Leonore Wigger, Marko Barovic, Andreas-David Brunner, Flavia Marzetta, Eyke Schöniger, Florence Mehl, Nicole Kipke, Daniela Friedland, Frederic Burdet et. al.

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

SLUB
Wir führen Wissen.

TECHNISCHE UNIVERSITÄT DRESDEN

Qucosa
Quality Content of Saxony

19

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

23

24 Leonore Wigger[1*], Marko Barovic[2,3,4*], Andreas-David Brunner[5*], Flavia Marzetta[1], Eyke

25 Schöniger[2,3,4], Florence Mehl[1], Nicole Kipke[2,3,4], Daniela Friedland[2,3,4], Frederic Burdet[1], Camille

26 Kessler[1], Mathias Lesche[6], Bernard Thorens[7], Ezio Bonifacio[3,4,8], Cristina Legido-Quigley[9,10],

27 Pierre Barbier Saint Hilaire[11], Philippe Delerive[12], Andreas Dahl[6], Christian Klose[13], Mathias J

28 Gerl[13], Kai Simons[13], Daniela Aust[14,15], Jürgen Weitz[16], Marius Distler[16], Anke M Schulte[17],

29 Matthias Mann[5#], Mark Ibberson[1#], Michele Solimena[2,3,4#]

30

31 [1]Vital-IT Group, SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland; [2]Department of

32 Molecular Diabetology, University Hospital and Faculty of Medicine, TU Dresden; [3]Paul

33 Langerhans Institute Dresden (PLID), Helmholtz Center Munich, University Hospital and Faculty

34 of Medicine, TU Dresden; [4]German Center for Diabetes Research (DZD e.V.), Neuherberg,

35 Germany; [5]Max Planck Institute of Biochemistry, Martinsried, Germany; [6]DRESDEN-

36 *concept* Genome Center, c/o Center for Molecular and Cellular Bioengineering, Technische

37 Universität Dresden, Germany; [7]Center for Integrative Genomics, University of Lausanne,

38 Lausanne Switzerland; [8]Center for Regenerative Therapies Dresden, Faculty of Medicine and

39 Center for Molecular and Cellular Bioengineering,Technische Universität Dresden, Dresden,

40 Germany; [9]Steno Diabetes Center Copenhagen, Gentofte, Denmark; [10]King's College London,

41 London, UK; 11DMPK Center, Technologie Servier, Orléans, France [12]Institut de Recherches

42 Servier, Pôle d'Innovation Thérapeutique Métabolisme, Suresnes, France; [13]Lipotype GmbH,

43 Dresden, Germany; [14]Department of Pathology, Medical Faculty, University Hospital Carl

44   Gustav Carus, Technische Universität Dresden, Dresden, Germany; [15]NCT Biobank Dresden,

45   University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany;

46   [16]Department of Visceral, Thoracic and Vascular Surgery, University Hospital Carl Gustav

47   Carus, Medical Faculty, Technische Universität Dresden, Dresden, Germany; [17]Sanofi-Aventis

48   Deutschland GmbH, Diabetes Research, Industriepark Höchst, 65926 Frankfurt am Main,

49   Germany.

50

51   * These authors contributed equally

52   # These authors jointly supervised this work

53

54   Corresponding authors:

55   Matthias Mann - mmann@biochem.mpg.de

56   Mark Ibberson - Mark.Ibberson@sib.swiss

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

58

59 Abstract

60

61 Existing studies do not sufficiently describe the molecular changes of pancreatic islet beta cells

62 leading to their deficient insulin secretion in type 2 diabetes (T2D). Here we address this

63 deficiency with a comprehensive multi-omics analysis of metabolically profiled

64 pancreatectomized living human donors stratified along the glycemic continuum from

65 normoglycemia to T2D. Islet pools isolated from surgical samples by laser-capture

66 microdissection had remarkably heterogeneous transcriptomic and proteomic profiles in

67 diabetics, but not in non-diabetic controls. Transcriptomics analysis of this unique cohort

68 revealed islet genes already differentially regulated in prediabetic individuals with impaired

69 glucose tolerance. Our findings demonstrate a progressive but disharmonic remodeling of

70 mature beta cells, challenging current hypotheses of linear trajectories toward precursor or

71 trans-differentiation stages in T2D. Further, integration of islet transcriptomics and pre-operative

72 blood plasma lipidomics data enabled us to define the relative importance of gene co-

73 expression modules and lipids positively or negatively associated with HbA1c levels, pointing to

74 potential prognostic markers.

75 # Abstract

76 Most research on human pancreatic islets is conducted on samples obtained from

77 normoglycemic or diseased brain dead donors and thus cannot accurately describe the

78 molecular changes of pancreatic islet beta cells as they progress towards a state of deficient

79 insulin secretion in type 2 diabetes (T2D). Here, we conduct a comprehensive multi-omics

80 analysis of pancreatic islets obtained from metabolically profiled pancreatectomized living

81 human donors stratified along the glycemic continuum, from normoglycemia to T2D. We find

82 that islet pools isolated from surgical samples by laser-capture microdissection display

83 remarkably more heterogeneous transcriptomic and proteomic profiles in patients with diabetes

84 than in non-diabetic controls. The differential regulation of islet gene expression is already

85 observed in prediabetic individuals with impaired glucose tolerance. Our findings demonstrate a

86 progressive, but disharmonic, remodeling of mature beta cells, challenging current hypotheses

87 of linear trajectories toward precursor or trans-differentiation stages in T2D. Furthermore,

88 through integration of islet transcriptomics with pre-operative blood plasma lipidomics, we define

89 the relative importance of gene co-expression modules and lipids that are positively or

90 negatively associated with HbA1c levels, pointing to potential prognostic markers.

91

92

93

94

95

Introduction

97

98   Type 2 diabetes (T2D) mellitus defines a cluster of genetically complex pathological states

99   characterized by persistent hyperglycemia, often leading to cardiovascular complications,

100 kidney failure, retinopathy and neuropathies. Affecting more than 450 million people, with rising

101 incidence rates over the past decades, this syndrome is a major threat for public health and

102 society globally[1]. Common determinant and ultimate cause of T2D is the inability of pancreatic

103 islet beta cells to secrete insulin in adequate amounts relative to insulin sensitivity, in the

104 absence of evidence for their autoimmune destruction or a monogenetic deficit. Beta cell failure

105 typically results from a lengthy process spanning many years. Remarkably, however, it can be

106 rapidly reverted upon bariatric surgery or severe caloric restriction[2,3]. These observations argue

107 against the occurrence of major beta cell apoptosis in T2D, especially since adult beta cells

108 hardly replicate, while robust evidence of beta cell neogenesis after puberty is also lacking.

109 Hence, the prevailing opinion is that persistent metabolic stress drives mature beta cells to

110 phenotypically de-differentiate into progenitor cells or trans-differentiate into other islet

111 endocrine cell types over time[4–6]. As the pathogenesis of beta cell dysfunction in T2D remains

112 largely unclear, the diagnosis of this disease relies on accepted, surrogate parameters and

113 cutoffs that have been primarily developed for clinical practice to optimize therapeutic

114 interventions[7].

115

116 Insight into molecular alterations associated with impaired insulin secretion in T2D has been

117 largely obtained from pancreatic islets isolated enzymatically from brain-dead or cadaveric

118 subjects classified according to a categorical division into non-diabetic and diabetic, rather than

119 on a continuum from euglycemia to steady hyperglycemia. This approach has multiple

5

120   shortcomings[8]. Briefly, islet researchers do not generally have access to extensive clinical and

121   laboratory information about the donors prior to their admission to an intensive therapy unit[9].

122   Moreover, the islet state is perturbed by the metabolic stress associated with a terminal

123   condition and the related pharmacological treatments[10,11]. Enzymatic isolation of islets and their

124   in vitro culture can further change their molecular profile[12,13]. In the attempt to overcome, at least

125   in part, these limitations, we established a complementary platform for the procurement of islets

126   which relies on the collection and analysis of pancreatic specimens from metabolically profiled

127   living donors undergoing pancreatectomy for a variety of disorders[8,14]. We showed that this

128   approach is very reproducible and scalable and provides a novel view on transcriptomic and

129   functional alterations in pancreatic islets of subjects with T2D[15–17]

130

131   The aim of the present study has been to profile in greater detail gene expression changes

132   occurring along the progression from euglycemia to long-standing T2D in human islets *in situ*

133   and to integrate this knowledge with clinical traits, circulating lipid levels and the islet proteome,

134   hence enabling inferences about the mechanisms driving islet dysfunction and the identification

135   of potential biomarkers for it.

136   # Results

137   ## Living donors enable islet studies along progression to T2D

138

139   To gain insight into the history of islet cell deterioration along the progression from normal

140   glycemic regulation to T2D, we collected surgical pancreatic tissue samples from 133

141   metabolically phenotyped pancreatectomized patients (PPP). Eighteen were non-diabetic (ND),

142   41 had impaired glucose tolerance (IGT), 35 Type 3c Diabetes (T3cD) and 39 T2D (Fig. 1A and

143   Fig. 1B). These group assignments were based on glycemic values at fasting and at the 2 h

144   time point of an oral glucose tolerance test (OGTT) using the thresholds defined in the

145 guidelines of the American Diabetes Association[7], or, when applicable, on a previously

146 established diagnosis of T2D. In this cohort, 51.9% were males and the mean age was

147 65.36±11.54 years, with ND PPP being on average younger than the other three groups (Fig.

148 1C and Supplementary Table 1). The body mass index (BMI) was significantly lower in ND

149 compared to IGT, T3cD and T2D PPP. The HbA1c value, as a parameter of longer-term

150 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

151 PPP (Fig. 1C and Supplementary Table 1). Moreover, based on histopathology, malignant

152 tumors occurred in 50%, 60.97%, 74.29% and 69.23% of ND, IGT, T3cD, and T2D PPP,

153 respectively (Supplementary Table 1).

154

155 Islet gene expression drifts with glycemic deterioration

156

157 Gene expression profiles of islets isolated by laser capture microdissection (LCM) from resected

158 and snap-frozen pancreas samples of ND, IGT, T3cD and T2D PPP were assessed by RNA

159 sequencing. After removal of genes with low expression levels, the overall islet transcriptome

160 encompassed 19,119 genes, of which 14,699±693 were present (raw read counts >0) in ND

161 PPP, 14,967±455 in IGT PPP, 14,939±493 in T3cD PPP and 14,997±428 in T2D PPP. Genes

162 with a fold change (FC)>1.5 and a false discovery rate (FDR)≤0.05 were considered to be

163 differentially expressed (DE) between the groups. Multiple group comparison by linear modeling

164 was performed (Supplementary Table 2). Subsequent pairwise group comparisons of IGT vs.

165 ND, T3cD vs. ND and T2D vs. ND revealed an exacerbation of gene dysregulation with

166 deterioration of glycemic control (Fig. 2A). Notably, no DE islet genes were identified between

167 IGT vs. ND PPP, while 161 and 650 DE genes were found between T3cD vs. ND PPP and T2D

168 vs. ND PPP, respectively (Fig. 2A and Supplementary Table 2).

169

170   Restricting the transcriptomic analysis to libraries in which insulin (*INS*) was the most expressed

171   gene resulted in the retention of islet datasets from 15 ND, 35 IGT, 21 T3cD and 24 T2D

172   subjects, without substantially affecting the overall composition of the cohort in regard to

173   diabetes status and major descriptive parameters (Supplementary Table 3). Deconvolution

174   analysis indicated that in 78.3% of retained samples the proportion of beta cells was >50%

175   (Extended data Fig. S1), supporting the choice of this strategy to discriminate samples

176   especially enriched in beta cell transcripts. This analysis further pointed to the overall

177   enrichment in beta cell content of the LCM isolated islets in comparison to another large study

178   based on islets isolated by enzymatic digestion (median beta cell:non-beta endocrine cell ratio

179   3.98:1 and 1.4:1)[18]. This enrichment can conceivably be attributed to the selectivity of the LCM

180   isolation procedure for beta cell rich areas due to their higher autofluorescence. Despite the

181   expected reduction in statistical power due to ~ 30% smaller size of this "restricted" cohort (92

182   samples retained from 133), the number of significantly DE genes increased in the multiple

183   group comparison, as well as in pairwise comparisons between islets of T2D vs. ND PPP by

184   51% to 984 (782 up, 202 down), and by 59% to 256 (209 up, 47 down) between islets of T3cD

185   vs. ND PPP (Fig. 2A, Supplementary Table 4). Seven of the 984 DE genes are among the

186   putative effectors of GWAS risk loci for T2D (https://t2d.hugeamp.org/), two upregulated

187   (*SGSM2* and *BCL2*) and five downregulated (*RASGRP1, G6PC2, SLC2A2, ZMAT4* and

188   *PLUT*)[19], while most of the remaining genes have not been previously reported to be altered in

189   islets of subjects with T2D[14,20].

190

191   Among the DE genes in islets of T2D PPP, *INF2* and *AKR7L* were negatively correlated in a

192   moderate fashion with duration of the disease measured in years (Spearman correlation

193   coefficient -0.32 and -0.41 respectively), albeit they were both upregulated relative to islets of

194   ND PPP. Most notably, this filtering step enabled, for the first time, the identification of 185 DE

195   genes between islets of IGT vs. ND PPP. Most of these DE genes were upregulated (181/185),

196 and 98 also differentially regulated with the same directionality (97 up, 1 down) between islets of

197 T2D vs. ND PPP. Intriguingly, and apparently at variance with previous findings[21], the proposed

198 T2D risk genes *ARAP1* and its neighboring gene *STARD10* were both upregulated and among

199 the 77 genes differentially regulated in islets of IGT PPP only. No islet cell type specific genes[22]

200 were enriched in any of the differential expression analyses. Furthermore, no shift of islet cell

201 type proportions with the progression of the disease was observed in the deconvolution analysis

202 (Extended Data Fig. S1A).

203

204 Concerning samples with the highest transcript other than insulin, these were not noteworthily

205 different from the other samples in any of the clinical parameters or anatomical part of the

206 pancreas the tissue originated from. Nine of them had *PRSS1* (coding for trypsin) as the most

207 enriched transcript, pointing to exocrine contamination and one was marked by *MALAT1* and

208 was therefore excluded as suspect for cancerous cell contamination. The remaining samples

209 were remarkable for expressing a non-beta-cell endocrine gene, specifically 13 samples with

210 predominant alpha cell (*GCG* or *TTR*) and 18 samples with predominant gamma cell (*PPY*)

211 characteristic genes. This is partially reflected by the results of the deconvolution analysis

212 (Extended Data Figure S1). This specific group of 41 samples was not analyzed further since

213 the number of subjects in each of the four glycemic groups was too small for statistical analysis.

214

215 For both the "restricted" and the full data set, heatmaps of gene expression levels in the four

216 patient groups were prepared as a visual complement to the statistical analysis (Fig. 2B and

217 Extended Data Fig. S2A). Despite the marked differences between the findings in the

218 "restricted" and complete cohort, upregulation prevailed as the direction of gene dysregulation in

219 both of them (Fig. 2A and Extended Data Fig. S2A). Based on these observations, pancreatic

220 tissue sections of 5 ND and 5 T2D PPP with the "restricted" cohort were immunostained with

221 antibodies specific for histone H3 and H4 lysine acetylation – an epigenetic modification

222   associated with greater access of transcription factors to promoter sites resulting in increased

223   gene expression. Notably, qualitative assessment by immunostaining indicated a remarkable

224   increase of the signals for acetylated H3 and H4 in the islets, and also in the surrounding

225   exocrine cells of T2D PPP, compared to ND PPP (Fig. 2D).

226   ## Gene pathways are progressively perturbed from IGT to T2D

227

228   We further analyzed differentially expressed gene functions by gene set enrichment analysis

229   using Gene Ontology terms and KEGG pathways (Fig. 2C, Extended Data Fig. S2B and

230   Supplementary Tables 5 and 6). Results obtained from the different gene set collections cross-

231   validated each other, since similar biological themes emerged. Islets of pre-diabetic and diabetic

232   subjects displayed upregulation of islet genes that were functionally related to cell-extracellular

233   matrix interaction, immune response and signaling pathways, while expression of genes related

234   to RNA processing, protein translation and mitochondrial oxidative phosphorylation were

235   downregulated. Importantly, the analysis performed on the "restricted" cohort, differently from

236   the full dataset, also revealed that the strength of the enrichment increased with progression of

237   the disease (Fig. 2C and Extended Data Fig. S2B). These data suggest that early dysregulation

238   of gene pathways exacerbates with the decline of beta cell function.

239   ## WGCNA identifies islet gene modules correlated with HbA1c

240

241   To globally interpret transcriptomic data and identify sets of genes likely to be functionally

242   related and co-regulated, we grouped genes based on similarities in their expression profiles

243   into modules using a network-based approach[23]. In the cohort of 133 PPP, we identified 36 co-

244   expressed gene modules, which were arbitrarily labeled M1 through M36. The expression

245   profiles of the genes in each module were summarized by a module eigengene, or first principal

246   component of the expression matrix. Module eigengenes were used to computationally relate

247   modules to one another and to genes or clinical variables. Correlation between module

248   eigengenes and diabetes-related clinical traits revealed modules M9 and M14 as those with the

249   highest positive and negative correlation with HbA1c, respectively (Fig. 3A and Supplementary

250   Table 7). The former consisted of a set of genes that showed similar patterns of increased

251   expression in most PPP with T2D (Fig. 3B), while the latter was mostly composed of genes with

252   coordinated down-regulation in diseased subject samples (Fig. 3C).

253

254   We next correlated the expression values of each gene contained in a module to the eigengene

255   of the module. The correlation coefficient from this calculation is denoted as the "module

256   membership" of the gene and serves as a quantitative measure of how representative a gene is

257   of the module it belongs to. Strong module memberships point to genes that are highly

258   connected in the underlying gene-gene similarity network of the WGCNA. This analysis allowed

259   us to identify highly connected genes or "hub" genes for HbA1c-related modules (Fig. 3D-E).

260   These included genes that we had previously identified as differentially expressed in subjects

261   with T2D, and which were correlated with HbA1c either positively, such as module M9 genes

262   *ALDOB* (FC=8.45 with adj. *p*<0.001 in T2D vs. ND in the "restricted" cohort) and *FAIM2*

263   (FC=7.11 with adj. *p*<0.001 in T2D vs. ND in the "restricted" cohort) or negatively, such as

264   module M14 genes *SLC2A2* (FC=-2.77 with adj. *p*<0.001 in T2D vs. ND in the "restricted"

265   cohort) and *TMEM37* (FC=-1.73 with adj. *p*<0.001 in T2D vs. ND in the "restricted" cohort). As in

266   other studies in mouse models of diabetes[24], we found *ALDOB* to be also upregulated in islets

267   from 13-week-old diabetic *db/db* mice compared to the heterozygous *db/+* littermate (Extended

268   Data Fig. S3A) as well as in a mouse beta, but not alpha, cell line upon exposure to high

269   glucose (Extended Data Fig. S3B). However, the overexpression of *ALDOB* in beta cells of T2D

270   PPP could not be verified by immunofluorescence on tissue sections due to the cross-reactivity

271   of the only available "specific" anti-ALDOB antibody with other aldolase isoforms (Extended

272   Data Fig. S3C).

273

The islet proteome is more heterogeneous in T2D

275

276 To verify and extend the transcriptomic data at the functional level of proteins, we analyzed the

277 mass spectrometry (MS)-based proteomic profiles of LCM pancreatic islets from five ND and

278 five T2D PPP (Supplementary Table 8). We chose these samples primarily based on tissue

279 availability and secondarily based on the levels of *ALDOB* found in RNA sequencing. Their islet

280 transcriptomics profiles closely resembled the one of the complete cohort: top regulated genes

281 in these 10 samples from T2D and ND PPP (Supplementary Table 2 and Extended Data Figure

282 S4C) were also among the most significantly differentially expressed islet genes in the entire

283 cohort. Using a very high sensitivity workflow on a novel trapped-ion mobility Time of Flight

284 mass spectrometer and miniaturized sample preparation[25], we identified 2,237±499 islet

285 proteins for ND PPP and 1,819±412 islet proteins for T2D PPP (Figure 4A). Quantitative

286 reproducibility between biological replicates was high with Pearson correlations ranging from

287 0.83 to 0.95 (Extended Data Fig. S4A). Principal component analysis (PCA) clustered the data

288 into two distinct groups matching the clinical stratification (Fig. 4B, see methods for detailed

289 data processing steps). Interestingly, islets of ND PPP clustered closely, indicating a very

290 similar proteome signature, while those of T2D PPP revealed substantial proteome

291 heterogeneity among each other. Differential expression analysis confirmed that islets of T2D

292 and ND PPP have very different proteomic profiles. The main differential drivers are well-

293 characterized markers of pancreatic islet cells, including SLC2A2[26], and many proteins

294 implicated in mitochondrial structure, translation, energy supply and amino acid or fatty

295 metabolism such as YMEL1, MRPL12, BA3(C14orf159), ACADS and its paralogue ACADSB,

296 which were highly depleted in islets of T2D PPP (Fig. 4C). Besides AKR7L, ACADS was the

297 only other upregulated and differentially expressed gene in islets of both IGT and T2D PPP,

298 while being also downregulated at the protein level. All differentially expressed mitochondrial

12

299   proteins are encoded by the nuclear genome (Fig. S4B). Intriguingly, the level of the

300   sulfonylurea receptor ABCC8 subunit[27] was also strongly reduced in islets of T2D PPP. This

301   downregulation might be an effect secondary to pharmacological treatment, as three among

302   these patients had been treated with anti-diabetic SUR1 antagonists glibenclamide (DP197),

303   glimepiride (DP118) or mitiglinide (DP087) (Extended Data Fig. S4C). Furthermore, we found

304   that transcriptome and proteome levels of pancreatic islets from the same donors are very

305   different (Extended Data Fig. S4D), as shown in another cellular system at single-cell level[28].

306   Nevertheless, we report the glycolytic enzyme ALDOB to be consistently upregulated

307   (Proteome: 4-fold, Transcriptome: FPKM: 76.16±50.82 in T2D PPP vs. 4.63±0.95 in ND PPP),

308   and the glucose transporter SLC2A2 to be downregulated (Proteome: 4-fold, Transcriptome: 4-

309   fold) in islets of T2D vs. ND PPP samples on both modalities (Extended Data Fig. S4E, 4C).

310   This is consistent with our transcriptomic data and that of previous studies[14,15] and our current

311   WGCNA analyses. Other proteins robustly overexpressed in islets of T2D PPP included the

312   alpha-L-fucosidase FUCA1 and the surface marker for hematopoietic stem cells THY1.

313

314   Next, we employed the proteomic ruler algorithm and annotations of subcellular localization to

315   compare the protein mass distribution of major cellular compartments[29](Fig. 4D). Islets of T2D

316   PPP lost an estimated protein mass of 6% in the Golgi apparatus, 24% in the endoplasmic

317   reticulum, and 27% in the mitochondria compared to those of ND PPP, while the cytoskeleton

318   protein mass was unchanged. Unsupervised hierarchical clustering of all 2,622 detected

319   proteins, clustered the data according to clinical categories (Fig. 4E). One-dimensional gene

320   ontology enrichment[30] revealed that two distinct clusters whose protein intensity levels

321   associated with the terms 'membrane attack complex' (p<2.18E-04) and 'Immunoglobulin C-

322   domain' (p<2.68E-06) were enriched by 2.27-fold and 2.36-fold in islets of T2D vs. ND PPP,

323   respectively. Proteins with the gene ontology-term 'differentiation' (p<3.09E-04) and

324   '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

327   Our study encompassed two independently generated lipidomics data sets. First, shotgun

328   lipidomics was performed on peripheral blood plasma samples of the aforementioned cohort (4

329   ND, 21 IGT and IFG, 13 T3cD and 17 T2D) (Supplementary Tables 9 and 10). Second,

330   sphingolipid profiling was performed on peripheral blood samples of subjects within the cohort

331   subjected to transcriptomic analysis (11 ND, 32 IGT and IFG, 26 T3cD and 32 T2D)

332   (Supplementary Tables 11 and 12). Prior to data analysis, lipidomics samples from PPP with

333   very high bilirubin values (>100 µmol/l) were removed to avoid bias in lipidomics profiles. In

334   each of the two data sets, all available samples from non-diabetic PPP (ND, as previously

335   defined) and the subset of IGT PPP with HbA1c<6.0 were combined into one group, which is

336   referred henceforth as ND for readability. The resulting sample sizes used in patient group

337   comparisons were as follows: 17 ND, 13 T3cD and 17 T2D in the shotgun lipidomics data set;

338   32 ND, 21 T3cD and 27 T2D in the targeted lipidomics data set.

340   Statistical tests included covariates to adjust for age, sex and BMI, similar to the transcriptomics

341   analysis. Pairwise comparisons of T2D vs ND and T3cD vs ND were performed. In shotgun

342   lipidomics, 113 lipid species from 11 classes were included in the data analysis. When

343   comparing T2D and T3cD to ND PPP, the majority of lipid classes displayed a remarkably

344   homogeneous downward-trend of the individual lipid species they comprised (Fig 5A-B). Most

345   prominently, plasma concentrations of lipids within the class of ether-linked

346   phosphatidylcholines (PC O-), a large class with 30 measured species, were lower in T2D

347   versus ND PPP. Fourteen lipids of this class were significantly decreased (adjusted p<0.05),

348   with all of them showing at least a 1.4-fold change. A few lipid species from smaller

349   phospholipid classes (one phosphatidylcholine (PC), one lysophosphatidylcholine (LPC) and

350   one phosphatidylinositol (PI)), as well as two from the sphingomyelin class (SM), were also

351 significantly less abundant in T2D than in ND PPP (PC 18:0;0/18:2;0: FC=-1.43, adj. *p*=0.040;

352 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=-

353 1.33, adj. *p*=0.037; SM 34:1;2:, FC=-1.24, adj. *p*=0.04). (Fig. 5A-B and Supplementary Table

354 13).

355

356 Next, we performed targeted sphingolipidomics on 14 distinct lipid species for accurate plasma

357 level estimation (ceramides, dihydroceramides and sphingoid bases) (Supplementary Table 16).

358 Plasma levels of ceramides d18:1/18:0 and d18:1/20:0 were increased in T2D compared to ND

359 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

360 testing correction). A similar trend towards elevation in T2D vs ND PPP was also observed in

361 the two dihydroceramide species having the same chain lengths as these ceramides, although

362 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;

363 DH Cer d18:0/20:0: FC=1.35, *p*=0.02). Thus, in our data set, plasma concentrations of some

364 ceramides and their precursor dihydroceramides appear to increase simultaneously in T2D. Cer

365 d18:1/24:0, but not the corresponding dihydroceramide, was observed to move in the opposite

366 direction being lower in T2D (FC=-1.28, *p*=0.017). Notably, ceramides were identified by LC-MS

367 (see methods) and, independently, by shotgun FT-MS and both profiles matched. Regarding the

368 LC-MS/MS analysis, the parent ion selected for dihydroceramides identification and

369 quantification was the protonated ion (without water loss). In FT-MS, we observed no significant

370 water loss from the ceramide standards. We therefore have reasons to believe that we detected

371 dihydroceramides, but not deoxyceramides, which are isomeric of the water loss form of the

372 dihydroceramides.[31]

373 Data integration identifies pathways for islet dysfunction

374

375 To identify a multivariate molecular profile that explains diabetes progression in the PPP cohort,

376 we performed a large-scale integrative multi-omics analysis combining clinical data with islet

377 transcriptomics and plasma lipidomics. Integration of transcriptomics and lipidomics data in the

378 same model enables to weigh the relative importance of lipid and gene expression features in

379 relationship to a chosen clinical trait. Hence, we explored the relationship between gene co-

380 expression modules and plasma lipids by computing a consensus orthogonal partial least

381 square (consensus OPLS)[32,33] model with HbA1c as the outcome. All three types of biological

382 data, namely gene co-expression modules, lipids from the shotgun analysis and sphingolipids

383 from the targeted analysis, contributed to the model (35%, 46.5% and 18.5%, respectively),

384 suggesting that they help to explain HbA1c levels in a complementary way. Among them,

385 different lipids and gene modules appear as the most relevant variables in the statistical

386 modelling of HbA1c levels (Fig. 6A, 6B and Supplementary Table 14). Importantly, the model

387 explained a large portion of data variance, highlighting a good fit with the experimental data (see

388 Methods for more details).

389

390 Among all considered biological data, the co-expression modules M1, M4, M8, M9, M30, M35

391 and M36 were the top predictive variables for high HbA1c levels, along with the two ceramide

392 species C20 and C18. TAGs were also contributing, although to a lesser extent (Fig 6A, right

393 hand side). Conversely, low levels of HbA1c were strongly related to the co-expression modules

394 M12 and M14 (Fig 6A, left hand side). However, the majority of the predominant predictors for

395 low HbA1c were lipid species, most importantly the PC O- class. This class was also found to

396 be lower in T2D compared to ND patient groups in differential abundance analysis, as shown in

397 Fig 5A. A number of SM, PI and PC lipid species were next in the importance ranking related to

398 low HbA1c, followed by the gene co-expression module M29. These results suggest that the

399 profile of patients with increased HbA1c is characterized by multiple molecular components,

400 some of which represent signals that were neither captured by differential abundance analyses

16

401 comparing diabetes status groups nor by correlating gene co-expression modules individually to

402 HbA1c. Most importantly, consensus OPLS multi-omics analysis pointed towards additional

403 gene co-expression modules that may play a role in glucose dysregulation.

404

405 Next, we used the results from the integrative data modelling to infer a network of key altered

406 biological pathways in dysfunctional beta cells. To this end, we pooled gene modules positively

407 associated with HbA1c levels (M1, M4, M8, M9, M30, M35 and M36) (Fig. 6A) and assessed

408 their overlap to KEGG pathways by over-representation analysis. We found that the biological

409 themes underlying these genes were very similar to the pathways upregulated in T2D and IGT

410 PPP and include cell-matrix interaction, cell signaling and immune response (Fig. 6C and

411 Supplementary Table 15). The same strategy was used to identify pathways associated with

412 genes from modules with a negative prediction score for HbA1c (M12, M14 and M29) (Fig. 6A),

413 revealing an enrichment for metabolic pathways (Fig. 6C and Supplementary Table 15). Of

414 note, several islet genes differentially regulated in T2D PPP were driving the enrichment of

415 these pathways. These include, for example, *ALDOB*, which stood out for its strong correlation

416 to HbA1c levels (Fig. 3D and Fig. 6C). These genes, or the proteins encoded by them, should

417 be regarded as putative candidate biomarkers for monitoring disease progression and

418 therapeutic intervention.

Discussion

419

420

421 This study provides an extensive analysis on islets *in situ* and plasma samples from the largest

422 cohort of in-depth metabolically profiled living donors. Multi-omics data were generated using

423 state-of-the-art approaches and integrated in a fashion not previously used in studies on islet

424 dysregulation in relation to hyperglycemia in humans. Our transcriptomic and proteomic data

425 from islets *in situ* of ND subjects represent a valuable reference for future investigations. More

426   broadly, this dataset would be a worthwhile addition to the growing number of islet resources on

427   type 1 and type 2 diabetes by different consortia, such as the Network for Pancreatic Organ

428   Donors with Diabetes (nPOD)[34], Human Pancreas Analysis Program (HPAP)[35], or the

429   Translational human pancreatic Islet Genotype tissue - Expression Resource (TIGER)

430   (https://www.t2dsystems.eu/tiger-database). Furthermore, we could identify for the first time a

431   set of islet genes altered in their expression already in subjects with impaired glucose tolerance.

432   This, in turn, enabled us to acquire an unprecedented cross-sectional overview of the

433   progression of islet gene dysregulation in parallel with the continuous elevation of HbA1c

434   values, beyond conventional thresholds for clinical classification of patients.

435

436   Pathways involved in RNA biology and especially in mitochondrial function emerged to be most

437   negatively perturbed - a conclusion which in the case of the latter was strongly corroborated by

438   the proteomic analysis, which enabled the identification of known and unknown differentially

439   expressed proteins in islets of T2D PPP. In this context, we emphasize the downregulation of

440   mitochondrial ACADS and its paralogue ACADSB, which catalyze the beta oxidation of short-

441   chain fatty acids, including sodium butyrate. This finding is intriguing in view of the ability of this

442   metabolite to broadly upregulate gene expression through inhibition of histone deacetylases.

443   Unlike in previous studies on isolated islets from brain-dead organ donors[14,19], but similar to

444   previous studies by us in human[14] and mouse models of diabetes[36] the vast majority of

445   differentially expressed genes in islets of T2D, but also IGT and T3cD PPP were upregulated.

446   Among those genes, *ALDOB* stands out being the one with the strongest correlation with the

447   islet gene module M9, which in turn has the strongest correlation with elevated HbA1c. Since

448   *ALDOB* is a marker of beta cell precursors[37], its overexpression could be interpreted as a sign

449   that in T2D, mature beta cells revert back to an immature stage of differentiation, or that a

450   compartment equivalent to the lifelong niche of virgin beta cells identified in adult mice expands

451   as a potential compensatory source of new beta cells[37]. However, no additional disallowed gene

452   of immature beta cells, markers of beta cell precursors or other islet cell types were differentially

453   regulated, while key determinants of mature beta cells, such as *PDX1*, *MAFA*, *NKX6.1* or *UCN3*

454   were unchanged, at least at the transcriptomic level. Retention of fractions of major islet cell

455   types (alpha, beta and delta) within the islet in T2D, consistent with recent imaging studies in

456   samples from pancreatectomized subjects (Cohrs et al)[17], was confirmed by deconvolution

457   analysis. Our global unbiased proteomic analysis, which corroborated the upregulation of

458   ALDOB, further showed that the expression profile of islet cells in T2D PPP is very divergent,

459   opposite to its remarkable homogeneity in islet cells of ND subjects. Hence, the regression of

460   beta cells toward a de-differentiated state following a linear trajectory recapitulating their

461   developmental path to maturation or their transdifferentiation into other islet cell types seems

462   less likely than a disharmonic relaxation of constraints on gene expression. Such processes,

463   although possibly reversible, could perturb the coordinated operation of islet cells, including beta

464   cells. In line with this, Lawlor *et al.* reported no evidence of beta cell

465   dedifferentiation/transdifferentiation and alterations in fractions of islet cells in the context of T2D

466   upon sequencing of single islet cells from a small cohort of ND and T2D organ donors, although

467   this conclusion has been more recently challenged[38]. While we strived to selectively enrich the

468   beta cell content of our omics data by laser capture microdissection of bulk islets based on the

469   lipofuscin-associated autofluorescence of beta cells and by subsequent deconvolution of the

470   data during their analysis, the unavoidable presence of other cell types in the samples

471   introduces a degree of uncertainty. Thus, for the future it would be important to assess whether

472   overexpression of ALDOB occurs indeed in beta cells and if it affects their glycolysis and

473   metabolism, taking into account that its paralogue ALDOA, whose RNA and protein levels were

474   unchanged, remains by far the predominant islet aldolase species. Attention may also be

475   directed toward understanding whether impaired oxidative phosphorylation, as a likely outcome

476   of the massively decreased expression of mitochondrial proteins, and thus energy balance

477   homeostasis, accounts, at least in part, for the observed less restrained gene expression.

478

479 The transcriptome and proteome of islets from subjects with T2D displayed the occurrence of an

480 immune response. At this time, however, we are not aware of factors which might readily

481 account for the presence of such signals. Specifically, patients with antibodies against known

482 autoantigens of T1D were excluded from the analysis. As in a previous report[15], histological

483 examination of the specimens did not reveal insulitis or macrophage infiltration. Pancreatitis was

484 more common among subjects with normoglycemia (22%) than with T2D (15%). The presence

485 of cancer cells in our islet preparations is also unlikely. Specifically, a qualified pathologist

486 routinely examined the surgical specimen to minimize the chance of contamination by

487 neoplastic tissue before it was taken for downstream processing. Histological survey of the

488 tissue did not reveal the presence of neoplastic cells in the islets. The transcriptomic analysis in

489 a previous study indicated that exocrine contamination of LCM islets from PPP was comparable

490 or less than in the case of enzymatically-isolated islets from organ donors (OD)[14]. Moreover, in

491 the same study we found no evidence for an enrichment of tumor cell transcripts in LCM islets

492 of PPP compared to islets of OD. Likewise, an enrichment analysis of pancreatic cancer specific

493 genes in the differentially expressed islet gene sets reported here using hypergeometric test

494 showed no enrichment for any of the four described pancreatic cancer subtypes (p = 0.87) as

495 reported in Bailey et al 2016[39]. Rather, the gene expression clustering was driven by the islet

496 isolation method and not by the origin of the tissue (OD vs. PPP). We still appreciate that in our

497 cohort cancer prevalence was higher in the T2D (69%) than in the ND (50%) group. Thus, we

498 cannot entirely rule out a metabolic pro-inflammatory impact of the cancer on islet gene and

499 protein expression or function.

500

501 Our lipidomics analyses revealed lowered phospholipid species (14 PC O-, one PC, one LPC,

502 one PI) and some elevated ceramides and dihydroceramides in T2D PPP. These findings match

503 observations reported in other recent studies on larger cohorts. Huynh et al (2019)[40] presented

504   a comprehensive shotgun lipidomics study on the AusDiab cohort, including 640 samples and

505   636 lipid species. In this work, many PC, PC O-, LPC and PI had a significant negative

506   association with blood glucose levels either after overnight fasting or at the 2-hour point of an

507   OGTT, including nine species that were found negatively associated with T2D in our own study

508   (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

509   ceramide Cer d18:1/18:0 and its precursor DH Cer d18:0/18:0 had both a significant positive

510   association with fasting glucose, supporting our notion that this lipid pair might be linked to

511   diabetes status. Furthermore, several prospective case-control studies reported significantly

512   decreased PC, PC O- and LPC plasma concentrations[41] or elevated dihydroceramide levels[42,43]

513   in progressors to T2D compared to non-diabetic controls. The congruency of these results

514   points to these lipids as potential biomarkers of beta cell function in T2D.

515

516   Finally, we use a data fusion method[32,33] to generate a model of how different molecular

517   features (islet gene co-expression, plasma shotgun lipidomics and targeted sphingolipidomics)

518   contribute to HbA1c levels in a continuum from healthy individuals to those with overt T2D. This

519   model allowed us to measure the *relative* importance of different molecular components in

520   explaining HbA1c variability, providing unique insights into the molecular profiles of individuals

521   as they lose glycemic control towards development of T2D. The rational for combining plasma

522   lipidomics with islet gene expression data was that the levels of some plasma lipids may affect

523   pancreatic islets and/or reflect changes occurring within them and thus be useful as biomarkers

524   to assess beta cell dysfunction in prediabetes and T2D. To our knowledge this is the first time

525   such an approach has been used in this field and we suggest that, by modelling multiple levels

526   of information at the same time in deeply phenotyped populations such as the one presented

527   here, we can gain a holistic view of the system and draw conclusions regarding key pathways,

528   targets and biomarkers in metabolic and other diseases.

Data availability

530 RNA Sequencing data was deposited in the NCBI Gene Expression Omnibus with GEO

531 accession number GSE164416. Human genome reference assembly GRCh38 is publicly

532 available.

533 The proteomics raw datasets and the MaxQuant output files generated and analyzed throughout

534 this study were deposited at the ProteomeXchange Consortium via the PRIDE partner

535 repository with the project accession number PXD022561 (https://www.ebi.ac.uk/pride/archive/).

536 Lipidomics data was deposited in the Zenodo database (zenodo.org,

537 doi:10.5281/zenodo.4716063).

Acknowledgements

550

551 Author contributions

552 J.W. and M.D., patient recruitment and surgery, provision of clinical data; E.S., N.K. and D.F.,

553 sample collection and processing, data entry; D.A., pathology; M.B., N.K. and E.S., patient

554 database management and selection; A-D.B. and M.M., proteomics; M.L., A.D., RNA

555 sequencing, C.L.Q., P.B.S.H, P.D., C. K., M. G., K.S., lipidomics and sphingolipidomics; L.W.,

556 M.B., A-D.B., F.Ma., F.Me., F.B. and Ca.K., analysis and integration of multi-omics data; E.B.,

557 autoantibody test; A.S., data in mouse tissue and cell lines; M.B., immunofluorescence stainings

558 and antibody validation; B.T., D.A., J.W., A.S., M.M., M.I. and M.S., conceptual insights and

559 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

560 manuscript. All authors read, revised and approved the final version of the manuscript.

561 Competing interests

562 KS is CEO of Lipotype GmbH. KS and CK are shareholders of Lipotype GmbH. MJG is an

563 employee of Lipotype GmbH. PBSH and PD are employees of Servier. AS is an employee of

564 Sanofi-Aventis Deutschland GmbH. The other authors declare no conflict of interest.

565 References

566 1. Saeedi, P. *et al.* Global and regional diabetes prevalence estimates for 2019 and

567   projections for 2030 and 2045: Results from the International Diabetes Federation

568   Diabetes Atlas, 9th edition. *Diabetes Res. Clin. Pract.* **157**, (2019).

569 2. Mizera, M. *et al.* Type 2 Diabetes Remission 5 Years After Laparoscopic Sleeve

570   Gastrectomy: Multicenter Cohort Study. *Obes. Surg.* 1–7 (2020). doi:10.1007/s11695-

571   020-05088-w

572 3. Lim, E. L. *et al.* Reversal of type 2 diabetes: Normalisation of beta cell function in

573   association with decreased pancreas and liver triacylglycerol. *Diabetologia* **54**, 2506–

574   2514 (2011).

575   4.   Talchai, C., Xuan, S., Lin, H. V., Sussel, L. & Accili, D. Pancreatic β cell dedifferentiation

576   as a mechanism of diabetic β cell failure. *Cell* **150**, 1223–1234 (2012).

577   5.   Wang, Z., York, N. W., Nichols, C. G. & Remedi, M. S. Pancreatic β cell dedifferentiation

578   in diabetes and redifferentiation following insulin therapy. *Cell Metab.* **19**, 872–882

579   (2014).

580   6.   Cinti, F. *et al.* Evidence of β-Cell Dedifferentiation in Human Type 2 Diabetes. *J. Clin.*

581   *Endocrinol. Metab.* **101**, 1044–1054 (2016).

582   7.   American Diabetes Association. Classification and diagnosis of diabetes: Standards of

583   Medical Care in Diabetes-2020. *Diabetes Care* **43**, S14–S31 (2020).

584   8.   Barovic, M. *et al.* Metabolically phenotyped pancreatectomized patients as living donors

585   for the study of islets in health and diabetes. *Molecular Metabolism* (2019).

586   doi:10.1016/j.molmet.2019.06.006

587   9.   Poitout, V. *et al.* A call for improved reporting of human islet characteristics in research

588   articles. *Diabetes* **68**, 209–211 (2019).

589   10.   Ebrahimi, A. *et al.* Evidence of stress in β cells obtained with laser capture

590   microdissection from pancreases of brain dead donors. *Islets* **9**, 19–29 (2017).

591   11.   Toyama, H., Takada, M., Suzuki, Y. & Kuroda, Y. Activation of macrophage-associated

592   molecules after brain death in islets. *Cell Transplant.* **12**, 27–32 (2003).

593   12.   Negi, S. *et al.* Analysis of Beta-Cell gene expression reveals inflammatory signaling and

594   evidence of dedifferentiation following human islet isolation and culture. *PLoS One* **7**, 1–

595   11 (2012).

596   13.   Weir, G. C. Glucolipotoxicity, β-cells, and diabetes: The emperor has no clothes.

597   *Diabetes* **69**, 273–278 (2020).

598   14.   Solimena, M. *et al.* Systems biology of the IMIDIA biobank from organ donors and

599   pancreatectomised patients defines a novel transcriptomic signature of islets from

600   individuals with type 2 diabetes. *Diabetologia* **61**, 641–657 (2018).

601   15.   Gerst, F. *et al.* The Expression of Aldolase B in Islets is Negatively Associated with

602   Insulin Secretion in Humans. *J. Clin. Endocrinol. Metab.* **103**, 4373–4383 (2018).

603   16.   Khamis, A. *et al.* Laser capture microdissection of human pancreatic islets reveals novel

604   eQTLs associated with type 2 diabetes. *Mol. Metab.* **24**, 98–107 (2019).

605   17.   Cohrs, C. M. *et al.* Dysfunction of Persisting β Cells Is a Key Feature of Early Type 2

606   Diabetes Pathogenesis. *Cell Rep.* **31**, (2020).

607   18.   Viñuela, A. *et al.* Genetic variant effects on gene expression in human pancreatic islets

608   and their implications for T2D. *Nat. Commun.* **11**, (2020).

609   19.   Mahajan, A. *et al.* Fine-mapping type 2 diabetes loci to single-variant resolution using

610   high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513

611   (2018).

612   20.   Taneera, J. *et al.* Identification of novel genes for glucose metabolism based upon

613   expression pattern in human islets and effect on insulin secretion and glycemia. *Hum.*

614   *Mol. Genet.* **24**, 1945–1955 (2014).

615   21.   Carrat, G. R. *et al.* Decreased STARD10 Expression Is Associated with Defective Insulin

616   Secretion in Humans and Mice. *Am. J. Hum. Genet.* **100**, 238–256 (2017).

617   22.   Xin, Y. *et al.* RNA Sequencing of Single Human Islet Cells Reveals Type 2 Diabetes

618   Genes. *Cell Metab.* **24**, 608–615 (2016).

619   23.   Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network

620   analysis. *BMC Bioinformatics* **9**, 559 (2008).

621   24.   Haythorne, E. *et al.* Diabetes causes marked inhibition of mitochondrial metabolism in

622   pancreatic β-cells. *Nat. Commun.* **10**, (2019).

623   25.   Meier, F. *et al.* Online parallel accumulation–serial fragmentation (PASEF) with a novel

624   trapped ion mobility mass spectrometer. *Mol. Cell. Proteomics* **17**, 2534–2545 (2018).

625   26.   Thorens, B. GLUT2, glucose sensing and glucose homeostasis. *Diabetologia* **58**, 221–

626   232 (2015).

627   27.   Pipatpolkai, T., Usher, S., Stansfeld, P. J. & Ashcroft, F. M. New insights into KATP

628   channel gene mutations and neonatal diabetes mellitus. *Nature Reviews Endocrinology*

629   **16**, 378–393 (2020).

630   28.   Brunner, A. D. *et al.* Ultra-high sensitivity mass spectrometry quantifies single-cell

631   proteome changes upon perturbation. *bioRxiv* 2020.12.22.423933 (2020).

632   doi:10.1101/2020.12.22.423933

633   29.   Wiśniewski, J. R., Hein, M. Y., Cox, J. & Mann, M. A 'proteomic ruler' for protein copy

634   number and concentration estimation without spike-in standards. *Mol. Cell. Proteomics*

635   **13**, 3497–3506 (2014).

636   30.   Cox, J. & Mann, M. 1D and 2D annotation enrichment: a statistical method integrating

637   quantitative proteomics with complementary high-throughput data. *BMC Bioinformatics*

638   **13 Suppl 16**, (2012).

639   31.   Zitomer, N. C. *et al.* Ceramide synthase inhibition by fumonisin B1 causes accumulation

640   of 1-deoxysphinganine. A novel category of bioactive 1-deoxysphingoid bases and 1-

641   deoxydihydroceramides biosynthesized by mammalian cell lines and animals. *J. Biol.*

642   *Chem.* **284**, 4786–4795 (2009).

643   32.   Boccard, J. & Rutledge, D. N. A consensus orthogonal partial least squares discriminant

644   analysis (OPLS-DA) strategy for multiblock Omics data fusion. *Anal. Chim. Acta* **769**, 30–

645   39 (2013).

646   33.   Boccard, J. & Rutledge, D. N. Iterative weighting of multiblock data in the orthogonal

647   partial least squares framework. *Anal. Chim. Acta* **813**, 25–34 (2014).

648   34.   Campbell-Thompson, M. *et al.* Network for Pancreatic Organ Donors with Diabetes

649   (nPOD): Developing a tissue biobank for type 1 diabetes. *Diabetes. Metab. Res. Rev.* **28**,

650   608–617 (2012).

651   35.   Kaestner, K. H., Powers, A. C., Naji, A. & Atkinson, M. A. NIH initiative to improve

652   understanding of the pancreas, islet, and autoimmunity in type 1 diabetes: The Human

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

654   36.   Ebrahimi, A. G. *et al.* Beta cell identity changes with mild hyperglycemia: Implications for

655   function, growth, and vulnerability. *Mol. Metab.* **35**, (2020).

656   37.   van der Meulen, T. *et al.* Virgin Beta Cells Persist throughout Life at a Neogenic Niche

657   within Pancreatic Islets. *Cell Metab.* **25**, 911-926.e6 (2017).

658   38.   Avrahami, D. *et al.* Single-cell transcriptomics of human islet ontogeny defines the

659   molecular basis of β-cell dedifferentiation in T2D. *Mol. Metab.* **42**, (2020).

660   39.   Bailey, P. *et al.* Genomic analyses identify molecular subtypes of pancreatic cancer.

661   *Nature* **531**, 47–52 (2016).

662   40.   Huynh, K. *et al.* High-Throughput Plasma Lipidomics: Detailed Mapping of the

663   Associations with Cardiometabolic Risk Factors. *Cell Chem. Biol.* **26**, 71-84.e4 (2019).

664   41.   Suvitaival, T. *et al.* Lipidome as a predictive tool in progression to type 2 diabetes in

665   Finnish men. *Metabolism.* **78**, 1–12 (2018).

666   42.   Wigger, L. *et al.* Plasma Dihydroceramides Are Diabetes Susceptibility Biomarker

667   Candidates in Mice and Humans. *Cell Rep.* **18**, 2269–2279 (2017).

668   43.   Mamtani, M. *et al.* Lipidomic risk score independently and cost-effectively predicts risk of

669   future type 2 diabetes: Results from diverse cohorts. *Lipids Health Dis.* **15**, (2016).

670

671 # Figure Legends

672 **Figure 1: Overview of the experimental procedures and cohort characteristics.** A)

673 Experimental procedures overview. Clinical data and peripheral blood were collected

674 preoperatively, and the snap-frozen surgical pancreatic tissue used for LCM of the islets of

675 Langerhans. Blood samples were analyzed for lipidomics, while LCM islets for transcriptomics

676 and proteomics. Omics datasets were individually evaluated in relationship to glycemic status

677 and further integrated with each other using Consensus Orthogonal Partial Least Squares

678   (OPLS) analysis. B) Waffle plot showing the structure of the cohort in terms of

679   glycemic/diabetes categories based on American Diabetes Association criteria. Absolute

680   numbers for each category are given in the legend boxes. C) Boxplots of four major clinical

681   parameters relevant for diabetes diagnosis and management. Statistically significant differences

682   from ND PPP were determined using the two-sided t-test (*p<0.05; **p<0.01). Boxplot spans

683   from 25th until 75th percentile, with centerline at median, whiskers extend to the most extreme

684   data point which is no more than 1.5 times the length of the box away from the box. Number

685   of observations in each comparison and category: age and BMI - 18 for ND, 41 for IGT(p=0.006

686   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);

687   fasting glycemia - 16 for ND, 38 for IGT (p=2.936-6), 34 for T3cD (p=1.249-7) and 33 for T2D

688   (p=2.692-7); glycemia at 2h point of OGTT: 15 for ND, 38 for IGT (p=1.486-6) and 23 for T3cD

689   (p=3.111-11).LCM: Laser Capture Microdissection; ND: Non-diabetic; IGT: Impaired Glucose

690   Tolerance; T3cD: Type 3c Diabetes; T2D: Type 2 Diabetes.

691   **Figure 2: Transcriptional changes between non-diabetic, pre-diabetic and diabetic**

692   **patients.** A) Number of DE genes identified by comparing glycemic groups of PPP in the entire

693   (all samples) or "restricted" cohort (*INS* filtered), using linear model with age, sex and BMI as

694   covariates. B) Gene expression profile of DE genes in the "restricted" cohort. Columns

695   represent patients grouped according to their glycemic status and ordered based on increasing

696   HbA1c levels. Rows, representing DE genes (variance stabilizing transformation normalized

697   counts), were clustered based on Euclidean distance. The colored side bar indicates in which

698   comparisons a gene was identified as differentially expressed. C) Gene Set Enrichment

699   Analysis of DE genes between IGT, T3cD or T2D and ND PPP in the "restricted" cohort. GO

700   terms and KEGG pathways are colored according to the normalized enrichment score.

701   Corresponding p-values are also indicated (*p<0.05, **p<0.01). D) Immunofluorescence for

702   insulin (green), acetylated histones H3 (left) and H4 (right) (magenta) in representative samples

703   of formalin fixed paraffin embedded pancreatic tissues from 5 ND and 5 T2D PPP. Scale bars

704   correspond to 20μm. DE: differentially expressed; ND: Non-diabetic; IGT: Impaired Glucose

705   Tolerance; T3cD: Type 3c Diabetes; T2D: Type 2 Diabetes.

706   **Figure 3: Identification of co-expressed gene modules related to diabetes traits.** A)

707   Correlation between module eigengenes and clinical traits including age, BMI, HbA1c, fasting

708   glucose, glucose at 2-hours after OGTT, HOMA2-B and HOMA2-IR. Each cell contains the

709   corresponding Spearman correlation coefficient and Student p value (in parenthesis). Cells are

710   colored according to their correlation to clinical traits. Modules are ordered based on their

711   correlation to HbA1c. B-C) Gene expression profiles of gene modules M9 (B) and M14 (C).

712   Columns, representing PPP, were grouped according to their glycemic status and ordered

713   based on increasing HbA1c levels. Rows, representing genes (variance stabilizing

714   transformation normalized counts), were clustered based on Euclidean distance. D-E) Scatter

715   plot of module membership vs. gene significance for HbA1c in modules M9 and M14. Genes

716   with the highest module membership and gene significance ("hub genes") are labeled. ND: Non-

717   diabetic; IGT: Impaired Glucose Tolerance; T3cD: Type 3c Diabetes; T2D: Type 2 Diabetes.

718   **Figure 4: Proteomics Analysis.** A) Number of identified proteins from pooled human

719   pancreatic islet cells isolated by LCM from PPP classified as non-diabetic (ND, N=5) or with

720   T2D (N=5). Boxplot spans from 25th until 75th percentile with centerline at median. Whiskers

721   extend to the most extreme data points in either direction. B) Principal Component Analysis

722   (PCA) of all grouped pancreatic islet measurements (ND=blue, T2D=orange). C) Volcano plot

723   comparing p values and $\log_2$-fold changes between islets of ND and T2D PPP. Multiple

724   hypothesis testing is controlled via Benjamini Hochberg correction at 5% False discovery rate.

725   D) Percentage distribution of total protein islet mass and its contribution per organelle between

726   ND and T2D PPP. The ND/T2D islet protein mass ratio in different organelles was normalized

727   by the nucleus protein mass. E) Hierarchical clustering of all islet proteins identified in the T2D

29

728   and ND PPP clusters. Log$_2$-transformed intensity values were normalized by z-scoring before

729   the clustering followed by one-dimensional gene ontology enrichment for cellular compartment

730   and keywords for each of the clusters. Distribution of systematically enriched clusters is shown

731   as the geometric mean at 95% confidence interval for each respective term in non-diabetic (ND,

732   N = 5) and type 2 diabetics (T2D, N = 5) with centerline at the geometric mean with 95%

733   confidence interval.

734   **Figure 5: Lipidomics differential analysis.** A-B) Shotgun lipidomics covering a variety of lipid

735   classes: Ceramides (Cer), Diacylglycerols (DAG), Lysophosphatidylcholines (LPC),

736   Lysophosphatidylethanolamines (LPE), Phosphatidylcholines (PC), Ether-linked

737   Phosphatidylcholines (PC O-), Phosphatidylethanolamines (PE), Ether-linked

738   Phosphatidylethanolamines (PE O-), Phosphatidylinositols (PI), Sphingomyelins (SM),

739   Triacylglyerols (TAG). Volcano plots represent comparisons of plasma lipid levels between ND

740   and T2D PPP. The X-axis shows direction and magnitude of the change; the Y-axis represents

741   the statistical significance of the change. Each point is a lipid species, colored by lipid class to

742   highlight class-specific trends. C) Targeted lipidomics on dihydroceramides (DH Cer),

743   ceramides (Cer) and Sphingoid bases (SB). Each heatmap column represents the comparisons

744   of plasma levels between ND and T2D PPP. Heatmap colors represent direction and magnitude

745   of the change. Log$_2$ Fold Change: ratio of mean lipid concentration in the two groups, log$_2$

746   transformed. Statistical model used for all panels: linear regression with age, sex and BMI as

747   covariates (p: p value); adjustment of p values across all lipid species by the Benjamini-

748   Hochberg method (adj. p: adjusted p value). T2D: Type 2 Diabetes; T3cD: Type 3 Diabetes; ND

749   & PD: non-diabetic and pre-diabetic (with impaired fasting glucose and/or impaired glucose

750   tolerance) with HbA1c<6.0.

751   **Figure 6: Multiblock data modeling of HbA1c.** A) Bar plot showing the variable importance in

752   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°C either natively or embedded in TissueTek OCT compound. Estimated warm and cold ischaemia time was on

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

788 and serum.

789 Transcriptomics

790 Islet procurement and RNA isolation

791 Pancreatic tissue was sectioned in a cryostat and mounted on UV pre-treated Zeiss

792 MembraneSlide 1.0 PEN slides. Laser capture microdissection (LCM) was done with a Zeiss

793 Palm MicroBeam system using autofluorescence to identify islets, as previously described[44].

794 RNA was isolated from approximately 20x6μm3 of islet tissue using the Arcturus PicoPure RNA

795 Isolation Kit. Only preparations with RNA Integrity Number ≥5 were used for RNA sequencing.

796 The entire handling of the tissue samples was done in a strictly RNAse free environment.

797 Library preparation, RNA Sequencing and alignment

798 Sequencing libraries were prepared from bulk RNA using the Illumina SmartSeq protocol. Single

799 ended 76bp sequencing was done with an Illumina HiSeq 2500 or Illumina HiSeq 500 at the

800 Next Generation Sequencing Core Facility of the CMCB Dresden, with the target depth of 35

801 million fragments per library. From FASTQ files, purity-filtered reads were trimmed with

802 Cutadapt to remove adapters and low-quality sequences (v. 1.8)[45]. Reads matching to

803 ribosomal RNA sequences were removed with fastq_screen (v. 0.11.1)[46]. Remaining reads

804 were further filtered for low complexity with reaper (v. 15-065)[47]. Reads were aligned against

805 Homo sapiens GRCh38.92 genome using STAR (v. 2.5.3a)[48]. The number of read counts per

806 gene locus was summarized with htseq-count (v. 0.9.1)[49] using Homo sapiens GRCh38.92 gene

807 annotation. Quality of the RNA-seq data alignment was assessed using RSeQC (v. 2.3.7)[50].

808 RNA Sequencing quality control, processing and differential expression analysis

809 RNA Sequencing datasets were screened for exocrine contamination in an initial quality control

810 (QC) step. Analysis of the absolute number of detected expressed genes, gene body coverage

33

811   and cumulative gene diversity assessment flagged a number of libraries to be of insufficient

812   quality for downstream analysis. Libraries were filtered for minimal expression by removal of

813   genes with less than 5 mean raw reads. Reads were normalized for library size and transformed

814   for variance stabilizing using tools from the DESeq2 Bioconductor package[51]. Further analysis

815   revealed 41 libraries in which transcripts other than insulin (*INS*) displayed the highest

816   normalized number of reads. Differential expression analysis across the clinical categories (ND,

817   IGT, T3cD, T2D) was performed using limma function with voom approach from the limma

818   Bioconductor package[52,53] on both the full dataset of 133 libraries which passed the QC analysis

819   as well as on the "restricted" dataset of 92 libraries featuring *INS* as the highest expressed gene

820   based on the linear model with age, sex and BMI as covariates. All analysis pertaining

821   transcriptomic data was done on R platform (version 3.6.3).

822   Gene set enrichment analysis of differentially expressed genes

823   Functional enrichment analyses of differentially expressed genes in IGT, T2D or T3cD

824   compared to ND patients were performed by weighted gene set enrichment analysis (GSEA) on

825   unfiltered gene lists ranked by decreasing differential expression test statistics. Gene Ontology

826   (GO) term and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway collections were

827   restricted to gene sets with a minimum and maximum sizes of 100 and 500, respectively. The

828   enrichment scores were normalized by gene set size and their statistical significance was

829   assessed by permutation tests (n=1,000). GO enrichment analyses were carried out using the

830   gseGO function from the R package clusterProfiler (version 3.10.1)[54]. GO terms enriched in at

831   least one comparison were identified using *p* value and normalized enrichment score thresholds

832   < 0.01 and > 2.5, respectively. Redundancy of enriched GO terms was removed using the

833   clusterProfiler simplify function (selecting the most representative term by *p* value) and

834   enrichment maps generated using the emapplot function from the R package enrichplot (version

835   1.2.0). KEGG pathway enrichment analyses were performed using the clusterProfiler gseKEGG

34

836   function. Results were filtered based on a *p* value threshold < 0.01 and a normalized enrichment

837   score threshold > 2. To simplify results visualization and interpretation, redundant KEGG

838   pathways were also collapsed into fewer biological themes using the enrichment map

839   visualizations.

Weighted Gene Correlation Network Analysis

840

Gene Co-expression Network Construction

841

842   The gene co-expression network was created following the weighted gene correlation network

843   analysis (WGCNA) protocol as implemented in the WGCNA package in R (version 1.68)[23], as

844   previously described[14]. WGCNA was performed on batch-corrected, normalized and variance

845   stabilizing transformed expression data from the full cohort of 133 subjects. The co-expression

846   network was constructed by calculating an adjacency matrix using Pearson correlation, pairwise

847   complete observations and unsigned method. The soft-threshold parameter was optimized with

848   the function pickSoftThreshold and the best threshold ($\alpha = 7$) selected by visual inspection. The

849   adjacency matrix was then computed into a topological overlap matrix (TOM), converted to

850   distances, and clustered by hierarchical clustering using average linkage clustering. Modules

851   were identified by dynamic tree cut using the hybrid method and parameters minClusterSize=20

852   and deepSplit=2. Similar modules were merged using a module eigengene distance of 0.15 as

853   the threshold.

Identification of co-expressed gene modules

854

855   We correlated the module eigengenes to clinical traits using Spearman correlation (pairwise

856   complete observations) and calculated the corresponding *p* values using the cor and

857   corPvalueStudent functions from the WGCNA package, respectively. Module-trait correlations

858   were represented as heatmap using the labeledHeatmap function from the WGCNA package.

859   The modules displaying the most positive or negative correlation to HbA1c were further

860 analysed. Normalized and variance stabilizing transformed gene counts for selected modules

861 were plotted as heatmap using the heatmap.2 function from the R gplots package (version

862 3.0.1.2). Rows (representing genes) were scaled and hierarchically clustered by Euclidean

863 distances. Columns, representing patients, were custom ordered as described in the legend of

864 figure 3. Module hub genes, such as highly connected genes within a module that could have a

865 strong influence on a phenotypic trait, were identified as those with highest correlations with the

866 particular trait and highest correlations with the module eigengene.

867 Significance of gene co-expression modules

868 We tested the significance of the co-expression modules by comparing their intramodular

869 connectivity (connectivity between nodes within the same module, as computed by the WGCNA

870 intramodularConnectivity function) to the background as follows. For each selected module of

871 size N, we calculated a Z-score as in equation 1:

872 $Z=(k-\mu)/\sigma$ (1)

873 where k is the intramodular connectivity and μ and σ are the mean and standard deviation of the

874 intramodular connectivity from 1,000 randomly sampled modules of size N respectively.

875 Empirical p values were then calculated as the fraction of random intramodular connectivity

876 values ≥ to the observed intramodular connectivity. For the modules with the highest variable

877 importance in projection score in the HbA1c multiblock model, all of the random intramodular

878 connectivity values were below the observed intramodular connectivity, suggesting that these

879 modules were more compact than modules assembled by randomly sampling the same number

880 of genes from the expression data (Supplementary Table 7).

881 Functional profiles of gene modules most predictive for HbA1c

882 The clusterProfiler enrichKEGG function was used to test for the over representation of selected

883 co-expressed gene modules in KEGG pathways using hypergeometric distribution. A *p* value

884 threshold < 0.01 was used to identify enriched terms. Enrichment map visualizations were used

36

885   to overcome gene set redundancy. Results were displayed as networks of enriched pathways

886   and overlapping genes using cytoscape (version 3.5.1).

887   ## Deconvolution analysis

888   In all samples a cell proportions matrix was produced using the R package DeconRNASeq

889   (v.1.26.0) on RPKM-transformed data. The signature file provided to DeconRNASeq comes

890   from Xin et al. (2016)[22], Supplementary Table S2A, obtained using single-cell data. It was

891   adapted to the human genome version 38 by excluding 15 obsolete genes.

892   ## Lipidomics

893

894   ### Sample availability and sample overlap with transcriptomics data

895   Pre-operative plasma lipidomics samples were obtained from a subset of the PPP cohort.

896   Shotgun lipidomics analysis was performed on plasma from 55 PPP. These included 53

897   subjects who also had their islet transcriptomics profile included in this study plus two PPP who

898   were not part of the transcriptomics analysis because the RNA-Seq data failed to pass the

899   quality control. Moreover, targeted sphingolipid analysis was performed on plasma from 101

900   PPP. These included 98 PPP whose transcriptomics data was also included in this study plus

901   three PPP whose RNA-Seq data was excluded for quality reasons. The number of samples in

902   the two types of lipidomics analysis was smaller than in islet transcriptomic analysis because of

903   the limited availability of plasma samples. The 55 PPP with shotgun lipidomics data were a

904   subset of the 101 PPP with targeted sphingolipid data, with the difference in sample numbers

905   being determined by plasma sample availability as well.

906

907   ### Shotgun lipidomics measurements

908   A streamlined mass-spectrometry (MS) -based platform for shotgun lipidomics developed by

909   Lipotype GmbH (Dresden, Germany) was used for lipidomic profiling of patient plasma samples.

910   Lipid extraction, internal standard addition and infusion into the mass spectrometer were

911   performed as previously described[55]. The internal standard mixture contained: cholesterol D6

912   (chol), cholesterol ester 20:0 (CE), ceramide 18:1;2/17:0 (Cer), diacylglycerol 17:0/17:0 (DAG),

913   phosphatidylcholine   17:0/17:0   (PC),   phosphatidylethanolamine   17:0/17:0   (PE),

914   lysophosphatidylcholine 12:0, (LPC) lysophosphatidylethanolamine 17:1 (LPE), triacylglycerol

915   17:0/17:0/17:0 (TAG) and sphingomyelin 18:1;2/12:0 (SM).

916

917   Samples were analyzed by direct infusion in a QExactive mass spectrometer (Thermo

918   Scientific) in a single acquisition. Tandem mass-spectrometry (MS/MS) was triggered by an

919   inclusion list encompassing corresponding MS mass ranges scanned in 1 Da increments. MS

920   and MS/MS data were combined to monitor CE, DAG and TAG ions as ammonium adducts;

921   PC, PC O-, as acetate adducts; and PE, PE O- and PI as deprotonated anions. MS only was

922   used to monitor LPE as deprotonated anion; Cer, SM and LPC as acetate adducts and

923   cholesterol as ammonium adduct.

924   Data post-processing and normalization were performed using an in-house developed data

925   management system. Only lipid identifications with a signal-to-noise ratio >5 and a signal

926   intensity 5-fold higher than in corresponding blank samples were considered for further analysis.

927   The median coefficient of lipid subspecies variation (RSD), as accessed by the repeated

928   analysis of reference samples, was 7.5%.

929   Targeted sphingolipid measurements

930   Ceramides (C16:0 cer, C18:0 cer, C18:1 cer, C20:0 cer, C22:0 cer, C24:0 cer and C24:1 cer),

931   Dihydroceramides (C16:0 DHcer, C18:0 DHcer, C18:1 DHcer, C20:0 DHcer, C22:0 DHcer,

932   C24:0 DHcer,C24:1 DHcer) and precursors (Sphingosine, Sphinganine, 1-Deoxysphinganine,1-

38

933   Methyldeoxysphinganine, SB) were quantified in plasma by liquid chromatography tandem

934   mass spectrometry (LC-MS/MS). In addition to samples, seven-point calibration curves and 3

935   levels of quality controls were made from pure standards in BSA 5%. Finally, reference plasma

936   spiked with analytes at two different levels were prepared as additional QC samples.

937   Lipid chromatographic separation was performed on a UPLC I-Class system (Waters) equipped

938   with an Acquity BEH C18, 100 x 2.1 mm, 1.7 μm column (Waters) heated at 60°C. Mobile phase

939   were A: 0.1 % (V/V) formic acid in water and B: 0.1 % (V/V) formic acid in acetonitrile/

940   isopropanol (60/40). Flow rate was set à 0.5 ml/min and a gradient was applied as follows:

941   0min: 45% A, 2min: 45% A, 3min: 15% A, 13min: 0% A, 14min: 45% A, 16 min: 45%A Mass.

942   Mass analysis was performed on an API 6500 system (Sciex) operating with an electrospray

943   source in positive mode. General parameters were set as follows: curtain gas: N2 (35 PSI), Ion

944   source gas 1: Air (50 PSI), Ion source gas 2: Air (50 PSI), ion source voltage: 5500 V,

945   temperature: 300°C, collision gas: N2 (7). Scheduled multiple reaction monitoring (MRM) mode

946   was used with a target scan time of 0.5s and an MRM detection window of 60s.

947

948   Data was acquired using Analyst 1.6.2 (Sciex) and data processing was performed with

949   MultiQuant 3.0 (Sciex). Peak area of analyte and internal standard were determined by the

950   MultiQuant 3.0 (Sciex) integration system. Analyte concentrations were determined using the

951   internal standard method. The standard curves were generated from the peak area ratios of

952   analyte/internal standard using linear regression analysis with 1/x2 weighting (except for C24

953   cer: quadratic regression analysis). Quantifications of analytes were accepted based on quality

954   control samples. A tolerance of 25% and 30% was applied for accuracy and precision of QC

955   samples and spiked plasma samples, respectively. All concentrations were reported in ng/mL.

956   Internal standards used are listed in the Supplementary Table 16.

957 Analysis of shotgun lipidomics and targeted sphingolipid data

958 The statistical analyses of the shotgun lipidomics and targeted sphingolipid data sets were kept

959 separate. Identical analysis steps were applied to the two data sets. Both sets had missing data

960 values. Lipid species with ≥25% missing values across all available plasma samples were

961 removed from the data set. This filtering resulted in 113 lipid species that were kept in the

962 shotgun data set (523 were removed) and 14 in the targeted data set (4 were removed). For the

963 lipids that remained in the data sets, missing values were imputed using a random forest

964 approach, applying the function missForest from the R package missForest, with default

965 parameters. In a next step, samples were filtered based on subject characteristics: individuals

966 with bilirubin levels ≥100 μmol/l were removed before all analysis; moreover, individuals

967 categorized as IGT with an HbA1c≥6% were excluded from the group comparisons in

968 differential analysis, but they were retained in other analyses involving lipidomics data. In

969 differential analysis, due to the limited number of available ND samples, the ND and the

970 included IGT samples were combined into a single group for comparison with other sample

971 groups, as described in the result section.

972

973 For differential analysis, linear models were applied, using the function lm from the R stats

974 package. For each comparison between two sample groups, a linear model that included

975 diabetes status as the main explanatory variable and age, sex and BMI as covariates was fitted

976 to the data from the two groups. P values for diabetes status were adjusted across all included

977 lipid species with the Benjamini-Hochberg method, separately for each comparison. In addition,

978 ANCOVA results from the three groups T2D, T3cD and ND (as defined above) with the same

979 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,

41

1006   which reflects the importance of variables both with respect to the response and to the

1007   projection quality. The most relevant features were selected using a VIP threshold > 1.2.

Proteomics

1008

Sample Preparation

1009

1010   Pooled pancreatic islet cells with an approximate surface area of 80,000 µm^2 were collected via

1011   Laser Capture Microdissection (LCM) onto adhesive cap tubes. Isolates were reconstituted in a

1012   20 µl lysis buffer (PreOmics, Germany) and transferred into PCR tubes[58]. Samples were boiled

1013   at 95°C for 1min to denature proteins and reduce and alkylate cysteines without shaking in a

1014   thermocycler (Eppendorf GmbH) followed by sonication at maximum power (Bioruptor,

1015   Diagenode, Belgium) for 10 cycles of 30sec sonication and 30sec cooldown each. Sample liquid

1016   was briefly spun down and boiled again for 10min without shaking. 20µl of 100mM TrisHCl pH

1017   8.5 (1:1 v/v) and 20ng Trypsin/LysC were added to each sample, followed by overnight

1018   digestion at 30°C without shaking. The next day, 40µl 99% Isopropanol 5% Trifluoroacetic acid

1019   (TFA) (1:1 v/v) was added to the solution and mixed by sonication. Samples were then

1020   subjected to stage-tip cleanup via styrenedivinylbenzene reversed-phase sulfonate (SDB-RPS).

1021   The sample liquid was loaded on one 14-gauge stage-tip plug. Peptides were cleaned up with

1022   2x200µl 99% Isopropanol 5% TFA and 2x200µl 99% ddH2O 5% TFA in an in-house made

1023   Stage-tip centrifuge at 2,000xg, followed by elution in 40µl 80% Acetonitrile, 5% Ammonia and

1024   dried at 45°C in a SpeedVac centrifuge (Eppendorf, Concentrator plus) according to the 'in-

1025   StageTip' protocol (PreOmics, Germany). Peptides were resuspended in 0.1% TFA, 2% ACN,

1026   97.9% ddH2O.

Liquid chromatography and mass spectrometry (LC-MS)

1027

1028   LC-MS was performed with an EASY nanoLC 1200 (Thermo Fisher Scientific) coupled online to

1029   a trapped ion mobility spectrometry quadrupole time-of-flight mass spectrometer (timsTOF Pro,

1030   Bruker Daltonik GmbH, Germany) via nano-electrospray ion source (Captive spray, Bruker

1031   Daltonik GmbH). Peptides were loaded on a 50cm in-house packed HPLC-column (75µm inner

1032   diameter packed with 1.9µm ReproSil-Pur C18-AQ silica beads, Dr. Maisch GmbH, Germany).

1033   Sample analytes were separated using a linear 120min gradient from 5-30% buffer B in 95min

1034   followed by an increase to 60% for 5min, and by a 5min wash at 95% buffer B at 300nl/min

1035   (Buffer A: 0.1% Formic Acid, 99.9% ddH2O; Buffer B: 0.1% Formic Acid, 80% CAN, 19.9%

1036   ddH2O). The column temperature was kept at 60°C by an in-house manufactured oven.

1037   Mass spectrometry analysis was performed in a data-dependent PASEF mode with 1 MS1

1038   survey TIMS-MS and 10 PASEF MS/MS scans per acquisition cycle. Ion accumulation and

1039   ramp time in the dual TIMS analyzer was set to 100ms each and we analyzed the ion mobility

1040   range from $1/K_0$ = 1.6 Vs cm^{-2} to 0.6 Vs cm^{-2}. Precursor ions for MS/MS analysis were isolated

1041   with 2Th windows for m/z<700 and 3Th for m/z>700 in a total m/z range of 100-1,700 by

1042   synchronizing quadrupole switching events with the precursor elution profile from the TIMS

1043   device. The collision energy was lowered linearly as a function of increasing mobility starting

1044   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

1045   were excluded with a polygon filter (otof control, Bruker Daltonik GmbH). Precursors for MS/MS

1046   were picked at an intensity threshold of 2.500 a.u. and resequenced until reaching a 'target

1047   value' of 20,000 a.u taking into account a dynamic exclusion of 40sec elution[25].

1048   Before MS analysis, the LC-MS setup was subjected to a rigorous quality control procedure.

1049   These criteria included protein- and peptide-identifications as well as general technical

1050   specifications like chromatography performance. If those thresholds were met (>5.500 protein

1051   groups, >38.000 peptides from 200 ng tryptic HeLa digest, chromatographic peak FWHM of <=9

1052   sec and peak base-to-base width <=17 sec on a 120 min liquid chromatography gradient;

1053   Quantitative reproducibility across two subsequent QC runs with a Pearson correlation of >0.97

1054   and coefficients of variation of <=10%) the project measurements were initiated. Furthermore,

1055   we subject our instruments to a rigorous weekly maintenance procedure (Maintenance of the

1056   Liquid chromatography platform and re-calibration of the mass spectrometer) to ensure highest

1057   overall performance and reproducibility.

1058   Proteomics raw file processing

1059   Raw files were searched against the human Uniprot databases (UP000005640_9606.fa,

1060   UP000005640_9606_additional.fa) MaxQuant (Version 1.6.7), which extracts features from

1061   four-dimensional isotope patterns and associated MS/MS spectra[59]. False-discovery rates were

1062   controlled at 1% both on peptide spectral match (PSM) and protein level. Peptides with a

1063   minimum length of seven amino acids were considered for the search including N-terminal

1064   acetylation   and   methionine   oxidation   as   variable   modifications   and   cysteine

1065   carbamidomethylation as fixed modification, while limiting the maximum peptide mass to 4,600

1066   Da. Enzyme specificity was set to trypsin cleaving c-terminal to arginine and lysine. A maximum

1067   of two missed cleavages were allowed. Maximum precursor and fragment ion mass tolerance

1068   were searched as default for TIMS-DDA data, while the main search peptide tolerance was set

1069   to 20ppm. The median absolute mass deviation for the data was 0.68ppm. Peptide

1070   identifications by MS/MS were transferred by matching four-dimensional isotope patterns

1071   between the runs with a 0.7-min retention-time match window and a 0.05 $1/K_0$ ion mobility

1072   window[60]. Label-free quantification was performed with the MaxLFQ algorithm and a minimum

1073   ratio count of 1[61].

1074   Bioinformatic analysis

1075   Bioinformatics analysis was performed in Perseus (version 1.6.7.0 and 1.5.5.0) and GraphPad

1076   Prism (version 8.2.1)[62]. Reverse database, contaminant, and only by site modification

1077   identifications were removed from the dataset. Data were grouped by analytical replicates and

1078   filtered to at least 70% data completeness in one group. Missing values were imputed from a

44

1079   data table specific normal distribution estimate with a downshift of 1.8 and a width of 0.3

1080   standard deviations after $\log_2$-transformation of the data. To represent the data reproducibility

1081   and variability, a principal component analysis was performed on the median data of analytical

1082   replicate measurements of each individual. Clinically classified T2D and ND individuals were

1083   tested for differences in their mean by a two-sided Student's t-test with S0=0.1 and a Benjamini-

1084   Hochberg correction for multiple hypothesis testing at an FDR of 0.05 preserving grouping of

1085   each individuals analytical replicate measurements, and presented as volcano plot. We then

1086   normalized the data by row-wise z-scoring followed by hierarchical clustering using Euclidean

1087   as the distance parameter for column- and row-wise clustering. 1D gene ontology enrichments

1088   of clustered and systematically changed proteins were performed with regards to their cellular

1089   compartment and keywords assignment[30]. Log_2 transformed LFQ data were used for the

1090   calculation of intensity shifts of the enriched keyword or cellular compartment term for each of

1091   the displayed clusters. Total protein copy number estimation of the median LFQ intensities for

1092   patients clinically classified as non-diabetic and diabetic were calculated using the Perseus

1093   plugin 'Proteomic ruler'[29]. Median LFQ intensity values for all T2D and ND were calculated. We

1094   annotated protein groups for the leading protein ID with the human Uniprot fasta file

1095   (UP000005640_9606.fa) and estimated the protein copy number with the following settings:

1096   Averaging mode. 'All columns separately', Molecular masses: 'Average molecular mass',

1097   Detectability correction: 'Number of theoretical peptides', Scaling mode: 'Histone proteomic

1098   ruler', Ploidy: '2', Total cellular protein concentration: '200g/l'. Proteins were annotated with

1099   regards to their cellular compartment by gene ontology. We calculated the median protein copy

1100   number for the samples from T2D and ND PPP separately and multiplied it by its protein mass.

1101   To calculate the subcellular protein mass contribution, we calculated the protein mass

1102   proportion for the GOCC terms 'Nucleus', 'Mitochondrion', 'Cytoskeleton', 'Golgi apparatus', and

1103   'Endoplasmic reticulum'. For calculating the organellar change between T2D and ND PPP,

1104   protein mass contributions of each organelle were normalized by its respective 'Nuclear part'

1105   contribution. Chromosomal annotation of significantly changed proteins between T2D and ND

1106   PPP was identified via Ensembl ID. For transcriptome to proteome correlation, the gene

1107   intersection of both data sets was scaled to 1E6 units, followed by log10-transformation.

Antibody validation

1108

1109   Rabbit polyclonal anti-ALDOB antibody (Proteintech, Cat.No. 18065-1-AP) was tested for

1110   specificity by western blotting of protein extracts of $ALDOB^{-/-}$ MIN6 cells generated with a

1111   CRISPR/Cas9 system, as described[63]. Primary antibodies against ALDOB, ALDOA (Abnova,

1112   Cat. No. H00000226-M01) and gamma tubulin as loading control (Sigma Aldrich, Cat.No. T-

1113   6557) were diluted in 5% non-fat milk 1:2000, 1:1000 and 1:5000, respectively. The knock-out of

1114   *ALDOB* was verified by Sanger sequencing of the target locus.

Isolated mouse islet and cell line experiments

1115

1116   Mouse (C57BL/6J, db/db (BKS.Cg-Dock7^{m} +/+ Lepr^{db}/J) and db/+ (Charles River Laboratories),

1117   3 animals/strain, male, age 13 weeks) islets were cultured for 1 day post isolation. Islet beta

1118   MIN6c4 (MIN6 clone 4, from Osaka University under Material License Agreement) and alpha

1119   αTC1-clone 6 (ATCC, CRL-2934) cell lines were harvested for RNA extraction using Qiagen

1120   RNeasy Mini Kit according to the manufacturer's instructions. After quality control, RNA samples

1121   were sequenced using the Illumina HiSeq 2000 platform and processed as previously

1122   described[51,64,65]. All animal experiments were done in accordance with the ethical approval of

1123   the Sanofi-Aventis Animal Welfare Office, Frankfurt/Main, Germany. The animals were housed

1124   at 20-24°C, by 45-65 % humidity setting in an artificial day / night (12hrs) rhythm.

Immunofluorescence microscopy

1125

1126   Immunofluorescence staining was done on formalin-fixed paraffin embedded 5µm thick sections

1127   of human pancreatic tissue. Acetylated histone H3 and H4 were detected in separate sections

1128 using rabbit polyclonal antibodies (Merck Millipore Cat.No. 06-598 and 06-599, respectively,

1129 dilution 1:100). A mouse monoclonal anti-insulin antibody (Thermo Fisher Scientific Cat.No. 53-

1130 9769-82, dilution 1:200) was used for co-staining, to identify the beta cell areas. Images were

1131 acquired using a Nikon C2+ confocal microscope with a 60x oil immersion objective, with

1132 acquisition parameters normalized to a negative control sample.

1133 Materials and methods references

1134 44. Sturm, D. *et al.* Improved Protocol For Laser Microdissection Of Human Pancreatic Islets

1135   From Surgical Specimens. *J. Vis. Exp.* 2–7 (2013). doi:10.3791/50231

1136 45. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing

1137   reads. *EMBnet.journal; Vol 17, No 1 Next Gener. Seq. Data Anal. - 10.14806/ej.17.1.200*

1138   (2011).

1139 46. Wingett, S. W. & Andrews, S. FastQ Screen: A tool for multi-genome mapping and quality

1140   control. *F1000Research* **7**, 1338 (2018).

1141 47. Davis, M. P. A., van Dongen, S., Abreu-Goodger, C., Bartonicek, N. & Enright, A. J.

1142   Kraken: A set of tools for quality control and analysis of high-throughput sequence data.

1143   *Methods* **63**, 41–49 (2013).

1144 48. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21

1145   (2013).

1146 49. Anders, S., Pyl, P. T. & Huber, W. HTSeq-A Python framework to work with high-

1147   throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

1148 50. Wang, L., Wang, S. & Li, W. RSeQC: Quality control of RNA-seq experiments.

1149   *Bioinformatics* **28**, 2184–2185 (2012).

1150 51. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion

1151   for RNA-seq data with DESeq2. *Genome Biol.* **15**, (2014).

1152 52. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing

1153   and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

1154   53.   Smyth, G. K. *et al.* RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR.

1155   *F1000Research* **5**, (2018).

1156   54.   Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: An R package for comparing

1157   biological themes among gene clusters. *Omi. A J. Integr. Biol.* **16**, 284–287 (2012).

1158   55.   Surma, M. A. *et al.* An automated shotgun lipidomics platform for high throughput,

1159   comprehensive, and quantitative analysis of blood plasma intact lipids. *Eur. J. Lipid Sci.*

1160   *Technol.* **117**, 1540–1549 (2015).

1161   56.   Smilde, A. K., Kiers, H. A. L., Bijlsma, S., Rubingh, C. M. & Van Erk, M. J. Matrix

1162   correlations for high-dimensional data: The modified RV-coefficient. *Bioinformatics* **25**,

1163   401–405 (2009).

1164   57.   Bylesjö, M., Rantalainen, M., Nicholson, J. K., Holmes, E. & Trygg, J. K-OPLS package:

1165   Kernel-based orthogonal projections to latent structures for prediction and interpretation

1166   in feature space. *BMC Bioinformatics* **9**, (2008).

1167   58.   Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated

1168   proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat.*

1169   *Methods* **11**, 319–324 (2014).

1170   59.   Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized

1171   p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.*

1172   **26**, 1367–1372 (2008).

1173   60.   Prianichnikov, N. *et al.* Maxquant software for ion mobility enhanced shotgun proteomics.

1174   *Mol. Cell. Proteomics* **19**, 1058–1069 (2020).

1175   61.   Cox, J. *et al.* Accurate proteome-wide label-free quantification by delayed normalization

1176   and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **13**, 2513–

1177   2526 (2014).

1178   62.   Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of
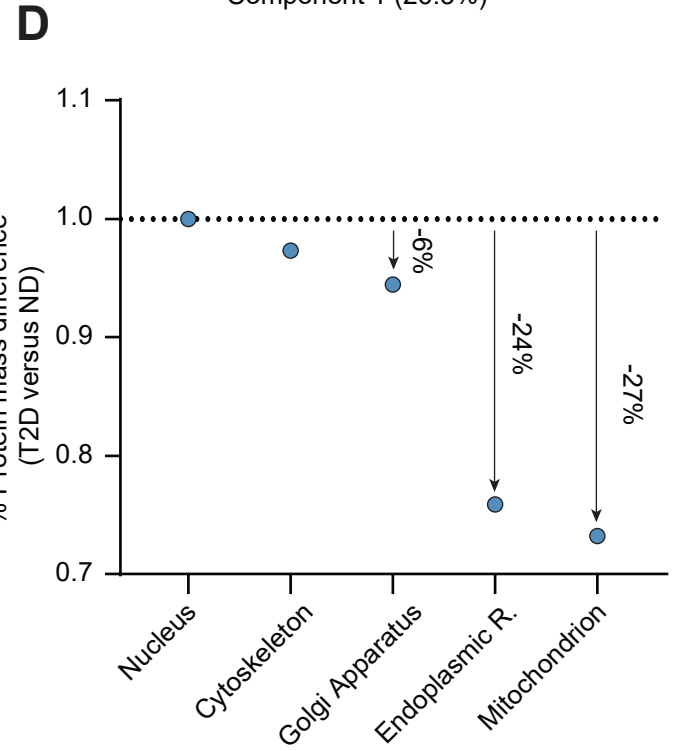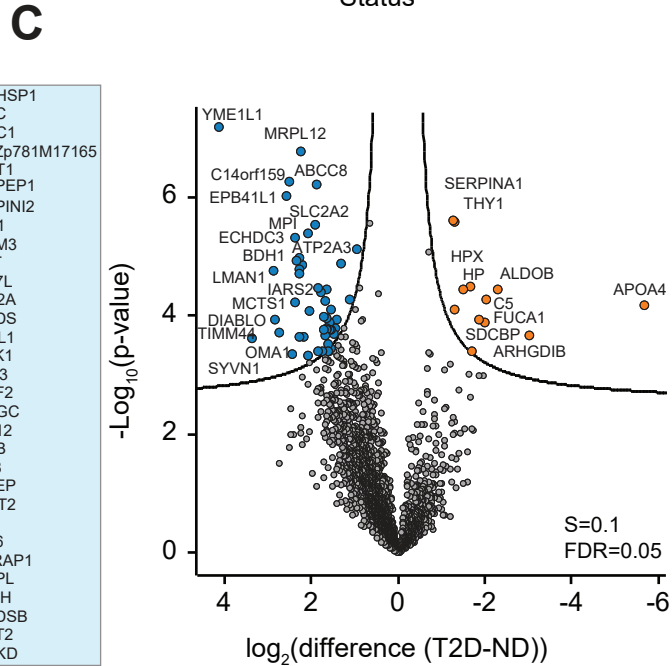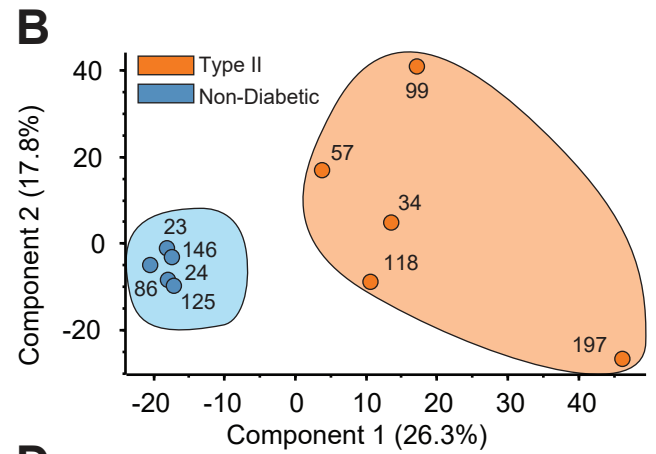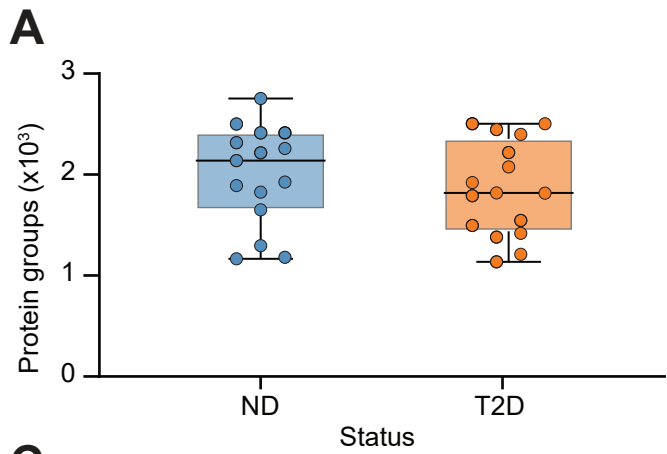
1179   (prote)omics data. *Nature Methods* **13**, 731–740 (2016).

1180   63.   Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**,

1181   2281–2308 (2013).

1182   64.   Hu, J., Ge, H., Newman, M. & Liu, K. OSA: A fast and accurate alignment tool for RNA-

1183   Seq. *Bioinformatics* **28**, 1933–1934 (2012).

1184   65.   Li, B., Ruotti, V., Stewart, R. M., Thomson, J. A. & Dewey, C. N. RNA-Seq gene

1185   expression estimation with read mapping uncertainty. *Bioinformatics* **26**, 493–500 (2009).

1186

49

A

B

C

A

B

C

D

A — Module-trait relationships

B — Module M9 genes (n=88); Diabetes status: ND, IGT, T3cD, T2D; Scaled gene expression

C — Module M14 genes (n=121); Diabetes status: ND, IGT, T3cD, T2D; Scaled gene expression

D — Correlation to HbA1c vs Correlation to M9; labeled genes: ALDOB, HSD17B13, AL357509.1, FAIM2, CDKN1C, SLC26A4, BTBD17, SPG21, KCNAB2, CERK

E — Correlation to HbA1c vs Correlation to M14; labeled genes: SLC2A2, ARG2, G3BP1, RRAGD, TMEM37, HHATL, AC013275.1, FBXO7, RNF6