Integrated transcriptomics identifies β -cell subpopulations and genetic networks associated with obesity and glycemic control in SM/J mice

Mario A Miranda^{1#}, Juan F Macias-Velasco^{1#}, Heather Schmidt¹, Heather A Lawson^{1##}.

¹Department of Genetics, Washington University School of Medicine, 660 South Euclid Ave, Saint Louis, MO, USA

[#]Authors contributed equally

^{##}Corresponding author

660 South Euclid Ave

Campus Box 8232

Saint Louis, MO, 63110

ph: 314-362-7269, fax: 314-362-7855

Keywords: single-cell RNAseq, bulk RNAseq, hyperglycemia, insulin, β -cell heterogeneity, diabetes, obesity, mouse model

52 Abstract

53 Understanding how heterogeneous β -cell function and stress response impact diabetic etiology is imperative for
 54 therapy development. Standard single-cell RNA sequencing analysis illuminates some genetic underpinnings
 55 driving heterogeneity, but new strategies are required to capture information lost due to technical limitations. Here,
 56 we integrate pancreatic islet single-cell and bulk RNA sequencing data to identify β -cell subpopulations based on
 57 gene expression and characterize genetic networks associated with β -cell function in high- and low-fat fed male and
 58 female SM/J mice at 20 and 30wks of age. Previous studies have shown that high-fat fed SM/J mice resolve
 59 glycemic dysfunction between 20 and 30wks. We identify 4 β -cell subpopulations associated with insulin secretion,
 60 hypoxia response, cell polarity, and stress response. Relative proportions of these cells are influenced by age, sex,
 61 and diet. Network analysis identifies fatty acid metabolism and β -cell physiology gene expression modules
 62 associated with the hyperglycemic-obese state. We identify subtype-specific expression of *Pdyn* and *Fam151a* as
 63 candidate regulators of genetic pathways associated with β -cell function in obesity. In sum, this study uses a novel
 64 data integration method to explore how β -cells respond to obesity and glycemic stress, helping to define the
 65 relationship between β -cell heterogeneity and diabetes, and shedding light on novel genetic pathways with
 66 therapeutic potential.

67

68

69

70

71

72

73

74

75

76

77

78 Introduction

79 Proper insulin secretion from pancreatic β -cells is required to maintain glycemic control. Obesity initially
80 promotes β -cell expansion, but prolonged glycemic stress and inflammation drive β -cell death and dysfunction,
81 resulting in type 2 diabetes¹⁻³. Without sufficient β -cell mass and insulin production, sustained hyperglycemia
82 increases risk for deadly metabolic diseases⁴⁻⁶. Currently, transplantation of cadaveric islets is the only method for
83 restoring β -cell mass in diabetes but is severely limited by donor availability and requires lifelong
84 immunosuppressant therapy^{7,8}. Differentiating induced pluripotent stem cells into insulin secreting cells may
85 alleviate this bottleneck, but current methods fail to recapitulate fine-tuned glucose sensing and insulin secretion *in*
86 *vivo*^{9,10}. There is urgent need for therapies that improve endogenous β -cell function. This requires understanding
87 how and why β -cells become dysfunctional in obesity.

88 Improving endogenous β -cell function in diabetes is complicated by cellular heterogeneity, because
89 individual β -cells vary significantly in function, gene expression, protein level, and stress response¹¹. Single cell
90 technologies permit interrogation of the molecular underpinnings of heterogeneity, addressing two fundamental
91 questions: Do functionally distinct subpopulations of β -cells exist? Can accounting for heterogeneity improve
92 diabetic therapy? Several research groups have proposed subpopulations based on clustering analysis using single
93 cell RNA sequencing (scRNA-seq), however, there is little agreement among studies¹²⁻¹⁵. High rates of gene dropout
94 and low read depth contribute to these problems, necessitating approaches that improve information capture without
95 losing cell type-specific information^{16,17}.

96 Integrating sc-RNAseq with bulk RNAseq data leverages bulk sequencing's high read depth, allowing for
97 capture of lowly expressed genes and robust expression analysis. Several tools can integrate sc- and bulk RNA-seq
98 data, focusing on deconvoluting bulk RNAseq data from heterogeneous tissue to account for differences in tissue
99 composition^{12,18,19}. These methods estimate and control for cell type abundance but do not identify cell type-specific
100 expression signatures in bulk datasets. Analytical strategies that identify cellular gene expression signatures in bulk
101 RNAseq data allow for robust cell type-specific differential expression analysis and complex network analysis that
102 is currently not feasible in scRNA-seq data.

Here, we characterize gene expression in β -cells from obese SM/J mice, who spontaneously transition from hyperglycemic to normoglycemic with improved β -cell function between 20 and 30 weeks of age^{20,21}. We assign functional identities to 4 subpopulations of β -cells using scRNA-seq. These subpopulations vary in proportion among hyperglycemic-obese, normoglycemic-obese, and normoglycemic-lean mice. We identify 316 genes specifically expressed by β -cells and establish a β -cell gene expression profile for each mouse. We leverage this information to focus our analyses of bulk-islet RNAseq and identify β -cell-specific differential expression and gene networks associated with the hyperglycemic-obese, normoglycemic-obese, and normoglycemic-lean states. Two novel potential regulators of β -cell function, *Pdyn* (Prodynorphin) and *Fam151a* (Family with sequence similarity 151 member A), are differentially expressed, highly connected within genetic networks, and primarily expressed by β -cell subpopulations associated with normoglycemia. This analysis demonstrates that integrating scRNAseq with bulk RNAseq is a powerful approach for exploring β -cell heterogeneity and identifying key genes and subpopulations that strongly associate with glycemic state. The genetic networks and β -cell subpopulation signatures we identify have high potential to lead to further research aimed at improving β -cell function in obesity.

Methods

Metabolic phenotyping. SM/J mice were obtained from The Jackson Laboratory (Bar Harbor, ME). Mouse colony was maintained at the Washington University School of Medicine and all experiments were approved by the Institutional Animal Care and Use Committee in accordance with the National Institutes of Health guidelines for the care and use of laboratory animals. Mice were weaned onto a high-fat diet (42% kcal from fat; Envigo Teklad TD88137) or isocaloric low-fat diet (15% kcal from fat; Research Diets D12284), as previously described²¹. At 20 or 30 weeks of age, mice were fasted for 4 hours, body weight measured, and blood glucose was measured via glucometer (GLUCOCARD). Mice were injected with sodium pentobarbital, followed by a firm toe pinch to ensure unconsciousness. Blood was collected via cardiac puncture and pancreas was collected. Blood was spun at 6000 rpm at 4°C for 20 minutes to collect plasma. Insulin ELISA (ALPCO 80-INSMR-CH01) was used to measure plasma insulin levels following manufacturer's instructions.

128 *Islet isolation and phenotyping.* Islets were isolated from pancreas as previously described²⁰ and rested overnight.
 129 5 Islets were equilibrated in KRBH buffer with 2.8 mM glucose for 30 minutes at 37°C, then placed in 150 µl
 130 KRBH containing 2.8 mM glucose at 37°C for 45 minutes, then 150 µl KRBH containing 11 mM glucose at 37°C
 131 for 45 minutes. Islets were then transferred into 150 µl acid ethanol. Islet content and secretion tubes were stored at
 132 -20°C overnight. Experiments were performed in duplicate per individual, and measurements are reported as the
 133 average of replicates. Mouse insulin ELISA (ALPCO 80-INSMU-E01) was performed, with the secretion tubes
 134 diluted 1:5, and content tubes diluted 1:100. Glucose stimulated insulin secretion was calculated by dividing insulin
 135 secretion at 11mM glucose by insulin secretion at 2.8 mM glucose. Basal insulin secretion was calculated by
 136 dividing insulin secretion at 2.8 mM glucose by islet insulin content. Total islet protein was measured using Pierce
 137 BCA Protein Assay kit (Thermo Scientific) according to manufacturer's instructions and read at 562 nm on the
 138 Synergy H1 Microplate Reader (Biotek). Islet insulin content was calculated by dividing the islet insulin level in
 139 the content tubes by total islet protein. All measurements were taken in duplicate, values reported are the average
 140 of replicates.

141 *Single cell RNA sequencing.* Single cell RNA sequencing (scRNA-seq) was performed on islets isolated from 15
 142 SM/J mice representing 6 cohorts: 20wk high-fat females (n=3), 20wk high-fat males (n=3), 30wk high-fat females
 143 (n=3), 30wk high-fat males (n=2), 20wk low-fat females (n=2), and 20wk low-fat males (n=2). Isolated islets were
 144 dissociated into single cell suspensions using Accumax cell/tissue dissociation solution (Innovative Cell
 145 Technologies). Libraries were prepped using the Chromium Single Cell 3' GEM, Library & Gel Bead Kit v3
 146 (10xGenomics) and sequenced at 2x150 paired end reads using a NovaSeq S4. After sequencing, reads were de-
 147 multiplexed and assigned to individual samples. Reads were aligned using 10x Genomics CellRanger (3.1.0) against
 148 our custom SM/J reference²². Samples that were prepped together were aggregated into batches using CellRanger
 149 aggregate. In the R environment (4.0.0), each aggregated batch was run through SoupX (1.5.0)²³ to estimate and
 150 correct for ambient RNA contamination. A contamination fraction of 0.05 was chosen. Removal of *Ins2* ambient
 151 RNA shown in **Supplemental Figure 1D-E**. Adjusted counts were imported into Seurat (3.2.2)²⁴, where cells were
 152 filtered for number of features detected (500-3000), total counts detected (1000-30000), percent mitochondrial

genes (0-30), visualized in **Supplemental Figure 1A**. For additional quality control, we excluded cells where nCount was not predictive of nFeature; the predictive error (residual) of a cell had to be within 3 standard deviations of the mean predictive error (~0). Cell counts for samples from one batch shown before and after filtering step shown in **Supplemental Figure 1B-C**. Expression was then normalized in Seurat (normalization.method = LogNormalize), batches were integrated, and clustered using a shared nearest neighbor approach. Using the top 10 principal components of the filtered expression data and a resolution of 0.14, we identified 9 clusters of cells using Clustree (0.4.3)²⁵. Cell types were assigned by identifying top over-expressed genes for each cluster relative to all other clusters using a Wilcoxon rank sum test, with an average log-fold-change threshold of ≥ 0.25 and requiring at least 25% of cells express the gene. Identities were assigned by comparing top over-expressed genes for each cluster with known cell-type specific markers for islet cells.

Identification of β -cell subtypes, differential expression, and composition. All β -cells were subset for further analysis. Using the top 10 principal components of the filtered expression data and a resolution of 0.09 (**Supplemental Figure 1F**), we identified 4 populations of β -cells, labeled Beta 1-4. Top over expressed genes for each population were identified using a Wilcoxon rank sum test, with an average log-fold-change threshold ≥ 0.1 compared to all other β -cells and an adjusted p-value < 0.01 (Bonferroni), shown in **Table 1**. Subtype-identifier genes were tested for gene set enrichment (see below) to determine presumed function of each population. Using these thresholds (average log-fold-change ≥ 0.1 and adjusted p-value < 0.01), differential expression was calculated between pairs of β -cell subpopulations (**Table 15-20**), across all β -cells between 20wk high-fat males and 30wk high-fat males (**Table 21**), across all β -cells between 20wk high-fat males and 20wk low-fat males (**Table 22**), and across Beta 1 cells between 20wk high-fat males and 20wk low-fat males (**Table 23**) using the “MAST” hurdle-model test²⁶. Relative proportions of β -cell subpopulations in each cohort was estimated using bootstrapping to calculate 95% confidence intervals by randomly sampling 1,000 cells from each cohort for 100 iterations.

Identification of β -cell-specific genes. To identify β -cell-specific gene expression signatures in the bulk sequencing data, we assume that for a given gene, the sum of expression in the scRNA-seq data, Y_{Total} , approximates the expression in bulk RNA sequencing data, Y_{Bulk} .

$$Y_{Bulk} \approx Y_{Total}$$

The Y_{Total} value can be re-written as the sum of expression from all the contributing cell types, where Y is the expression from a given cell type, and N is the total number of cells of that type.

$$E[Y_{Total}] \sim E[Y_{\beta-cell}] * N_{\beta-cell} + E[Y_{\alpha-cell}] * N_{\alpha-cell} + \dots + E[Y_{CellType}] * N_{CellType}$$

Therefore, the expected relative contribution of β -cells ($Q_{\beta-cell}$), to total expression (Y_{Total}), can be written as:

$$Q_{\beta-cell} \sim \frac{E[Y_{\beta-cell}] * N_{\beta-cell}}{E[Y_{Total}]}$$

When $Q_{\beta-cell}$ is high in all 6 cohorts, we are confident that gene expression in the bulk data is nearly exclusively coming from β -cells. To determine the contribution of β -cells and all other cell types to gene expression, the distribution of normalized expression was assessed using a Cullen and Frey graph using the fitdistrplus package (1.1-3)²⁷ in R and identified as beta distributed (**Supplemental Figure 2A-B**).

scRNA-seq data is characterized by a buildup of expression values at 0 due to cells that do not express a gene or due to gene-dropout. This distribution required us to employ a beta-hurdle model described by three parameters: Pr(Dropout), α , and β . We treat the probability of attaining an expression value of 0 (Pr(Dropout)) and the distribution of non-zero values separately for a given gene. We fit the expression of each gene in each cell type using a beta-hurdle model optimized by maximum goodness of fit estimation. To fit the first part of the beta-hurdle mode, the Kolmogorov-Smirnov statistic using the ks.test package (4.05)²⁸ in R was used to select the optimal α and β shape parameters that best fit the expression distribution of non-zero-expressing cells by minimizing the distance between the cumulative distribution of the real data and theoretical beta model data (Ks) (**Supplemental Figure 2C**). To fit the second part of the beta-hurdle mode, Pr(Dropout) was iterated between 0 and 1, selecting the Pr(Dropout) that minimized Ks (**Supplemental Figure 2D**). In most cases, assuming 100% of cells not expressing the gene was due to gene drop out provided the best fit (**Supplemental Figure 2E**). From the best fit beta-hurdle model, the expected value for each gene within a cell type $E[Y_{CellType}]$ can be calculated using the fit α and β shape parameters as:

$$E[Y_{CellType}] = \frac{1}{1 + \left(\frac{\beta}{\alpha}\right)}$$

Multiplying the expected value by the total number of cells of that cell type provides the total expression contribution of that cell type. Summing this value across all cell types provides total expression and allows assessment of the contribution of β -cell gene expression to the genes' total expression. The cell type contribution of all genes in each of the 6 cohorts are shown in **Table 3-8**. We required β -cells to account for $\geq 80\%$ of total gene expression in each of the six cohorts analyzed, resulting in 316 " β -cell-specific genes", shown in **Table 9**.

Bulk RNA sequencing. Islets from 32 mice were sequenced: 4 males and 4 females from each diet (high-fat and low-fat) and each age (20wk and 30wk), $n=32$. Islet RNA was extracted using the RNeasy MinElute Cleanup kit (Qiagen), RNA concentration was measured via Nanodrop and RNA quality/integrity was assessed with a BioAnalyzer (Agilent). Libraries were prepped using the SMARTer cDNA synthesis kit (Takara Bio) and sequenced at 2x150 paired end reads using a NovaSeq S4. After sequencing, reads were de-multiplexed and assigned to individual samples. FASTQ files were trimmed and filtered to remove low quality reads and aligned against a SM/J custom genome using STAR^{22,29,30}. Read counts for β -cell-specific genes were normalized via TMM normalization and pairwise differential expression between cohorts was performed using edgeR³¹. Differential expression analysis for all 316 β -cell-specific genes across select cohort comparisons reported in **Table 11**.

Co-expression network analysis. Weighted Gene Co-Expression Network Analysis (WGCNA) identifies co-expression modules and correlates them with phenotypic traits³². Briefly, edgeR-normalized counts for β -cell-specific genes were converted to standard normal, and an adjacency matrix was created from bi-weight mid-correlations calculated between all genes in all individuals and raised to a power β of 8, chosen based on a scale-free topology index above 0.9, to emphasize high correlations³³. The blockwiseModules function created an unsigned Topological Overlap Measure using the adjacency matrix to identify modules of highly interconnected genes. Eigengenes were calculated as the first principal component for each module, and Pearson's correlations were calculated between eigengene expression and phenotype to estimate module-trait relationships. Module-trait correlations were considered significant at an FDR-corrected p-value < 0.05 . Each gene's module membership and

correlation with metabolic phenotypes are reported in **Table 13**. The adjacency matrix was then used to calculate the connectivity of each gene with other genes within its module (k_{within}). Adjacency matrices were subset for each age x diet x sex cohort, used to calculate cohort-specific connectivity for each gene (**Table 14**), then used to calculate differential connectivity between cohorts. We report differential connectivity between 20wk high-fat males and 30wk high-fat males, HM , calculated as

$$kDiff(HM) = \frac{HM20 - HM30}{HM20 + HM30}$$

and differential connectivity between 20wk high-fat males and 20wk low-fat males, $M20$, calculated as

$$kDiff(M20) = \frac{HM20 - LM20}{HM20 + LM20}$$

where HM20, HM30, and LM20 are cohort-specific k_{within} values for each gene. This provides each gene in the comparison with a value between -1 and 1, where $|kDiff| > 0.5$ is considered differentially connected. Genes with positive differential connectivity are more highly connected in 20wk high-fat males than 30wk high-fat males or 20wk low-fat males, depending on comparison.

Gene set enrichment analysis. Over-representation analysis (ORA) was performed using the WEB-based Gene Set Analysis Toolkit v2019³⁴ on genes overexpressed in individual β -cell subtypes (**Table 2**), β -cell-specific genes (**Table 10**), and genes within blue and brown expression modules (**Table 12**), using all genes expressed in the sc-RNAseq data set as a reference set. Analysis included gene ontologies (biological process, cellular component, molecular function), pathway (KEGG), and phenotype (Mammalian Phenotype Ontology).

Results

SM/J islets contain four β -cell subpopulations. We developed an analysis pipeline that integrates sc- and bulk RNA-seq data to characterize the transcriptional changes in β -cells from obese SM/J mice as they improve glycemic control with age (**Figure 1A**). We identified islet cells (α , β , and δ), exocrine cells (acinar and ductal), vascular cells (smooth muscle and endothelial), and immune cells (B cells and macrophages) (**Figure 1B**). Subsequent clustering

248 of β -cells revealed 4 subpopulations, labeled Beta 1-4 (**Figure 1C, Supplemental Figure 1F**). 20wk hyperglycemic
249 high-fat males have a significantly larger Beta 1 population compared to 30wk normoglycemic high-fat males and
250 20wk low-fat males, at the expense of a diminished Beta 2 population (**Figure 1 D-E**).

251

252 **Figure 1.** SM/J islets contain four subpopulations of β -cells. **(A)** Data analysis pipeline identifies subpopulations
253 of β -cells and integrates bulk and single cell RNA sequencing data. **(B)** Single cell RNA sequencing UMAP plot of
254 islet tissue. VSMC – vascular smooth muscle cell, VEC – vascular endothelial cell. **(C)** Single cell RNA sequencing
255 UMAP of 4 subpopulations of β -cells. **(D)** Bootstrap analysis quantifies relative proportions of β -cell
256 subpopulations across cohorts. Black dot represents estimated proportion, colored bars illustrate 95% confidence
257 interval. Cell type composition estimates with non-overlapping confidence intervals are significantly different. **(E)**

Ternary plot illustrates differences in subtype composition of Beta 1-3 in high-fat SM/J mice. HF20 – 20wk high-fat female, HF30 – 30wk high-fat female, HM20 – 20wk high-fat male, HM30 – 30wk high-fat male, LF20 – 20wk low-fat female, LM20 – 20wk low-fat male.

β-cell subpopulations have unique expression signatures. To identify genes that are overexpressed in each subpopulation, we performed differential expression analysis between each subpopulation and all other β-cells (Table 1). The top ten highest differentially expressed genes within each subtype are visualized in Figure 2. Gene enrichment analysis on the overexpressed genes in each subpopulation reveals potential specialization: Beta 1 – insulin secretion, Beta 2 – hypoxia response, Beta 3 – cell polarity, Beta 4 – stress response (Table 2).

Figure 2. β-cell subpopulations have unique gene expression signatures. (A) Heat map for top 10 differentially expressed genes in each β-cell subpopulation. Yellow is highly expressed; purple is lowly expressed.

SM/J β-cells uniquely express 316 genes. To identify genes primarily expressed by β-cells in scRNA-seq data, we employed a beta hurdle model, which allowed us to estimate the relative contribution of each cell type to total gene expression (Supplemental Figure 2, Tables 3-8). To be considered a β-cell-specific gene, we required β-cells to

272 account for at least 80% of total expression within each cohort (**Figure 3A**). We identified 316 β -cell-specific genes
 273 (**Table 9**), comprised of genes canonically associated with β -cell identity, including *Ucn3*, *Gcgr*, and *Slc2a2*
 274 (**Figure 3B-E**) and genes with unknown function in β -cells (**Figure 3F-G**). Overrepresentation analysis revealed
 275 enrichment for terms associated with β -cell function, including mature onset diabetes of the young (MODY) and
 276 carbohydrate homeostasis, along with terms related to neuron function including neuroactive ligand-receptor
 277 interaction, neuron projection terminus, and axon part (**Table 10**). We then sought to discover if β -cell-specific
 278 genes were overexpressed within any of the β -cell subpopulations. We identified 20 β -cell-specific genes
 279 overexpressed in Beta 1 cells, far more than would be expected by chance (**Figure 3H**).

280 *β -cell genes are differentially expressed by age, diet, and sex.* To robustly characterize β -cell gene expression in
 281 SM/J mice, we normalized bulk RNA-seq data from 20- and 30-week high- and low-fat males and females (4 per
 282 cohort, n=32) using only the 316 β -cell-specific genes. Principal components analysis on the normalized expression
 283 data revealed high-fat males separated from other cohorts (**Figure 4A**). This is consistent with our previous studies
 284 showing that high-fat fed SM/J males show a more extreme glycemic response than other cohorts^{20,21}. Pairwise
 285 comparison revealed 8 differentially expressed genes between 20- and 30wk high-fat males and 2 differentially
 286 expressed females between 20- and 30wk high-fat females (**Figure 4B**). 20wk male mice differentially express 104
 287 genes between diets, while 30wk male mice differentially expressed 17. Females differentially expressed a largely
 288 consistent set of genes between diets in 20- and 30wk cohorts. Differential expression analysis is reported in **Table**
 289 **11**. We focused subsequent analyses between 20wk and 30wk high-fat males (**Figure 4C**), and 20wk high and low-
 290 fat males (**Figure 4E**). These analyses allow comparison between hyper- and normoglycemic-obese mice, and
 291 between hyperglycemic-obese and normoglycemic-lean mice. We highlight two genes with unknown roles in β -
 292 cell function, *Pdyn* and *Fam151a*, as differentially expressed between 20- and 30wk high-fat males (**Figure 4D**),
 293 and 20wk high and low-fat males (**Figure 4F**), respectively.

294

295

Figure 3. Single cell RNA sequencing identifies β -cell-specific genes. (A) Total gene expression and β -cell-specific contribution to gene expression in 20wk low-fat female β -cells. Gold line indicates threshold for β -cell-specific expression. Red dots identify genes that pass the threshold cutoff in this cohort. Arrows identify highly expressed genes associated with β -cell identity. Genes must pass threshold in all cohorts to be considered β -cell-specific in the rest of the analysis. (B) UMAP plot of all islet cell types. UMAP plots for β -cell-specific expression of known β -cell markers (C) *Ucn3*, (D) *Gcgr*, (E) *Slc2a2*. UMAP plots for β -cell-specific expression of novel β -cell markers (F) *Pdyn*, and (G) *Fam151a*. (H) Permutation analysis of β -cell-specific genes in Beta 1 overexpressed genes. Distribution shows expected number of β -cell-specific genes in Beta 1 overexpressed gene set based on chance ($n = 1000$ permutations), red line indicates real number of β -cell-specific genes in Beta 1 overexpressed gene set. P-value indicates probability of number of β -cell-specific genes in Beta 1 overexpressed gene set due to chance.

306

Figure 4. β -cell-specific genes are differentially expressed by age, diet, and sex. **(A)** Principal component analysis of bulk RNA sequencing data normalized with only β -cell-specific genes. **(B)** Number of differentially expressed genes in females and males, across diets and age. **(C)** Differentially expressed genes between 20wk high-fat males and 30wk high-fat males. **(D)** *Pdyn* is differentially expressed in high-fat males. **(E)** Differentially expressed genes between 20wk high-fat males and 20wk low-fat males. **(F)** *Fam151a* is differentially expressed in 20wk males. Blue genes are significantly under-expressed in comparison, red genes are significantly over-expressed. Vertical golden lines indicate threshold for significance based on average log fold change, horizontal line indicates threshold for significance based on FDR corrected p-value. **p-value < 0.01, **** p-value < 0.0001 in FDR corrected pairwise comparison. HF20 – 20wk high-fat female, HF30 – 30wk high-fat female, HM20 – 20wk high-fat male, HM30 – 30wk high-fat male, LF20 – 20wk low-fat female, LM20 – 20wk low-fat male.

317 *β-cell gene expression networks correlate with metabolic traits.* We performed network analysis to characterize
 318 how groups of genes behaved across age and dietary cohorts. We performed weighted gene co-expression network
 319 analysis (WGCNA), which groups similarly co-expressed genes into discrete modules, then correlates these
 320 modules with phenotypes³². We collected metabolic phenotypes in the bulk RNA-seq mice including body weight,
 321 blood glucose level, serum insulin, and islet-specific phenotypes including glucose-stimulated insulin secretion,
 322 basal insulin secretion, and islet insulin content (**Supplemental Figure 3**). Performing co-expression analysis
 323 identified 9 discrete modules of genes, 6 of which correlated significantly with at least one phenotype (**Table 13**,
 324 **Supplemental Figure 4**). We highlight the blue and brown modules for their correlation with phenotypes relevant
 325 to hyperglycemic obese 20wk high-fat mice.

326 *Blue module associated with fatty acid metabolism correlates with blood glucose level and is altered by age in*
 327 *obese mice.* The blue module contains 42 genes, and over-representation analysis reveals enrichment for genes
 328 related to fatty acid metabolism (**Table 12**). Blue network eigengene expression correlates with blood glucose level
 329 across all mice (**Figure 5A**, **Supplemental Figure 5F**). This correlation is driven by the resolution of
 330 hyperglycemia in high-fat mice (**Figure 5A**, **Supplemental Figure 5B**). Strength of individual gene membership
 331 within the blue module correlates with strength of correlation to blood glucose levels (**Figure 5B**). The blue network
 332 is visualized as the sum of gene-pair correlations (**Supplemental Figure 5D**). This network was then subset into
 333 20wk high-fat males (**Figure 5D**) and 30wk high-fat males (**Figure 5E**). Four genes were identified as differentially
 334 connected between networks, although none were individually differentially expressed between 20- and 30wk high-
 335 fat males (**Figure 5C**, **Table 14**).

336

337 **Figure 5.** Blue module network structure is altered by age in high-fat male mice. **(A)** Blue module heatmap,
 338 eigengene expression, and blood glucose levels across all individuals. **(B)** Correlation between strength of module
 339 membership and blood glucose levels for blue module. **(C)** Total and differential connectivity between blue module
 340 genes in 20wk and 30wk high-fat male mice. Vertical golden lines indicate threshold for differential connectivity.
 341 **(D)** Blue module network structure in 20wk high-fat males and **(E)** 30wk high-fat males. Size and color of node
 342 indicates overall connectivity within the network, thickness of edges indicates strength of correlation between gene
 343 pairs.

344 *Brown module associated with β -cell function correlates with insulin levels and is influenced by diet.* The brown
 345 module contains 39 genes, and over-representation analysis reveals enrichment for genes related to abnormal β -cell
 346 physiology (**Table 12**). Brown network eigengene expression negatively correlates with serum insulin levels across
 347 all mice (**Figure 6A, Supplemental Figure 5E**). This correlation is driven in part by diet, where high-fat mice have

high levels of circulating insulin (**Figure 6A, Supplemental Figure 5A**). Strength of gene membership in the brown module negatively correlates with strength of correlation to serum insulin level, including *Fam151a*, which is highly connected in the brown network (**Figure 6B**). Further, expression of *Fam151a* negatively correlates with serum insulin levels across all cohorts, independent of network construction (**Figure 6C**). The brown network is visualized as the sum of gene-pair correlations (**Supplemental Figure 5C**). The network was subset into 20wk high-fat males (**Figure 6D**) and 20wk low-fat males (**Figure 5E**), revealing strong network structure in 20wk low-fat males. Differential connectivity was calculated between 20wk high and low-fat males to identify genetic network connectivity that was altered between cohorts (**Figure 6F, Table 14**). Differential connectivity was contrasted with differential expression between 20wk high and low-fat males, revealing *Fam151a* to be both differentially connected and differentially expressed (**Figure 6G**).

Single cell expression of Pdyn and Fam151a are influenced by age and diet. We sought to determine if the differential expression of *Pdyn* and *Fam151a* seen in the bulk analysis is recapitulated at the single cell level. We compared expression of *Pdyn* in all β -cells between 20wk high-fat males and 30wk high-fat males (**Figure 7A, Table 21**) and found significant differential expression, driven by differences in the proportion of β -cells expressing *Pdyn* (**Figure 7B**). Likewise, *Fam151a* is differentially expressed across all β -cells between 20wk high-fat males and 20wk low-fat males (**Figure 7C, Table 22**), driven by differences in the proportion of β -cells expressing *Fam151a* (**Figure 7D**). To determine if differential expression could be attributed to changes in β -cell subpopulation proportions, we calculated differential expression between the two most prominent subtypes, Beta 1 and Beta 2, revealing significant differential expression of *Pdyn* and *Fam151a* (**Figure 7E, Table 15**). While *Pdyn* is overexpressed in Beta 2 cells, which are prominent in 30wk high-fat males, *Fam151a* is overexpressed in Beta 1 cells, which are prominent in 20wk high-fat males. This suggests differential expression of *Pdyn* can be attributed to differences in subpopulation proportions, but not for differential expression of *Fam151a*. Assessing differential expression in Beta 1 cells between 20wk high-fat males and 20wk low-fat males revealed *Fam151a* was very significantly differently expressed (**Figure 7F, Table 23**), driven by an increase in the proportion of β -cells

expressing *Fam151a* (Figure 7G), which could contribute to the differential expression of *Fam151a* between 20wk high-fat males and 20wk low-fat males seen in the bulk data.

374

375

Figure 6. Brown module network structure is altered by diet in 20wk male mice. (A) Brown module heatmap, eigengene expression, and serum insulin levels across all individuals. (B) Correlation between strength of module membership and serum insulin levels for brown module. Gene of interest, *Fam151a*, highlighted as a highly connected member of the brown module and highly correlative with serum insulin levels. (C) Expression of *Fam151a* correlates with serum insulin levels across all individuals independent of module membership. Light red – 20wk high-fat mice, dark red – 30wk high-fat mice, light blue – 20wk low-fat mice, dark blue – 30wk low-fat fed mice. (D) Brown module network structure in 20wk high-fat males and (E) 20wk low-fat males. Gene of interest, *Fam151a*, is highlighted in yellow in each network. Size and color of node indicates overall connectivity within the

network, thickness of edges indicates strength of correlation between gene pairs. (F) Total and differential connectivity between brown module genes in high and low-fat male SM/J mice. Vertical golden lines indicate threshold for significant differential connectivity. Gene of interest, *Fam151a*, is highlighted as a differentially connected gene. (G) Differential connectivity and differential expression of brown module genes between 20wk high and 20wk low-fat males. FDR – false discovery rate of p-values associated with differential expression. Vertical golden lines indicate threshold for differential connectivity, Horizontal golden line indicates threshold for differential expression. Gene of interest, *Fam151a*, is highlighted as differentially connected and differentially expressed.

Figure 7. *Pdyn* and *Fam151a* are differentially expressed in β -cell subtypes. (A) Expression of β -cell-specific genes across all β -cells between 20wk and 30wk high-fat males. (B) Percent of β -cells in 20wk and 30wk high-fat males expressing β -cell-specific genes. (C) Expression of β -cell-specific genes across all β -cells between 20wk high and low-fat males. (D) Percent of β -cells in 20wk high and low-fat males expressing β -cell-specific genes. (E) Comparison of *Pdyn* and *Fam151a* expression across β -cell subtypes. (F) Expression of β -cell-specific genes in beta 1 cells between 20wk high-fat and 20wk low-fat fed mice. (G) Percent of beta 1 cells expressing beta specific genes between 20wk high-fat and 20wk low-fat fed mice. Blue genes are significantly under-expressed in comparison, red genes are significantly over-expressed. Vertical golden lines indicate threshold for significance based on average log fold change, horizontal line indicates threshold for significance based on Bonferroni corrected p-value.

404 Discussion

405 Pancreatic β -cell heterogeneity has been studied extensively using single cell technology because of the cell type's
 406 unique insulin-secreting capabilities and central role in diabetes etiology. Several groups have proposed the
 407 existence of functionally distinct β -cell subpopulations^{35–38}, but their existence and relevance to insulin homeostasis
 408 remain controversial. We identified 4 β -cell populations (**Figure 1-2, Supplemental Figure 1**), Beta 1-4, across
 409 hyperglycemic-obese, normoglycemic-obese, and normoglycemic-lean mice. While the functional roles assigned
 410 to each population are based on gene enrichment analysis and open to interpretation (**Supplemental Table 2**), the
 411 relative proportion of each population changes in conjunction with islet function across cohorts, suggesting a
 412 functional relevance to overall β -cell population structure.

413 Beta 1 cells are most prevalent in hyperglycemic obese mice and are associated with elevated basal insulin secretion
 414 and low insulin content. Beta 1 cells overexpress mature markers including *Ucn3*³⁹, *Pdx1*⁴⁰, and *Acly*⁴¹, and negative
 415 regulators of glucose-stimulated insulin secretion including *Abcc8*⁴², *G6pc2*⁴³, and *Ucp2*⁴⁴. Terms associated with
 416 Beta 1 cells include signal release and hormone transport. Beta 1 cells appear to be primed for insulin release, but
 417 not to perform glucose-stimulated insulin secretion. Farack *et al*³⁶ identified a subpopulation of “extreme” β -cells
 418 that specialize in basal insulin secretion, with high insulin expression and low mature insulin content. These cells
 419 have high expression of markers *Ucn3*, *Acly*, and *Pdx1*, and increase in proportion in obese mice. The Beta 1 cells
 420 we identified are similar to “extreme” β -cells based on their mature profile, potentially limited glucose response,
 421 and abundance in diabetic obese mice. Validation is required to determine if these β -cells function by over-
 422 expressing insulin secretion pathway components while suppressing glucose response mechanisms.

423 Beta 2 cells are prevalent in normoglycemic-obese and normoglycemic-lean mice, and are associated with high
 424 GSIS and high islet insulin content. Beta 2 cells are enriched for response to oxygen levels, pyruvate metabolism,
 425 and nucleotide phosphorylation, each associated with protection from hypoxia^{45,46}. These cells are equally prevalent
 426 in obese 30wk high-fat males and lean 20wk low-fat males, which differ greatly in islet mass, suggesting Beta 2
 427 cells are not a hypoxic population. Further, 20wk high-fat males have similar islet mass to 30wk high-fat males, but
 428 a much smaller Beta 2 population. Importantly, 30wk high-fat males have healthier glycemic parameters than 20wk

high-fat males^{20,21}. Several groups have identified a subset of heavily vascularized islets that have elevated oxygen consumption and superior GSIS at the cost of susceptibility to hypoxia. We suspect Beta 2 cells represent a highly functional β -cell population that confer protection against hypoxia^{47,48}. Given their abundance in functional islets from both obese and lean normoglycemic mice, we hypothesize Beta 2 cells represent a mature, and possibly stress-resistant β -cell population.

Beta 3 cells are prevalent in normoglycemic obese mice which have elevated β -cell mass, high GSIS, and high insulin content. Beta 3 cells overexpress plasma membrane components, including several claudin family members, and ribosome components. Claudins provide structural integrity to tight junctions, which maintain cell polarity⁴⁹. Polarity in β -cells serves as both a driver and a characteristic of mature, highly functional β -cells^{35,50,51}. *Cldn4*, *Cldn3*, and *Cldn7* are all upregulated in Beta 3 cells and associated with mature β -cell function⁵². Beta 3 cells upregulate 7 ribosome genes and elevated ribosome biogenesis is associated with increased β -cell apoptosis⁵³, β -cell replication³⁵, and mature β -cell function⁵⁴, making the functional relevance of overexpression unclear. One gene, *Rpl7*, is associated with mature β -cell function⁵⁵, while *Rpl23* protects against apoptosis⁵⁶. From this we conclude that Beta 3 cells represent a polarized and mature β -cell population.

Beta 4 cells are the lowest in abundance and elevated in lean males compared to obese males. Terms associated with Beta 4 cells include response to stress and response to topologically incorrect protein. In addition to overexpression of UPR components, Beta 4 cells have significant down regulation of *Ins1* and *Ucn3*. Several groups have identified β -cell subpopulations based on a stress response signature^{13,57,58}, which have coalesced into a theory about stress response cycling, where β -cells go through periods of UPR activation and low insulin production to clear misfolded proteins^{15,59}. Interestingly, 20wk low-fat mice have a similar proportion of stress response cycling cells compared to other groups (~7%), however this population is significantly smaller in high-fat males, suggesting that stress response cycling may be suppressed by obesity.

Efforts to describe β -cell transcriptional heterogeneity are marred by lack of consensus across scRNA-seq studies⁶⁰. This is attributed to low read depth, which results in differences in gene capture driven by technical artifacts and random chance rather than biological variation¹⁶. Further, it is unclear what level of gene expression fold change is

454 meaningful between individual cells, rendering single cell data poorly suited to assess differential expression across
455 different cohorts. To address these problems, we developed a technique to integrate single cell data with bulk RNA-
456 seq data (**Figure 3, Supplemental Figure 2**), which is not inhibited by these technical limitations, to assess β -cell-
457 specific differential expression across cohorts of high- and low-fat fed male and female SM/J mice. While the list
458 of genes identified as β -cell-specific is only 316 genes (**Supplemental Table 9**), we are highly confident of their
459 specificity to β -cells, allowing for robust assessment of differential expression and network analysis. Sub-setting
460 the bulk RNA-seq data by β -cell-specific genes provided a small but highly-focused search space to identify genes
461 that influence β -cell function in hyperglycemic-obesity, normoglycemic-obesity, and healthy mice.

462 We identified *Pdyn* as a novel candidate gene associated with improved glycemic control in males. *Pdyn* encodes
463 Proenkephalin-b, which is cleaved into Metenkephalin, an opioid with a short half-life and low bioavailability⁶¹.
464 While *Pdyn*'s role in β -cell function is unknown, many opioids modulate insulin secretion⁶². Further, *Pdyn* expression
465 is associated with regulation by *Pax6* and *Lkbl*, hallmarks of mature β -cells^{63,64}. *Pdyn* is over expressed in Beta 2
466 cells, which are increased in normoglycemic-obese mice, and may link improved β -cell function with hypoxia
467 response, as *Pdyn* provides protection from hypoxia in cardiomyocytes and neuronal tissue^{65,66}.

468 We used weighted gene co-expression network analysis (WGCNA) to identify networks of co-expressed β -cell
469 genes and to assess how networks correlate with metabolic and islet phenotypes (**Supplemental Figure 3**). Previous
470 efforts to construct networks in islet RNA-seq data failed to account for cell-type specific gene expression, making
471 it difficult to determine if these networks operate within an individual cell⁶⁷⁻⁶⁹. Our analysis revealed 6 β -cell-
472 specific modules that correlate with phenotypes, two of which were enriched for intriguing ontology terms
473 (**Supplemental Figure 4**). We chose to explore these networks in depth to assess the context in which they were
474 relevant to β -cell function in obesity.

475 The blue module is highly expressed in 20wk high-fat males, strongly correlating with blood glucose concentration
476 (**Figure 5, Supplemental Figure 5**). The blue module is enriched with fatty acid metabolism genes, which are
477 heavily implicated in obesity and diabetes (**Supplemental Table 12**). In β -cells, fatty acid metabolism provides a
478 glucose-independent stimulus for insulin release^{70,71}. While short term fatty acid exposure enhances GSIS,

479 prolonged exposure decreases GSIS and reduces insulin content^{70,72}. Fatty acid metabolism protects against
 480 glucotoxicity-induced damage, allowing cells to convert excess glucose into glycerol, which bypasses ROS
 481 generation⁷³. How this network connects to the β -cell dysfunction seen in obese, hyperglycemic 20wk high-fat mice
 482 is not immediately clear. Since fatty acid metabolism can stimulate insulin release, it is possible that prolonged
 483 activation of this network promotes elevated basal insulin secretion but inhibits GSIS. Expression of this network
 484 may be related to the abundance of Beta 1 cells in 20k high-fat males, which are geared toward basal insulin release.
 485 One gene in the blue module, *Entpd3*, confers susceptibility to diet-induced obesity and hyperglycemia⁷⁴. This fatty
 486 acid metabolism gene network may provide a link between the increased proportion of Beta 1 cells and the β -cell
 487 dysfunction observed in 20wk high-fat males.

488 The brown module is highly expressed in lean mice and negatively correlates with serum insulin levels (**Figure 6**,
 489 **Supplemental Figure 5**). Unsurprisingly, this module is enriched for genes associated with β -cell physiology
 490 (**Supplemental Table 12**). We focused on how the network behaved in 20wk high and low-fat males, as they
 491 provided strong contrast for serum insulin levels and β -cell function. Core members of this network include known
 492 markers of β -cell maturation including *Ucn3*, *Slc2a2*, and *Pcsk1*. This network is suppressed in 20wk high-fat males,
 493 who have dysfunctional β -cells. We homed in on *Fam151a*, a gene highly connected within the network, that
 494 independently correlates with serum insulin levels. *Fam151a* is both differentially expressed and differentially
 495 connected within the brown module between 20wk high-fat and 20wk low-fat mice, suggesting an important role
 496 within the network and a potential link between serum insulin levels and β -cell function. *Fam151a* is a membrane
 497 embedded protein of unknown function, but it has been associated with dysregulated insulin secretion in *Isl1*
 498 knockout mice, and falls within a QTL associated with diabetes⁷⁵⁻⁷⁷. Further work will explore how *Fam151a*
 499 influences β -cell function in the context of obesity.

500 We explored if subpopulation composition contributes to differential expression of *Pdyn* and *Fam151a* (**Figure 7**).
 501 Total *Pdyn* expression increased across all β -cells between 20 and 30wk high-fat fed males, in agreement with the
 502 bulk RNA-seq data. *Pdyn* is overexpressed in Beta 2 cells, which increase in proportion between 20 and 30 weeks
 503 in high-fat mice, suggesting differential expression of *Pdyn* is driven by a change in subpopulation proportions.

Investigation into how *Pdyn* influences β -cell function must consider that knockout/overexpression may affect subpopulations differently, and that subpopulation structure changes based on metabolic context. Likewise, total *Fam151a* increased across all β -cells between 20wk high and low-fat males, in agreement with the bulk RNA-seq data. However, *Fam151a* is strongly overexpressed in Beta 1 cells, which are most abundant in 20wk high-fat males. To explore this discrepancy, we performed differential expression analysis in Beta 1 cells between 20wk high and 20wk low-fat males and found *Fam151a* expression is significantly higher in 20wk low-fat male Beta 1 cells. Thus, while the proportion of Beta 1 cells is lower in 20wk low-fat mice, the expression of *Fam151a* is significantly increased in their Beta 1 cells. These findings underscore the complexity of gene expression and suggest it is important to consider both subpopulation composition and expression within subpopulations when exploring β -cell heterogeneity in diabetes and obesity.

In summary, we identified 4 β -cell subpopulations whose relative proportions change depending on metabolic state. Beta 1 cells are primed for basal insulin secretion and proportionally high in hyperglycemic obese mice. Beta 2 cells are primed for protection from hypoxia associated with enhanced function and are abundant in normoglycemic obese and normoglycemic lean mice at the expense of Beta 1 cells. In conjunction, hyperglycemic obese mice express a highly connected genetic network associated with fatty acid metabolism, which is lost as glycemic control improves. The interplay between changing β -cell subpopulations and decreased fatty acid metabolism likely contributes to the improved β -cell function and subsequent restoration of glycemic control seen in obese SM/J mice^{20,21}. This study provides a road map for exploring cellular heterogeneity by integrating sc- and bulk RNA-seq data, allowing for robust characterization of subpopulation structure, differential expression, and network analysis associated with obesity and glycemic stress.

Acknowledgements

Author Contributions

HAL and MAM designed the experiments. MAM and HS performed mouse phenotyping and ELISAs. MAM and JFMV performed computational analysis and figure generation. MAM wrote the manuscript. All authors edited and approved the final draft.

529 Funding

530 This work was supported by the Washington University Department of Genetics, the Diabetes Research Center at
 531 Washington University (grant P30DK020579), the NIH NIDDK (grant K01 DK095003 to HAL, grant T32
 532 DK108742 to MAM, grant F31 DK125023-01 to MAM), GTAC@MGI Symposium Pilot Project Funding to HAL,
 533 and the NIH NHGRI (grant T32-GM007067 to JFMV).The authors declare no conflicts of interest.

534

535 Supplemental Figures

536
 537 **Supplemental Figure 1.** Single cell RNA sequencing quality control. (A) Total counts, features, and mitochondrial
 538 RNA expression plotted for each cell. Cells within blue box and without mitochondrial gene expression were
 539 included in analysis. (B) Number of cells identified in individual mice prior to quality control, (C) and after quality
 540 control. (D) Expression of *Ins2* across all cells prior to ambient RNA removing using SoupX, (E) and after ambient
 541 RNA removal. (F) Resolution analysis to determine the number of β -cell subpopulations. Selected resolution
 542 highlighted with red box.

543

544 **Supplemental Figure 2.** Quantifying expected expression for β -cell expression of *Dcd*. **(A)** Distribution of
545 expression across all β -cells. **(B)** Cullen and Frey analysis identifies *Dcd* expression to be beta distributed. **(C)**
546 Kolmogorov-Smirnov test identifies alpha and beta parameters that minimize Ks between real (blue line) and
547 simulated (red line) distribution of cells expressing *Dcd*. **(D)** Estimating percent of cells not expressing *Dcd* due to
548 gene drop out by iterating alpha and beta parameters that minimize Ks between real and simulated data. **(E)** Density
549 plot visualizing estimated parameters for distribution of *Dcd* expression. Red line shows distribution of actual data,
550 black line shows distribution of simulated data based on optimal alpha and beta parameters. From these parameters,
551 expected expression is calculated.

Supplemental Figure 3. Metabolic and islet phenotypes from bulk RNA sequencing mice. **(A)** body weight, **(B)** blood glucose, **(C)** serum insulin levels collected after 4-hour fast. **(D)** Glucose-stimulated insulin secretion (GSIS), **(E)** Basal insulin secretion, **(F)** Islet insulin content collected after isolated islets were rested overnight. N = 4 mice per age X sex X diet cohort. N= 10 islets per individual for islet phenotypes. Middle bar represents mean, box represents 25th and 75th quartile, whiskers represent minimum and maximum values.

559

560 **Supplemental Figure 4.** Correlation between phenotype and module eigengene expression. Number of genes
561 within each module reported on y-axis, phenotypic trait reported on x-axis. For each module-trait relationship, the
562 Pearson correlation between eigengene expression and phenotype value is reported (top) along with an FDR-
563 corrected p-value for the correlation (bottom). Color of box indicates strength of correlation. NS = non-significant
564 association based on FDR-corrected p-value.

Supplemental Figure 5. Overview of brown and blue modules. (A) Principal component analysis of gene expression within the brown module across all cohorts, segregated by diet. (B) Principal component analysis of gene expression within the blue module, segregated by cohort. (C) Brown module network structure in across all cohorts. Gene of interest, *Fam151a*, is highlighted in yellow. Size and color of node indicates overall connectivity within the network, thickness of edges indicates strength of correlation between gene pairs. (D) Blue module network structure in across all cohorts. Size and color of node indicates overall connectivity within the network, thickness of edges indicates strength of correlation between gene pairs. (E) Correlation between brown module eigengene expression and serum insulin levels across all cohorts. (F) Correlation between blue module eigengene expression and blood glucose levels across all cohorts.

576 Table Legends

577 **Supplemental Table 1.** Differential gene expression between β -cell subtypes. Average log fold change
578 (Avg_logFC) is calculated by dividing expression within subtype by expression in all other β -cells. Pct.1 is percent
579 of cells in subtype expressing the gene, pct.2 is percent of all other β -cells expressing the gene. P_Val_adj is
580 Bonferroni corrected p-value. Genes with Avg_logFC > 0.1 and p_val_adj <0.01 are considered significantly
581 enriched in the subpopulation.

582 **Supplemental Table 2.** Type five results for over-representation analysis (ORA) significantly enriched genes
583 within each β -cell subpopulation, including gene ontology terms.

584 **Supplemental Table 3.** Gene expression contribution for all cell types, in 20 week high-fat females.

585 **Supplemental Table 4.** Gene expression contribution for all cell types, in 30 week high-fat females.

586 **Supplemental Table 5.** Gene expression contribution for all cell types, in 20 week high-fat males.

587 **Supplemental Table 6.** Gene expression contribution for all cell types, in 30 week high-fat males.

588 **Supplemental Table 7.** Gene expression contribution for all cell types, in 20 week low-fat females.

589 **Supplemental Table 8.** Gene expression contribution for all cell types, in 20 week low-fat males.

590 **Supplemental Table 9.** β -cell-specific genes identified in single cell RNA sequencing analysis.

591 **Supplemental Table 10.** Top five results for over-representation analysis (ORA) for β -cell-specific genes,
592 including gene ontology and mammalian phenotype ontology terms.

593 **Supplemental Table 11.** Bulk RNA sequencing differential expression analysis for β -cell-specific genes. For each
594 analysis, log fold change, counts per million, p-value, and fdr corrected p-value are reported. hf = 20wk high-fat
595 females/30wk high-fat females, hm = 20wk high-fat males/30wk high-fat males, lf = 20wk low-fat females/30wk
596 low-fat females, lm = 20wk low-fat males/30wk low-fat males, f20 = 20wk high-fat females/20wk low-fat females,
597 m20 = 20wk high-fat males/20wk low-fat males, f30 = 30wk high-fat females/30wk low-fat females, m30 = 30wk
598 high-fat males/30wk low-fat males, age = all 20wk mice/all 30wk mice, diet = all high-fat mice/ all low-fat mice,
599 sex = all females/all males.

600 **Supplemental Table 12.** Type five results for over-representation analysis (ORA) significantly enriched genes
601 within brown and blue module genes, including gene ontology and KEGG pathway terms.

602 **Supplemental Table 13.** Gene significance and module membership for weighted gene co-expression analysis. For
603 each gene, module membership is reported. Gene significance (GS) for each trait is reported as a Pearsons
604 correlation, along with p-value (p.GS) associated with correlation. Module membership (MM) for each module
605 reported as Pearsons coefficient along with p-value (p.MM) associated with correlation. BG = Blood glucose, GSIS
606 = Glucose-stimulated insulin secretion, Basal.INS = Basal insulin secretion, INS.content = islet insulin content.

607 **Supplemental Table 14.** Network connectivity for all β -cell genes. Strength of connectivity for each gene within
608 its assigned module reported as k-within value.

609 **Supplemental Table 15.** Differential gene expression between beta 1 and beta 2 subtypes. Average log fold change
610 (Avg_logFC) is calculated by dividing expression within Beta 1 by expression in Beta 2. Pct.1 is percent of Beta 1
611 cells expressing the gene, pct.2 is percent of Beta 2 cells expressing the gene. P_Val_adj is Bonferroni corrected p-
612 value. Genes with Avg_logFC > 0.1 and p_val_adj <0.01 are considered significantly differentially expressed.

613 **Supplemental Table 16.** Differential gene expression between beta 1 and beta 3 subtypes. Average log fold change
614 (Avg_logFC) is calculated by dividing expression within Beta 1 by expression in Beta 3. Pct.1 is percent of Beta 1

615 cells expressing the gene, pct.2 is percent of Beta 3 cells expressing the gene. P_Val_adj is Bonferroni corrected p-
616 value. Genes with Avg_logFC > 0.1 and p_val_adj <0.01 are considered significantly differentially expressed.

617 **Supplemental Table 17.** Differential gene expression between beta 1 and beta 4 subtypes. Average log fold change
618 (Avg_logFC) is calculated by dividing expression within Beta 1 by expression in Beta 4. Pct.1 is percent of Beta 1
619 cells expressing the gene, pct.2 is percent of Beta 4 cells expressing the gene. P_Val_adj is Bonferroni corrected p-
620 value. Genes with Avg_logFC > 0.1 and p_val_adj <0.01 are considered significantly differentially expressed.

621 **Supplemental Table 18.** Differential gene expression between beta 2 and beta 3 subtypes. Average log fold change
622 (Avg_logFC) is calculated by dividing expression within Beta 2 by expression in Beta 3. Pct.1 is percent of Beta 2
623 cells expressing the gene, pct.2 is percent of Beta 3 cells expressing the gene. P_Val_adj is Bonferroni corrected p-
624 value. Genes with Avg_logFC > 0.1 and p_val_adj <0.01 are considered significantly differentially expressed.

625 **Supplemental Table 19.** Differential gene expression between beta 2 and beta 4 subtypes. Average log fold change
626 (Avg_logFC) is calculated by dividing expression within Beta 2 by expression in Beta 4. Pct.1 is percent of Beta 2
627 cells expressing the gene, pct.2 is percent of Beta 4 cells expressing the gene. P_Val_adj is Bonferroni corrected p-
628 value. Genes with Avg_logFC > 0.1 and p_val_adj <0.01 are considered significantly differentially expressed.

629 **Supplemental Table 20.** Differential gene expression between beta 3 and beta 4 subtypes. Average log fold change
630 (Avg_logFC) is calculated by dividing expression within Beta 3 by expression in Beta 4. Pct.1 is percent of Beta 3
631 cells expressing the gene, pct.2 is percent of Beta 4 cells expressing the gene. P_Val_adj is Bonferroni corrected p-
632 value. Genes with Avg_logFC > 0.1 and p_val_adj <0.01 are considered significantly differentially expressed.

633 **Supplemental Table 21.** Differential gene expression between all β -cells in 20wk high-fat fed males and 30wk
634 high-fat fed males. Average log fold change (Avg_logFC) is calculated by dividing expression within 20wk high-
635 fat fed males by expression in 30wk high-fat fed males. Pct.1 is percent of 20wk high-fat fed male β -cells expressing
636 the gene, pct.2 is percent of 30wk high-fat fed male β -cells expressing the gene. P_Val_adj is Bonferroni corrected
637 p-value. Genes with Avg_logFC > 0.1 and p_val_adj <0.01 are considered significantly differentially expressed.

638 **Supplemental Table 22.** Differential gene expression between all β -cells in 20wk high-fat fed males and 20wk
639 low-fat fed males. Average log fold change (Avg_logFC) is calculated by dividing expression within 20wk high-
640 fat fed males by expression in 20wk low-fat fed males. Pct.1 is percent of 20wk high-fat fed male β -cells expressing
641 the gene, pct.2 is percent of 20wk low-fat fed male β -cells expressing the gene. P_Val_adj is Bonferroni corrected
642 p-value. Genes with Avg_logFC > 0.1 and p_val_adj <0.01 are considered significantly differentially expressed.

643 **Supplemental Table 23.** Differential gene expression between Beta 1 cells in 20wk high-fat fed males and 20wk
644 low-fat fed males. Average log fold change (Avg_logFC) is calculated by dividing expression within beta 1 cells
645 of 20wk high-fat fed males by expression in beta 1 cells 20wk low-fat fed males. Pct.1 is percent of Beta 1 cells in
646 20wk high-fat fed males expressing the gene, pct.2 is percent of Beta 1 cells in 20wk low-fat fed males expressing
647 the gene. P_Val_adj is Bonferroni corrected p-value. Genes with Avg_logFC > 0.1 and p_val_adj <0.01 are
648 considered significantly differentially expressed.

649

650 References

- 651 1. Butler, A. E. *et al.* β -cell deficit and increased β -cell apoptosis in humans with type 2 diabetes. *Diabetes*
652 **52**, 102–110 (2003).
- 653 2. Talchai, C., Xuan, S., Lin, H. V., Sussel, L. & Accili, D. Pancreatic β Cell Dedifferentiation as a
654 Mechanism of Diabetic β Cell Failure. *Cell* **150**, 1223–1234 (2012).
- 655 3. Weir, G. C. & Bonner-Weir, S. Five Stages of Evolving Beta-Cell Dysfunction During Progression to
656 Diabetes. *Diabetes* **53**, S16–S21 (2004).
- 657 4. Giovannucci, E. *et al.* Diabetes and Cancer: A consensus report. *Diabetes Care* **33**, 1674–1685 (2010).
- 658 5. Leon, B. M. Diabetes and cardiovascular disease: Epidemiology, biological mechanisms, treatment

- recommendations and future research. *World J. Diabetes* **6**, 1246 (2015).
6. Chen, R., Ovbiagele, B. & Feng, W. Diabetes and Stroke: Epidemiology, Pathophysiology, Pharmaceuticals and Outcomes. *Am. J. Med. Sci.* **351**, 380–386 (2016).
7. Pepper, A. R., Gala-Lopez, B., Ziff, O. & Shapiro, A. J. Current status of clinical islet transplantation. *World J. Transplant.* **3**, 48 (2013).
8. Hering, B. J. *et al.* Phase 3 trial of transplantation of human islets in type 1 diabetes complicated by severe hypoglycemia. *Diabetes Care* **39**, 1230–1240 (2016).
9. Velazco-Cruz, L. *et al.* Acquisition of Dynamic Function in Human Stem Cell-Derived β Cells. *Stem Cell Reports* **12**, 351–365 (2019).
10. Rezaian, A. *et al.* Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. *Nat. Biotechnol.* **32**, 1121–1133 (2014).
11. Miranda, M. A., Macias-Velasco, J. F. & Lawson, H. A. Pancreatic β -cell heterogeneity in health and diabetes: classes, sources, and subtypes. *Am. J. Physiol. Metab.* **320**, E716–E731 (2021).
12. Baron, M. *et al.* A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. *Cell Syst.* **3**, 346–360.e4 (2016).
13. Muraro, M. J. *et al.* A Single-Cell Transcriptome Atlas of the Human Pancreas. *Cell Syst.* **3**, 385–394.e3 (2016).
14. Segerstolpe, Å. *et al.* Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes. *Cell Metab.* **24**, 593–607 (2016).
15. Xin, Y. *et al.* Pseudotime Ordering of Single Human β -Cells Reveals States of Insulin Production and Unfolded Protein Response. *Diabetes* **67**, 1783–1794 (2018).
16. Mawla, A. M. & Huisin, M. O. Navigating the Depths and Avoiding the Shallows of Pancreatic Islet Cell Transcriptomes. *Diabetes* **68**, 1380–1393 (2019).
17. Stegle, O., Teichmann, S. A. & Marioni, J. C. Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* **16**, 133–145 (2015).
18. Jew, B. *et al.* Accurate estimation of cell composition in bulk expression through robust integration of single-cell information. *Nat. Commun.* **11**, 1–11 (2020).
19. Wang, X., Park, J., Susztak, K., Zhang, N. R. & Li, M. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat. Commun.* **10**, 1–9 (2019).
20. Miranda, M. A. *et al.* Spontaneous restoration of functional β -cell mass in obese SM/J mice. *Physiol. Rep.* **8**, (2020).
21. Carson, C. *et al.* Brown Adipose Expansion and Remission of Glycemic Dysfunction in Obese SM/J Mice. *Cell Rep.* **33**, (2020).
22. Nikolskiy, I. *et al.* Using whole-genome sequences of the LG/J and SM/J inbred mouse strains to prioritize quantitative trait genes and nucleotides. *BMC Genomics* **16**, 415 (2015).
23. Young, M. D. & Behjati, S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. **9**, 1–10.
24. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21 (2019).
25. Zappia, L. & Oshlack, A. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *Gigascience* **7**, (2018).
26. Finak, G. *et al.* MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. (2015) doi:10.1186/s13059-015-0844-5.
27. Delignette-Muller, M. L. & Dutang, C. fitdistrplus: An R package for fitting distributions. *J. Stat. Softw.* **64**, 1–34 (2015).
28. Marsaglia, G., Tsang, W. W. & Wang, J. Evaluating Kolmogorov’s distribution. *J. Stat. Softw.* **8**, 1–4 (2003).
29. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
30. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (2010).
31. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–40 (2010).

- 710 32. Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC*
711 *Bioinformatics* **9**, 559 (2008).
- 712 33. Fuller, T. F. *et al.* Weighted gene coexpression network analysis strategies applied to mouse weight.
713 *Mamm. Genome* **18**, 463–472 (2007).
- 714 34. Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z. & Zhang, B. WebGestalt 2019: gene set analysis toolkit with
715 revamped UIs and APIs. *Nucleic Acids Res.* **47**, W199–W205 (2019).
- 716 35. Bader, E. *et al.* Identification of proliferative and mature β -cells in the islets of Langerhans. *Nature* **535**,
717 430–434 (2016).
- 718 36. Farack, L. *et al.* Transcriptional Heterogeneity of Beta Cells in the Intact Pancreas. *Dev. Cell* **48**, 115–
719 125.e4 (2019).
- 720 37. Johnston, N. R. *et al.* Beta Cell Hubs Dictate Pancreatic Islet Responses to Glucose. *Cell Metab.* **24**, 389–
721 401 (2016).
- 722 38. Smukler, S. R. *et al.* The Adult Mouse and Human Pancreas Contain Rare Multipotent Stem Cells that
723 Express Insulin. *Cell Stem Cell* **8**, 281–293 (2011).
- 724 39. Blum, B. *et al.* Functional beta-cell maturation is marked by an increased glucose threshold and by
725 expression of urocortin 3. *Nat. Biotechnol.* **30**, 261–264 (2012).
- 726 40. Bastidas-Ponce, A. *et al.* Foxa2 and Pdx1 cooperatively regulate postnatal maturation of pancreatic β -cells.
727 *Mol. Metab.* **6**, 524–534 (2017).
- 728 41. Chu, K. Y. *et al.* ATP-Citrate Lyase Reduction Mediates Palmitate-induced Apoptosis in Pancreatic Beta
729 Cells. *J. Biol. Chem.* **285**, 32606–32615 (2010).
- 730 42. Bennett, K., James, C. & Hussain, K. Pancreatic β -cell K ATP channels: Hypoglycaemia and
731 hyperglycaemia. *Rev. Endocr. Metab. Disord.* **2010 113** **11**, 157–163 (2010).
- 732 43. Pound, L. D. *et al.* G6PC2: A Negative Regulator of Basal Glucose-Stimulated Insulin Secretion. *Diabetes*
733 **62**, 1547–1556 (2013).
- 734 44. Zhang, C.-Y. *et al.* Uncoupling Protein-2 Negatively Regulates Insulin Secretion and Is a Major Link
735 between Obesity, β Cell Dysfunction, and Type 2 Diabetes. *Cell* **105**, 745–755 (2001).
- 736 45. Kim, J., Tchernyshyov, I., Semenza, G. L. & Dang, C. V. HIF-1-mediated expression of pyruvate
737 dehydrogenase kinase: A metabolic switch required for cellular adaptation to hypoxia. *Cell Metab.* **3**, 177–
738 185 (2006).
- 739 46. Losenkova, K. *et al.* Endothelial cells cope with hypoxia-induced depletion of ATP via activation of
740 cellular purine turnover and phosphotransfer networks. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1864**,
741 1804–1815 (2018).
- 742 47. Lau, J., Svensson, J., Grapensparr, L., Johansson, Å. & Carlsson, P.-O. Superior beta cell proliferation,
743 function and gene expression in a subpopulation of rat islets identified by high blood perfusion.
744 *Diabetologia* **55**, 1390–1399 (2012).
- 745 48. Ullsten, S., Lau, J. & Carlsson, P.-O. Vascular heterogeneity between native rat pancreatic islets is
746 responsible for differences in survival and revascularisation post transplantation. *Diabetologia* **58**, 132–
747 139 (2015).
- 748 49. Meda, P. Protein-Mediated Interactions of Pancreatic Islet Cells. *Scientifica (Cairo)*. **2013**, 1–22 (2013).
- 749 50. Collares-Buzato, C. B., Carvalho, C. P. F., Furtado, A. G. & Boschero, A. C. Upregulation of the
750 expression of tight and adherens junction-associated proteins during maturation of neonatal pancreatic
751 islets in vitro. *J. Mol. Histol.* **2004 358** **35**, 811–822 (2004).
- 752 51. Gan, W. J. *et al.* Cell polarity defines three distinct domains in pancreatic β -cells. *J. Cell Sci.* **130**, 143–151
753 (2017).
- 754 52. Li, H., John, A. N., Nagatake, T., Hamazaki, Y. & Jiang, F. Claudin 4 in pancreatic β cells is involved in
755 regulating the functional state of adult islets. *FEBS Open Bio* **10**, 28 (2020).
- 756 53. S, A., T, M., Y, K. & M, K. Increased ribosomal biogenesis induces pancreatic beta cell failure in mice
757 model of type 2 diabetes. *Biochem. Biophys. Res. Commun.* **381**, 367–371 (2009).
- 758 54. Kobiita, A. *et al.* The Diabetes Gene JAZF1 Is Essential for the Homeostatic Control of Ribosome
759 Biogenesis and Function in Metabolic Stress. *Cell Rep.* **32**, 107846 (2020).
- 760 55. Ni, Q. *et al.* Raptor regulates functional maturation of murine beta cells. *Nat. Commun.* **2017 81** **8**, 1–13

- (2017).
56. Qi, Y. *et al.* Ribosomal protein L23 negatively regulates cellular apoptosis via the RPL23/Miz-1/c-Myc circuit in higher-risk myelodysplastic syndrome. *Sci. Reports* 2017 71 7, 1–12 (2017).
 57. Fang, Z. *et al.* Single-Cell Heterogeneity Analysis and CRISPR Screen Identify Key β -Cell-Specific Disease Genes. *Cell Rep.* **26**, 3132–3144.e7 (2019).
 58. Baron, M. *et al.* A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. *Cell Syst.* **3**, 346–360.e4 (2016).
 59. Szabat, M. *et al.* Kinetics and genomic profiling of adult human and mouse β -cell maturation. *Islets* **3**, 175–187 (2011).
 60. Wang, Y. J. & Kaestner, K. H. Single-Cell RNA-Seq of the Pancreatic Islets—a Promise Not yet Fulfilled? *Cell Metab.* **29**, 539–544 (2019).
 61. Chavkin, C., Shoemaker, W. J., McGinty, J. F., Bayon, A. & Bloom, F. E. Characterization of the Prodynorphin and Proenkephalin Neuropeptide Systems in Rat Hippocampus'. *J. Neurosci* **5**, 606–616 (1985).
 62. Singh, A., Gibert, Y. & Dwyer, K. M. The adenosine, adrenergic and opioid pathways in the regulation of insulin secretion, beta cell proliferation and regeneration. *Pancreatology* **18**, 615–623 (2018).
 63. Swisa, A. *et al.* PAX6 maintains β cell identity by repressing genes of alternative islet cell types. *J. Clin. Invest.* **127**, 230 (2017).
 64. Kone, M. *et al.* LKB1 and AMPK differentially regulate pancreatic β -cell identity. *FASEB J.* **28**, 4972 (2014).
 65. S, P., OW, van den B., EG, L. & RP, X. Cross-talk of opioid peptide receptor and beta-adrenergic receptor signalling in the heart. *Cardiovasc. Res.* **63**, 414–422 (2004).
 66. da Silveira Scarpellini, C., Gargaglioni, L. H., Branco, L. G. S. & Bicego, K. C. Role of preoptic opioid receptors in the body temperature reduction during hypoxia. *Brain Res.* **1286**, 66–74 (2009).
 67. Keller, M. P. *et al.* Genetic Drivers of Pancreatic Islet Function. *Genetics* **209**, 335 (2018).
 68. Koltes, J. E. *et al.* A gene expression network analysis of the pancreatic islets from lean and obese mice identifies complement 1q like-3 secreted protein as a regulator of β -cell function. *Sci. Reports* 2019 91 9, 1–19 (2019).
 69. Perez-Alcantara, M. *et al.* Patterns of differential gene expression in a cellular model of human islet development, and relationship to type 2 diabetes predisposition. *Diabetol.* 2018 617 **61**, 1614–1622 (2018).
 70. Kristinsson, H., Smith, D. M., Bergsten, P. & Sargsyan, E. FFAR1 Is Involved in Both the Acute and Chronic Effects of Palmitate on Insulin Secretion. *Endocrinology* **154**, 4078–4088 (2013).
 71. Acosta-Montano, P. & García-González, V. Effects of Dietary Fatty Acids in Pancreatic Beta Cell Metabolism, Implications in Homeostasis. *Nutrients* **10**, (2018).
 72. Oprescu, A. I. *et al.* Free Fatty Acid-Induced Reduction in Glucose-Stimulated Insulin Secretion. *Diabetes* **56**, 2927–2937 (2007).
 73. Rotondo, F. *et al.* Glycerol is synthesized and secreted by adipocytes to dispose of excess glucose, via glycerogenesis and increased acyl-glycerol turnover. *Sci. Rep.* **7**, 8983 (2017).
 74. B, S. *et al.* Global deletion of NTPDase3 protects against diet-induced obesity by increasing basal energy metabolism. *Metabolism.* **118**, (2021).
 75. Woods, L. C. S. *et al.* Identification of Genetic Loci Involved in Diabetes using a Rat Model of Depression. *Mamm. Genome* **20**, 486 (2009).
 76. Ediger, B. N. *et al.* Islet-1 is essential for pancreatic β -cell function. (2014).
 77. H, K. *et al.* Examination of OLETF-derived non-insulin-dependent diabetes mellitus QTL by construction of a series of congenic rats. *Mamm. Genome* **13**, 558–562 (2002).