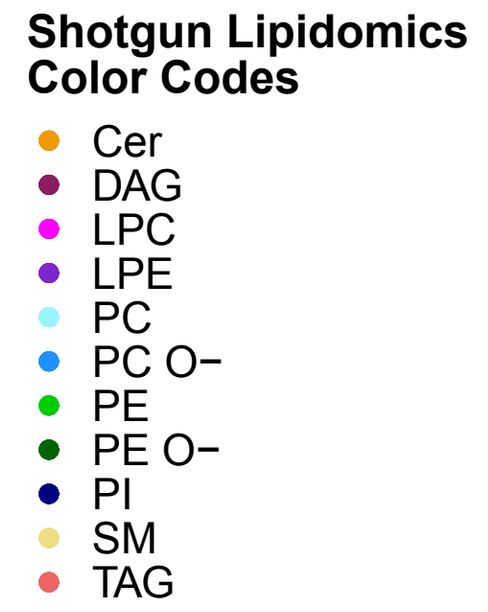

A
Protein groups (x10³)
Status
ND
T2D
B
Type II
Non-Diabetic
Component 2 (17.8%)
Component 1 (26.3%)
C
CARHSP1
GCLC
EXOC1
DKFZp781M17165
GSPT1
XPNPEP1
SERPINI2
ACO1
GSTM3
LYST
AKR7L
AMY2A
ACADS
MRPL1
PDZK1
MTIF3
NUTF2
RRAGC
RAB12
EIF5B
RPL8
ANPEP
RHOT2
FXN
ACP6
LDLRAP1
PREPL
HIBCH
ACADSB
BCAT2
CARKD
-Log10(p-value)
log2(difference (T2D-ND))
S=0.1
FDR=0.05
D
% Protein mass difference (T2D versus ND)
-6%
-24%
-27%
Nucleus
Cytoskeleton
Golgi Apparatus
Endoplasmic R.
Mitochondrion
E
Non-diabetics
Type-2 diabetics
Z-Score(log2)
Membrane attack complex
4 Proteins
Δ1.18
Immunoglobulin C-domain
7 Proteins
Δ1.24
Differentiation
37 Proteins
Δ0.72
Mitochondrion
200 Proteins
Δ0.83
log2(Intensity)
ND
T2D

A)

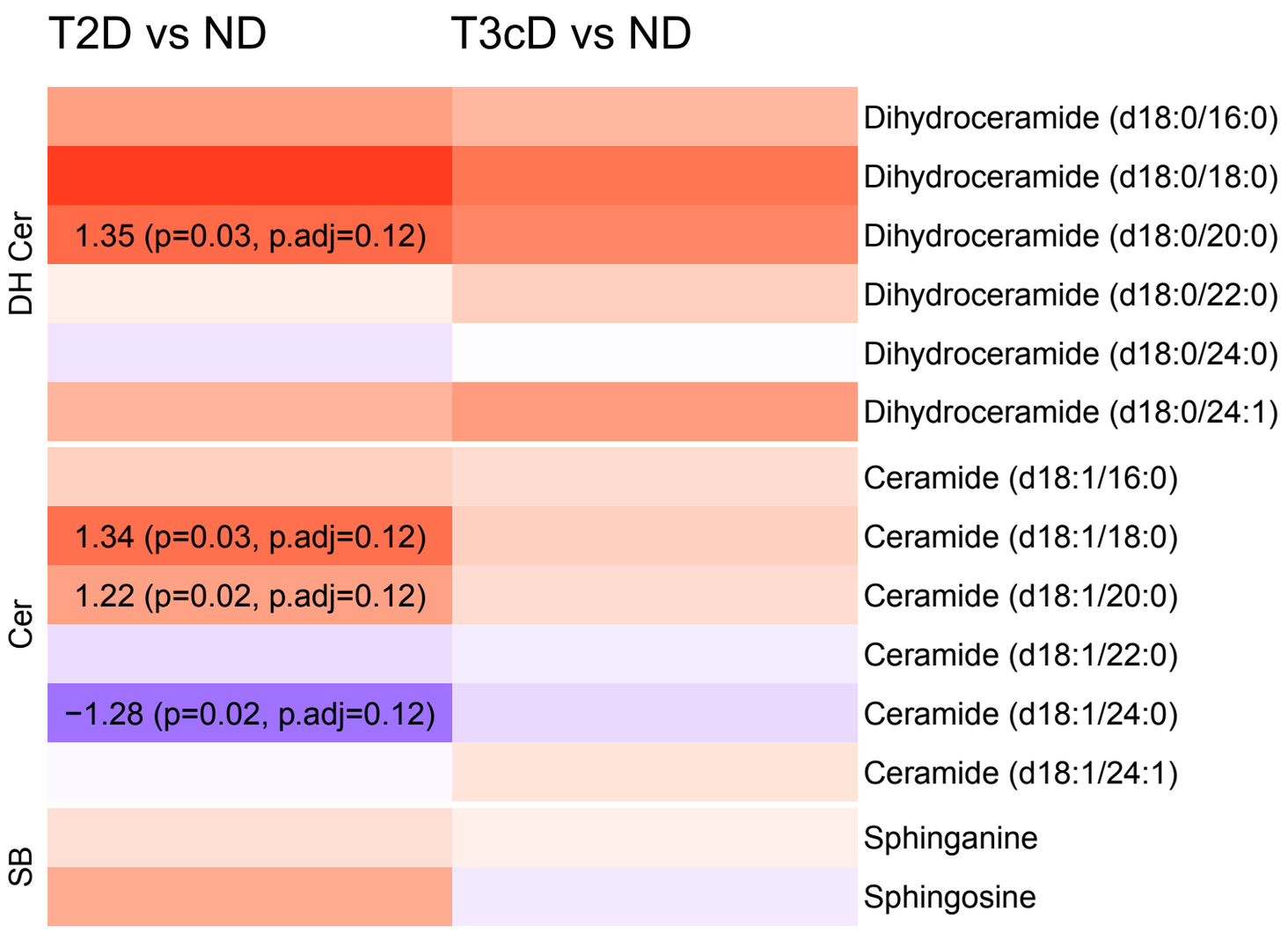

T2D vs ND
−log10(p)
log2 Fold Change
adj.p=0.05
adj.p=0.1
p = 0.05
p = 0.1
0
1
2
3
−1
0
1

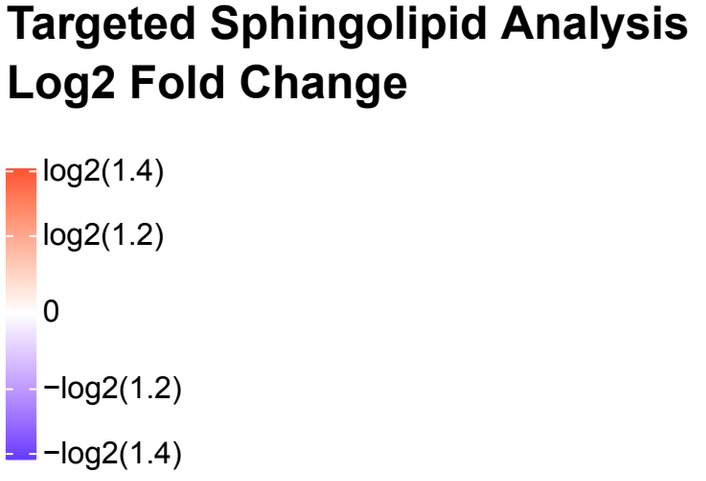
B)

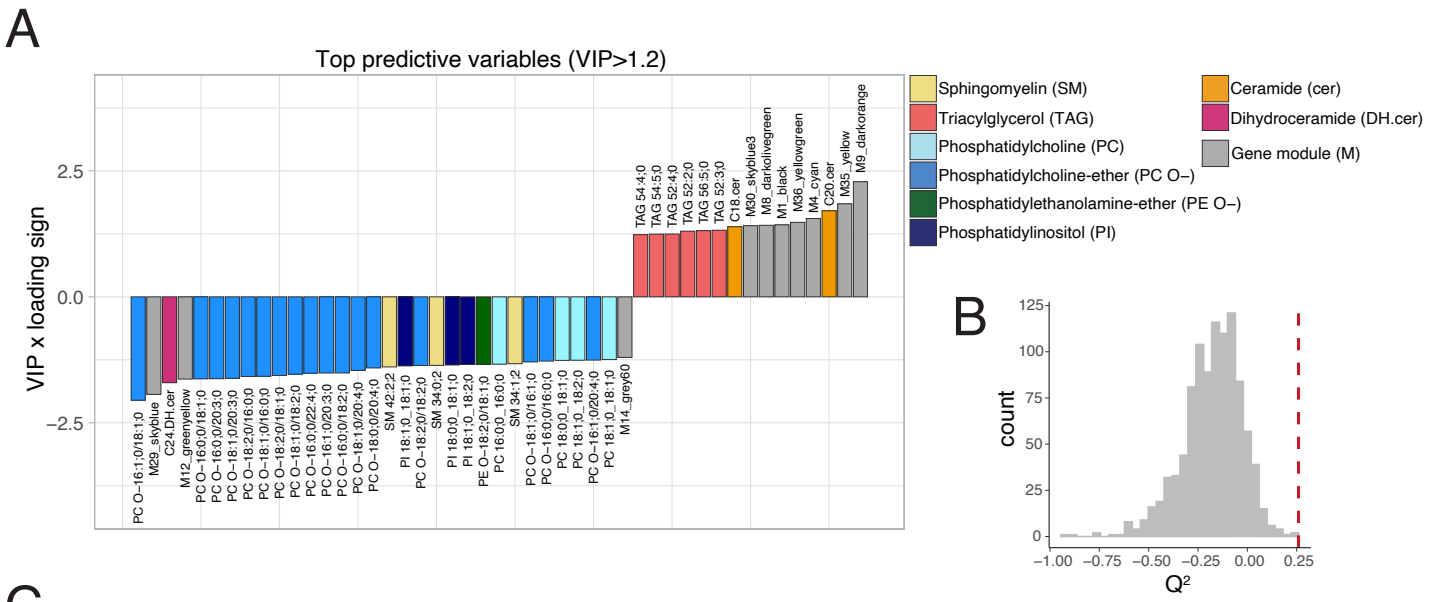

T3cD vs ND
−log10(p)
log2 Fold Change
p = 0.05
p = 0.1
0
1
2
3
−1
0
1

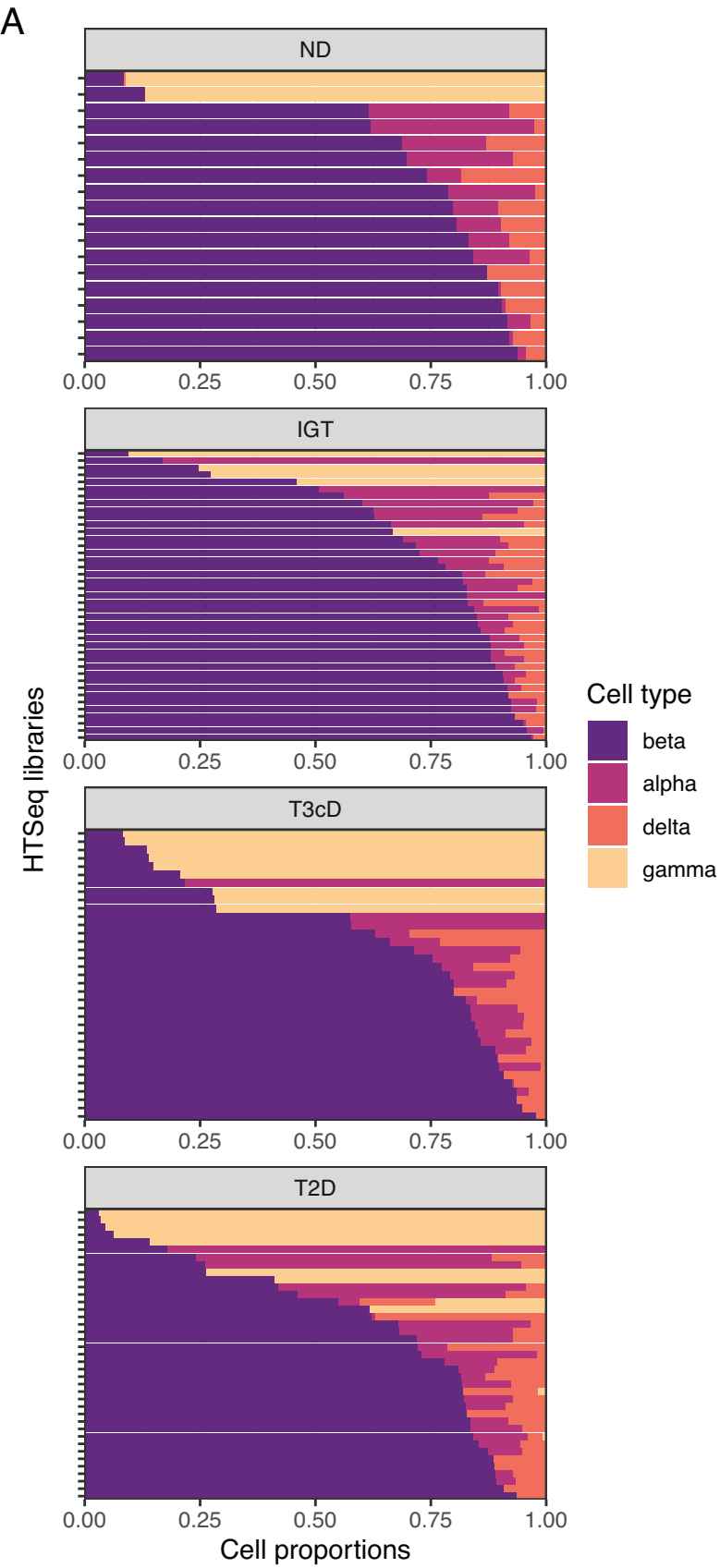

Shotgun Lipidomics
Color Codes
Cer
DAG
LPC
LPE
PC
PC O−
PE
PE O−
PI
SM
TAG

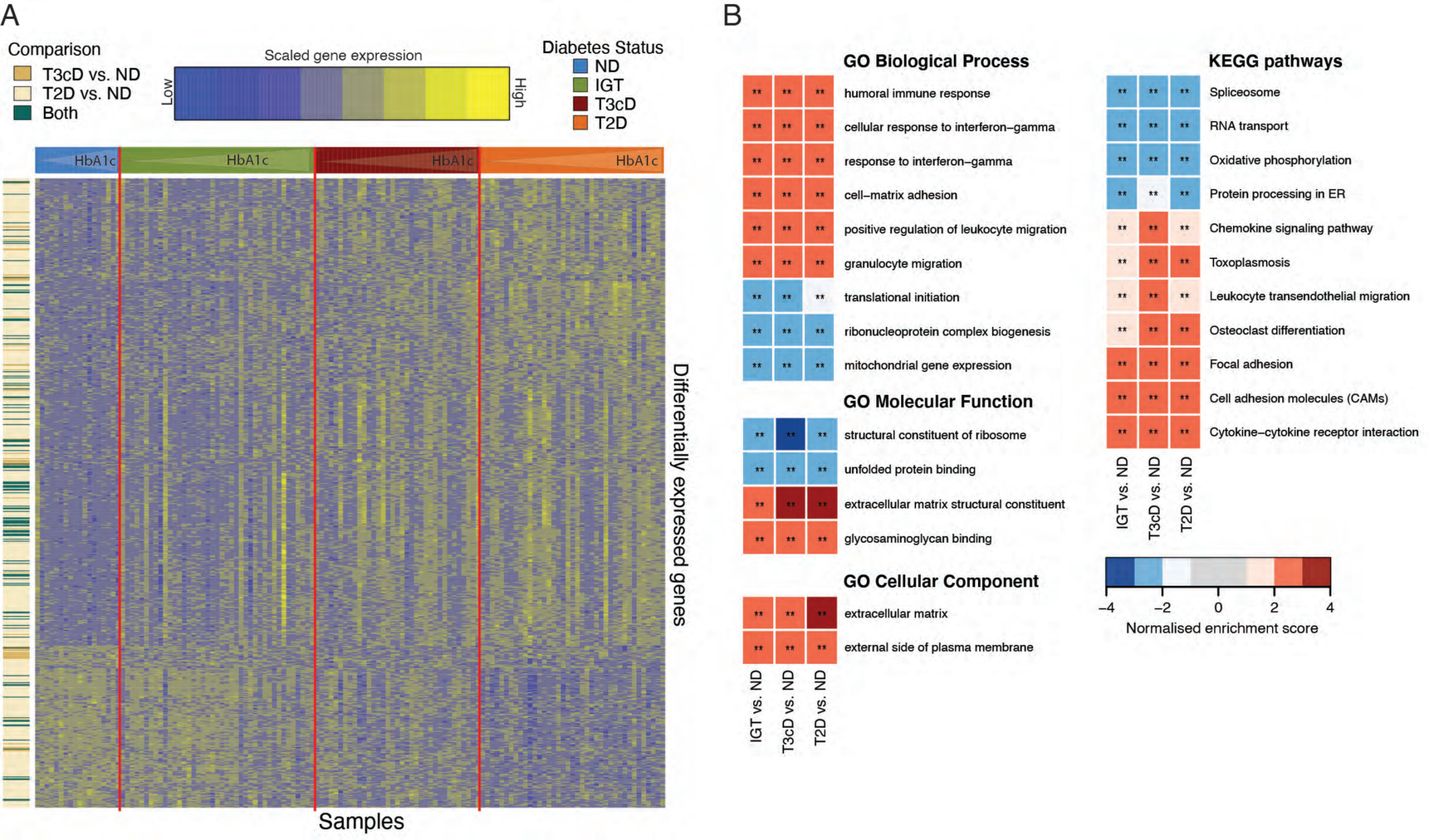
C)

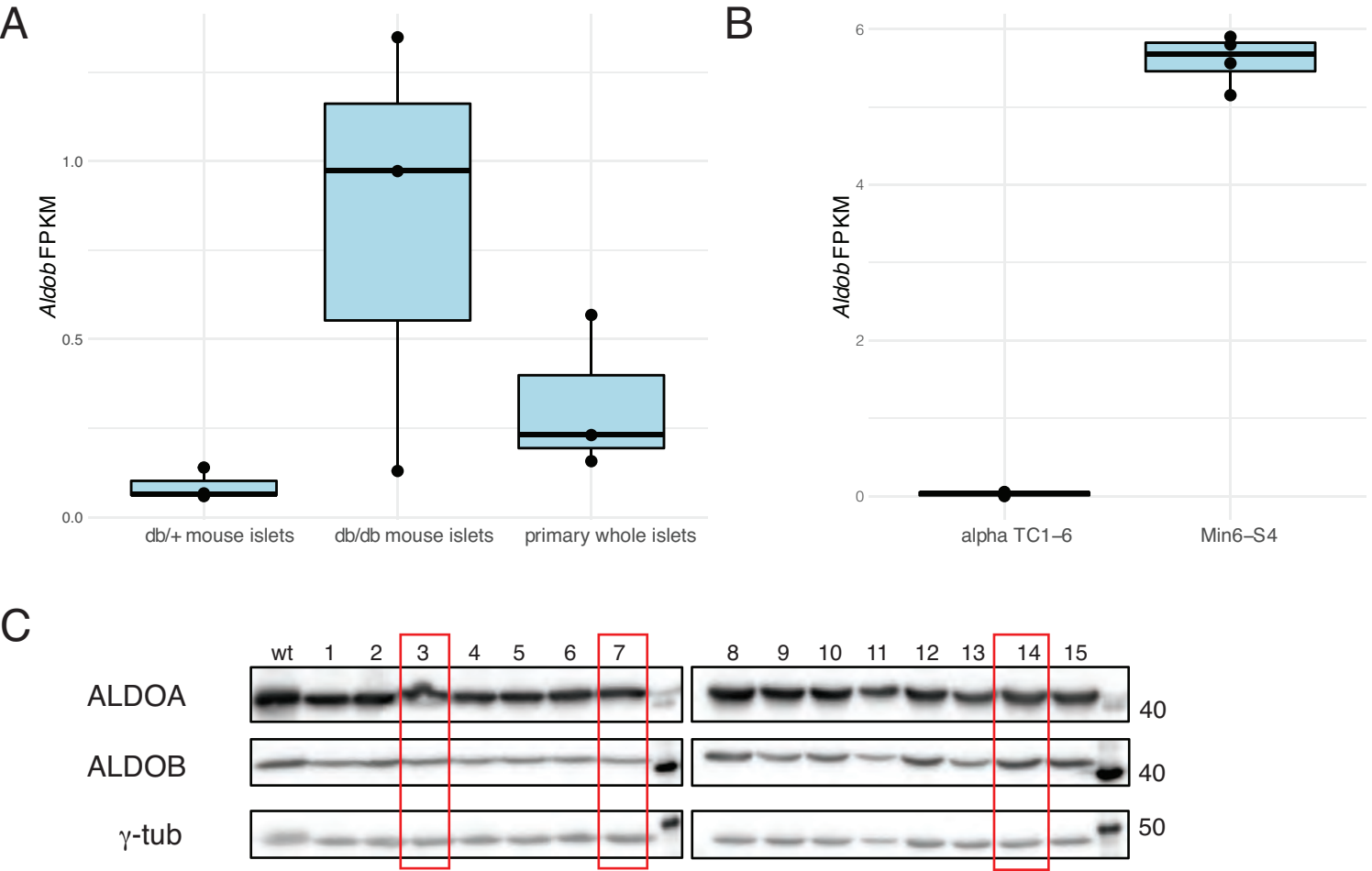

T2D vs ND
T3cD vs ND
DH Cer
Cer
SB
1.35 (p=0.03, p.adj=0.12)
1.34 (p=0.03, p.adj=0.12)
1.22 (p=0.02, p.adj=0.12)
−1.28 (p=0.02, p.adj=0.12)
Dihydroceramide (d18:0/16:0)
Dihydroceramide (d18:0/18:0)
Dihydroceramide (d18:0/20:0)
Dihydroceramide (d18:0/22:0)
Dihydroceramide (d18:0/24:0)
Dihydroceramide (d18:0/24:1)
Ceramide (d18:1/16:0)
Ceramide (d18:1/18:0)
Ceramide (d18:1/20:0)
Ceramide (d18:1/22:0)
Ceramide (d18:1/24:0)
Ceramide (d18:1/24:1)
Sphinganine
Sphingosine

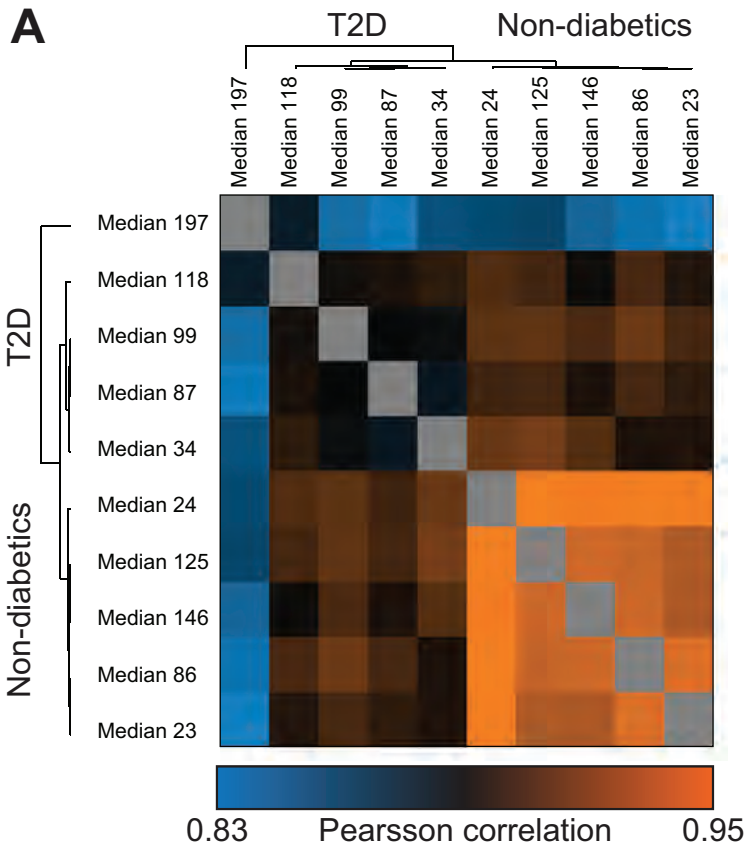

Targeted Sphingolipid Analysis
Log2 Fold Change
log2(1.4)
log2(1.2)
0
−log2(1.2)
−log2(1.4)

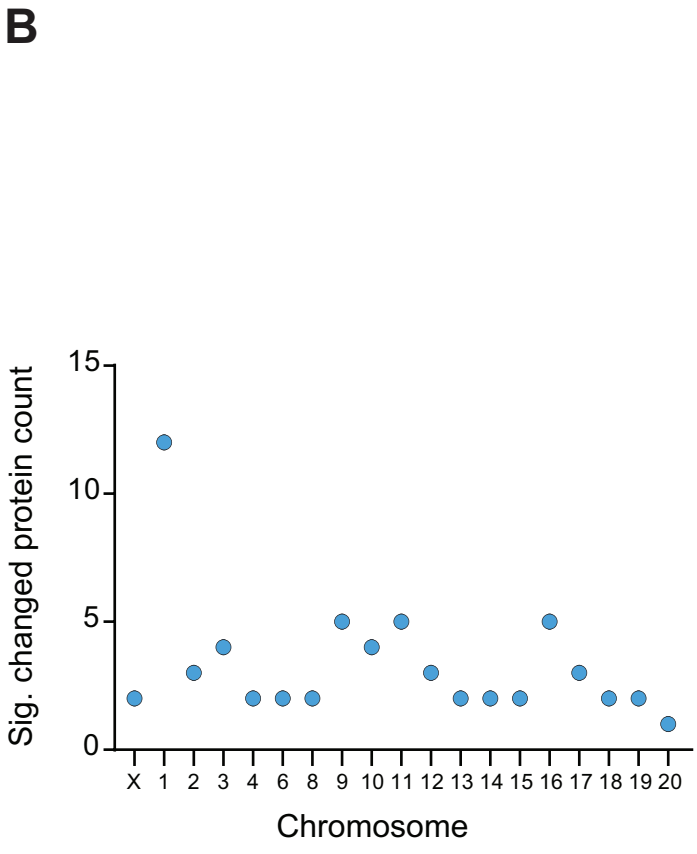

A
Top predictive variables (VIP>1.2)
VIP x loading sign
2.5
0.0
−2.5
PC O−16:1;0/18:1;0
M29_skyblue
C24:DH.cer
M12_greenyellow
PC O−16:0;0/18:1;0
PC O−16:0;0/20:3;0
PC O−18:1;0/20:3;0
PC O−18:2;0/16:0;0
PC O−18:1;0/16:0;0
PC O−18:2;0/18:1;0
PC O−18:1;0/18:2;0
PC O−16:0;0/22:4;0
PC O−16:1;0/20:3;0
PC O−16:0;0/18:2;0
PC O−18:1;0/20:4;0
PC O−18:0;0/20:4;0
SM 42:2;2
PI 18:1;0_18:1;0
PC O−18:2;0/18:2;0
SM 34:0;2
PI 18:0;0_18:1;0
PI 18:1;0_18:2;0
PE O−18:2;0/18:1;0
PC 16:0;0_16:0;0
SM 34:1;2
PC O−18:1;0/16:1;0
PC O−16:0;0/16:0;0
PC 18:0;0_18:1;0
PC 18:1;0_18:2;0
PC O−16:1;0/20:4;0
PC 18:1;0_18:1;0
M14_grey60
TAG 54:4;0
TAG 54:5;0
TAG 52:4;0
TAG 52:2;0
TAG 56:5;0
TAG 52:3;0
C18.cer
M30_skyblue3
M8_darkolivegreen
M1_black
M36_yellowgreen
M4_cyan
C20.cer
M35_yellow
M9_darkorange
Sphingomyelin (SM)
Triacylglycerol (TAG)
Phosphatidylcholine (PC)
Phosphatidylcholine-ether (PC O−)
Phosphatidylethanolamine-ether (PE O−)
Phosphatidylinositol (PI)
Ceramide (cer)
Dihydroceramide (DH.cer)
Gene module (M)
B
count
125
100
75
50
25
0
−1.00
−0.75
−0.50
−0.25
0.00
0.25
Q²
C
Modules M12+M14+M29
Metabolic reactions
Carbon metabolism
ENO2
CS
ENO1
Sphingolipid metabolism
SGPL1
RNA degradation
Biosynthesis of amino acids
ARG2
Glycolysis / Gluconeogenesis
G6PC2
Circadian rhythm
CTSV
Spliceosome
PPIE
SF3B3
PRKAG1
Lysosome
HNRNPC
GNPTG
LAPTM4A
AMPK signaling pathway
M6PR
AP1M2
PFKFB2
SCD5
mRNA surveillance pathway
CREB3L2
CSTF2T
Huntington disease
Viral carcinogenesis
PSMC3
MAPKAPK2
POLR2C
MRPS18B
TUBA4A
Bacterial invasion of epithelial cells
ELMO2
Gene Node
Size
Color
Correlation to HbA1c
low
high
DE in T2D vs ND (padj < 0.05)
UP
DOWN
Modules M1+M4+M8+M9+M30+M35+M36
Cell signaling
NF-kappa B signaling pathway
Toll-like receptor signaling pathway
EDARADD
NOD-like receptor signaling pathway
IL-17 signaling pathway
TXNIP
TGF-beta signaling pathway
CARD16
CCL21
TAB3
INHBA
T cell receptor signaling pathway
Cytokine-cytokine receptor interaction
HSP90AB1
Chemokine signaling pathway
Malaria
TNF signaling pathway
LRP1
B cell receptor signaling pathway
AGE-RAGE signaling pathway in diabetic complications
INPP5D
PRKCB
MMP2
MAPK signaling pathway
Insulin resistance
HSPA8
SLC2A1
PTPRC
CSF2RB
HIF-1 signaling pathway
FAS
CASP10
Prostate cancer
ALDOB
PFKFB3
PGK1
Cell adhesion molecules (CAMs)
Primary immunodeficiency
VCAN
SDC3
CDKN1A
ADA
CD3D
PDCD1LG2
Apoptosis
Th17 cell differentiation
BCL2
Hematopoietic cell lineage
VAV2
Transcriptional misregulation in cancer
CD37
Th1 and Th2 cell differentiation
RUNX1
CD5
Natural killer cell mediated cytotoxicity
Antigen processing and presentation
HLA-E
HLA-A
NGFR
HLA-C
PRF1
TUBA1B
HLA-DMA
HLA-F
Cell-matrix interaction
Type I diabetes mellitus
HLA-DQA2
HLA-B
Immune response
PI3K-Akt signaling pathway
Phagosome
GAD1
ECM-receptor interaction
Focal adhesion
C1R
PARVG
LAMB2
COL6A1
C3
Regulation of actin cytoskeleton
COL4A5
Complement and coagulation cascades
ACTG1
Protein digestion and absorption
C1QA
C1QC
COL5A1
COL5A2
Platelet activation
Hippo signaling pathway
CFD
C2
COL14A1
FERMT3
WNT3

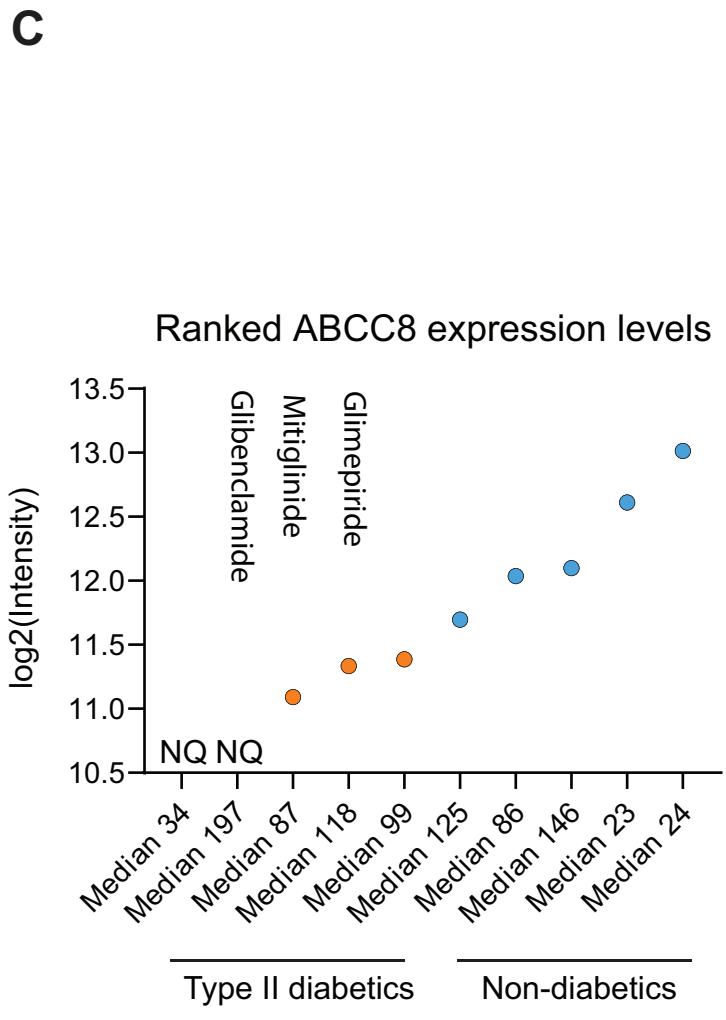

A
ND
IGT
T3cD
T2D
HTSeq libraries
0.00
0.25
0.50
0.75
1.00
Cell proportions
Cell type
beta
alpha
delta
gamma
B
Beta−cells
INS
All other
Alpha cells
GCG
TTR
Delta cells
Gamma cells
PPY
Most expressed gene in the sample
Cell proportions

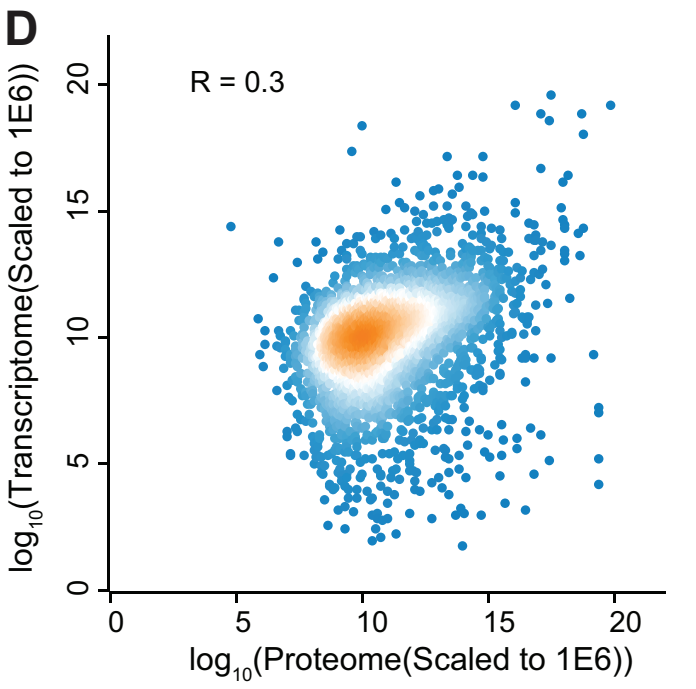

A
Comparison
T3cD vs. ND
T2D vs. ND
Both
Scaled gene expression
Low
High
Diabetes Status
ND
IGT
T3cD
T2D
HbA1c
HbA1c
HbA1c
HbA1c
Differentially expressed genes
Samples
B
GO Biological Process
humoral immune response
cellular response to interferon−gamma
response to interferon−gamma
cell−matrix adhesion
positive regulation of leukocyte migration
granulocyte migration
translational initiation
ribonucleoprotein complex biogenesis
mitochondrial gene expression
GO Molecular Function
structural constituent of ribosome
unfolded protein binding
extracellular matrix structural constituent
glycosaminoglycan binding
GO Cellular Component
extracellular matrix
external side of plasma membrane
IGT vs. ND
T3cD vs. ND
T2D vs. ND
KEGG pathways
Spliceosome
RNA transport
Oxidative phosphorylation
Protein processing in ER
Chemokine signaling pathway
Toxoplasmosis
Leukocyte transendothelial migration
Osteoclast differentiation
Focal adhesion
Cell adhesion molecules (CAMs)
Cytokine−cytokine receptor interaction
IGT vs. ND
T3cD vs. ND
T2D vs. ND
−4
−2
0
2
4
Normalised enrichment score

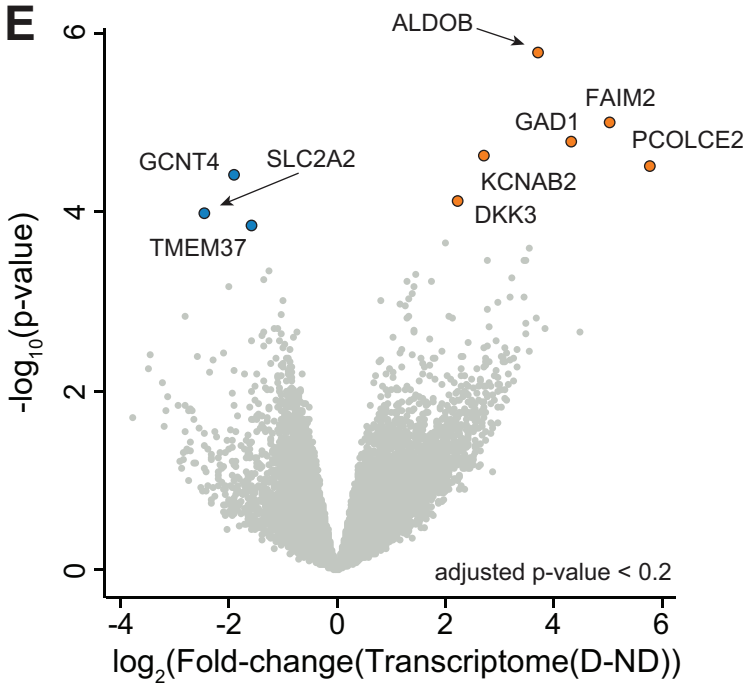

A
Aldob FPKM
1.0
0.5
0.0
db/+ mouse islets
db/db mouse islets
primary whole islets
B
Aldob FPKM
6
4
2
0
alpha TC1–6
Min6–S4
C
wt 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
ALDOA
ALDOB
γ-tub
40
40
50


A
T2D
Non-diabetics
Median 197
Median 118
Median 99
Median 87
Median 34
Median 24
Median 125
Median 146
Median 86
Median 23
T2D
Non-diabetics
0.83
Pearsson correlation
0.95


B
Sig. changed protein count
15
10
5
0
X 1 2 3 4 6 8 9 10 11 12 13 14 15 16 17 18 19 20
Chromosome


C
Ranked ABCC8 expression levels
log2(Intensity)
13.5
13.0
12.5
12.0
11.5
11.0
10.5
Glibenclamide
Mitiglinide
Glimepiride
NQ NQ
Median 34
Median 197
Median 87
Median 118
Median 99
Median 125
Median 86
Median 146
Median 23
Median 24
Type II diabetics
Non-diabetics


D
R = 0.3
log10(Transcriptome(Scaled to 1E6))
log10(Proteome(Scaled to 1E6))
0
5
10
15
20


E
ALDOB
FAIM2
GAD1
PCOLCE2
GCNT4
SLC2A2
KCNAB2
DKK3
TMEM37
-log10(p-value)
adjusted p-value < 0.2
log2(Fold-change(Transcriptome(D-ND))
-4
-2
0
2
4
6