Supplementary Information.

Figure S1.

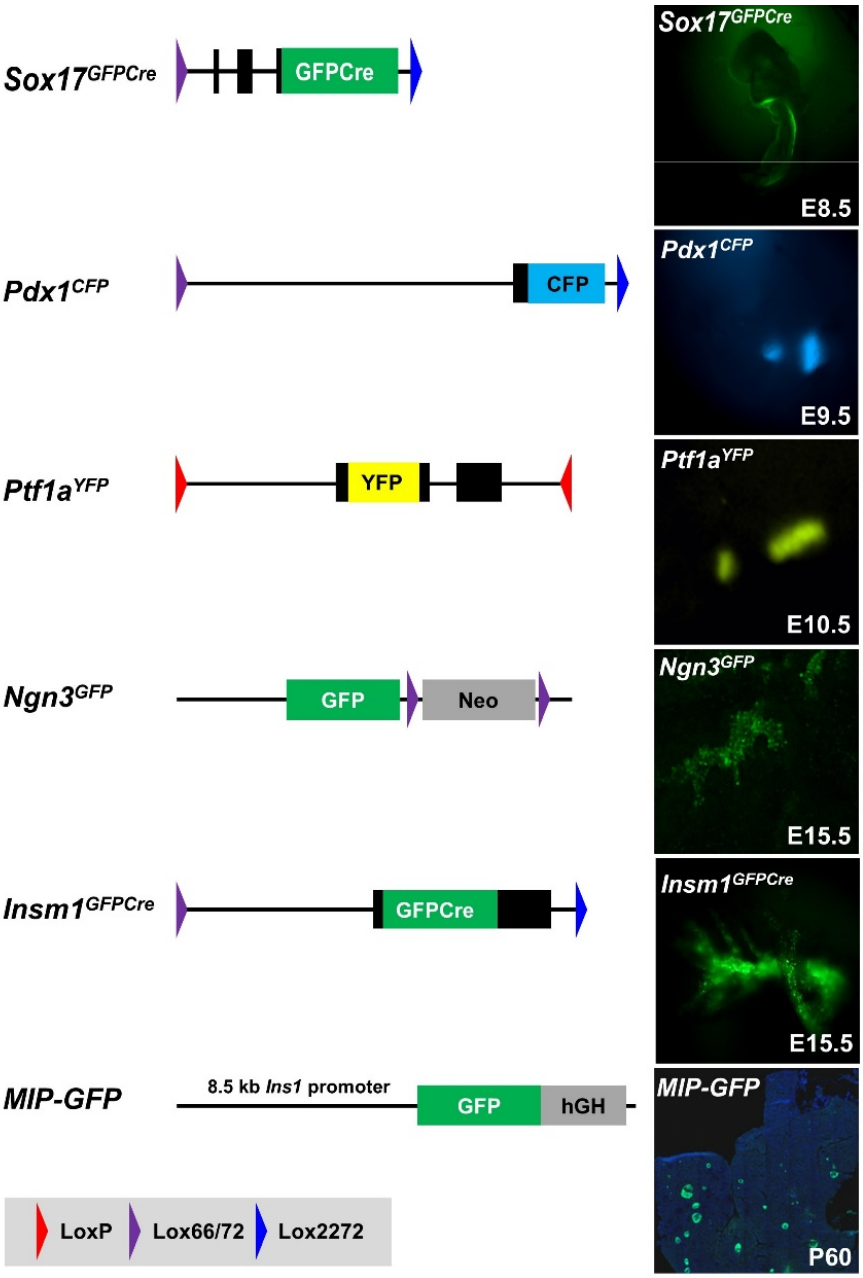


Figure S1. Mouse alleles used for FACS-isolation of specific cell populations during pancreas development. Schematic representations of fluorescent allele designs followed by direct fluorescence images demonstrating specific areas of fluorescent protein expression during pancreas development. *Sox17*^{GFPCre} expression marks endoderm at E8.0-8.5; *Pdx1*^{CFP} marks foregut endoderm at E9.5 (with higher levels of expression in ventral and dorsal pancreatic buds); *Ptf1a*^{YFP} marks pancreatic multipotent progenitor cells in branching pancreatic epithelium at E10.5; *Neurog3*^{GFP} marks endocrine progenitor cells at E15.5; *Insm1*^{GFPCre} marks progenitor and committed endocrine cells at E15.5; and the *MIP-GFP* transgene marks mature β-cells at P60.

Figure S2.

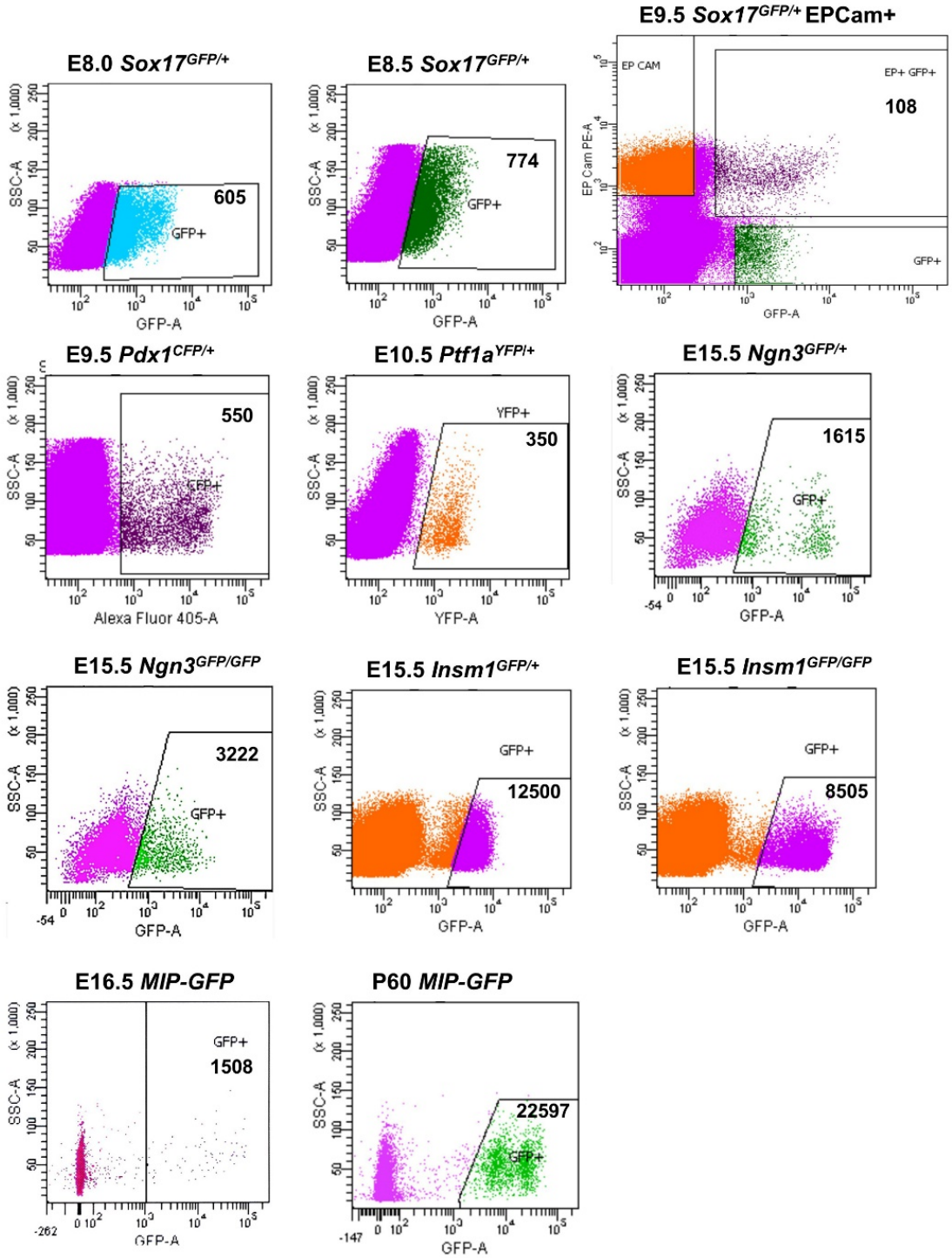
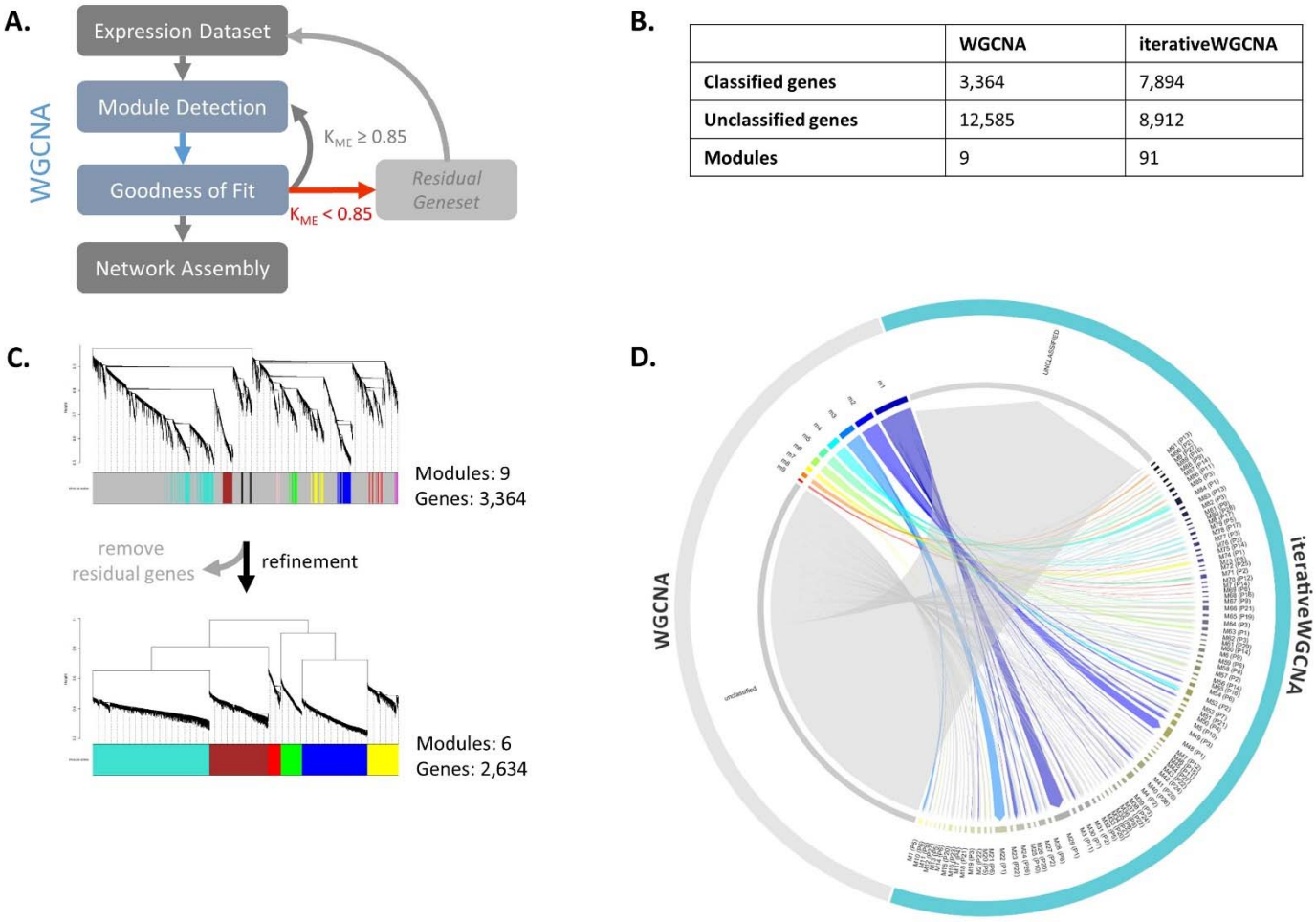


Figure S2. Representative FACS profiles of pancreatic cell populations isolated at different stages of development. Endodermal and pancreatic cell populations were isolated using the indicated mouse alleles and transgene at different developmental time points. Dispersed cells were sorted based on fluorescent protein expression. GFP and YFP were excited with a 488 nm laser and emission detected with a 502 nm long pass and 530/30 bandpass filter. CFP was excited using 405 nm laser and emission detected with a 470 nm long pass and 510/80 bandpass filter. The numbers inside the gates indicate the average number of fluorescent protein–positive cells obtained per embryo (n ≥ 3).

Figure S3.



	WGCNA	iterativeWGCNA
Classified genes	3,364	7,894
Unclassified genes	12,585	8,912
Modules	9	91

Figure S3. The use of iterativeWGCNA improves gene classification into β-cell co-expression network. **A.** A diagram of iterativeWGCNA process. Starting with a complete co-expression dataset, the first pass (and iteration) of iterativeWGCNA uses WGCNA's blockwise modules function to construct a tentative network and detect gene modules. Then, WGCNA is used to evaluate the goodness of fit for each gene against its assigned module, based on eigengene connectivity (kME). Poorly-fitting genes (kME <0.85) are set aside into a residual geneset and the remaining well-fitting genes are re-run through the network detection and refinement, until there are no poorly-fitting genes removed from assigned modules. Upon completion of such a pass of iterativeWGCNA, the residual geneset (poorly fitting genes pruned through multiple iterations within a single pass) is then used as a new expression dataset, and the process is repeated until no additional modules are detected. The network is then assembled after a final evaluation of goodness of fit of all genes to all predicted modules. Genes are assigned to the module to which they showed the best fit. Genes exhibiting a kME <0.85 to all module eigengenes are left unclassified. **B.** A table comparing network statistics after blockwise WGCNA and iterativeWGCNA analyses of the 15,949 genes expressed in developmental β-cell transcriptional datasets. **C.** Hierarchical clustering of genes generated from blockwise WGCNA versus one pass (6 iterations) of

iterativeWGCNA. The colored strips under each dendrogram depict the module assignment for the gene associated with the corresponding dendrogram branch. Clustering resulting from WGCNA analysis shows long intra-cluster branch lengths and the poor association between module memberships and gene clusters. The improved clustering resulting from one complete pass of iterativeWGCNA allows for better classification of the genes, by removing those that do not fit well with the primary signal in the expression dataset. Gene clusters are distinct (separated by long branches) and group genes assigned to single modules. **D.** A chord diagram showing gene classification changes between blockwise WGCNA (lowercase modules labels) and iterativeWGCNA (UPPERCASE module labels) applied to β-cell developmental transcriptional datasets. A large proportion of genes that were unclassified after blockwise WGCNA are assigned to modules by iterativeWGCNA. Furthermore, many genes were re-assigned from their initial blockwise WGCNA modules to better fitting modules.

Figure S4.

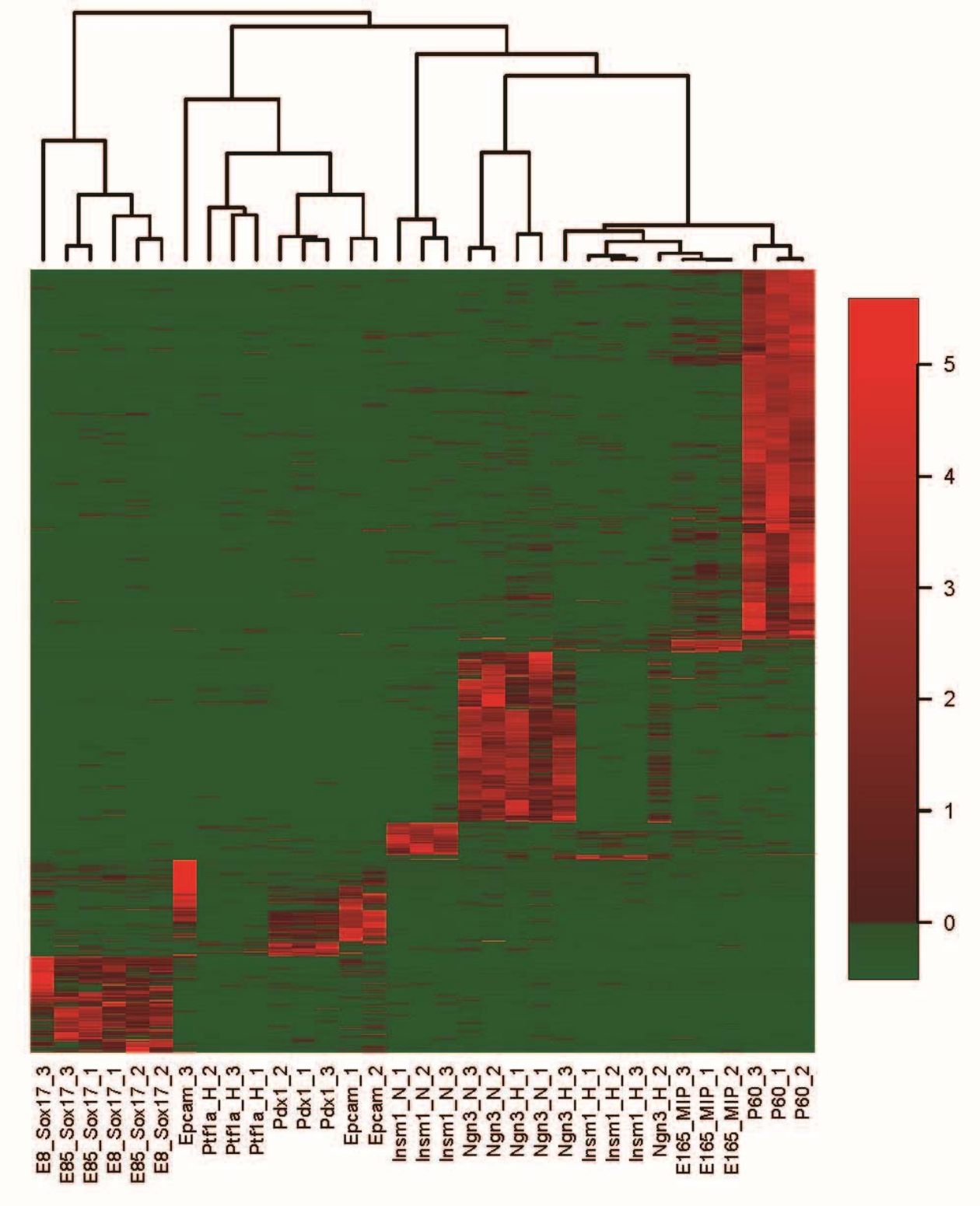


Figure S3. Heatmap illustrating filtering for tissue specificity. Expression heatmap for the (1656) exons with H<1.5. Rows are exons and columns are samples. Both exons and samples were clustered with hierarchical clustering (average linkage method) with similarity measured by Pearson correlation. The heatmap was generated with the R package Heatplus.

Figure S5.

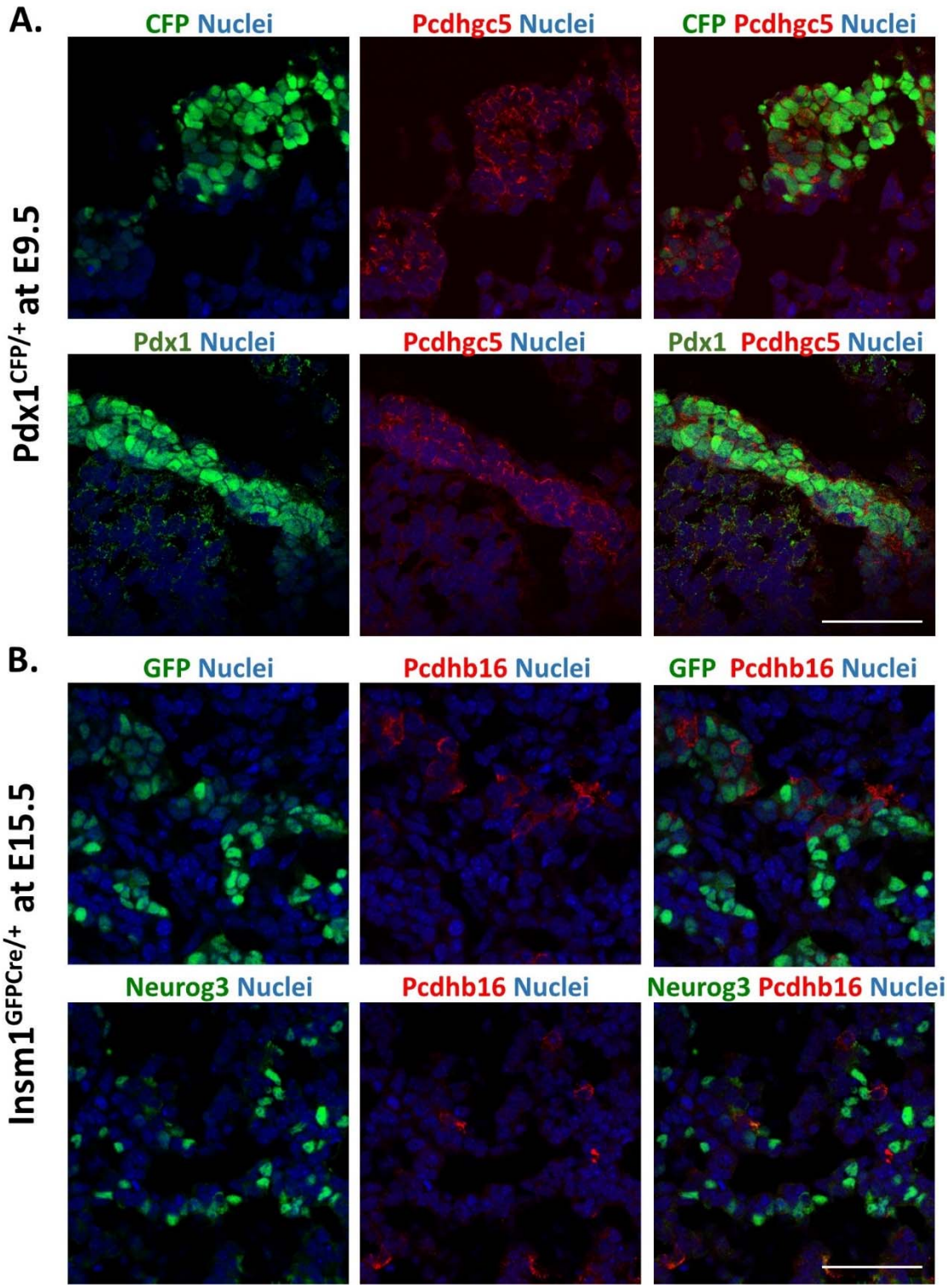


Figure S5. Co-expression of clustered protocadherins with pancreatic lineages. A. Immunofluorescent staining of pancreatic endoderm from *Pdx1*$^{CFP/+}$ mice at E9.5. Cluster γ protocadherin PCDHGC5 (red) is enriched in CFP-positive as well as PDX1-positive cells (green). **B.** Immunofluorescent staining of pancreatic tissue from *Insm1*$^{GFP/+}$ mice at E15.5. Cluster β protocadherin PCDHB16 (red) partially co-localizes with Insm1.GFP-positive and NEUROG3-positive endocrine progenitor cells (green). Nuclei are stained with To-Pro-3 dye (blue). Scale bar 50 μM.

Figure S6.

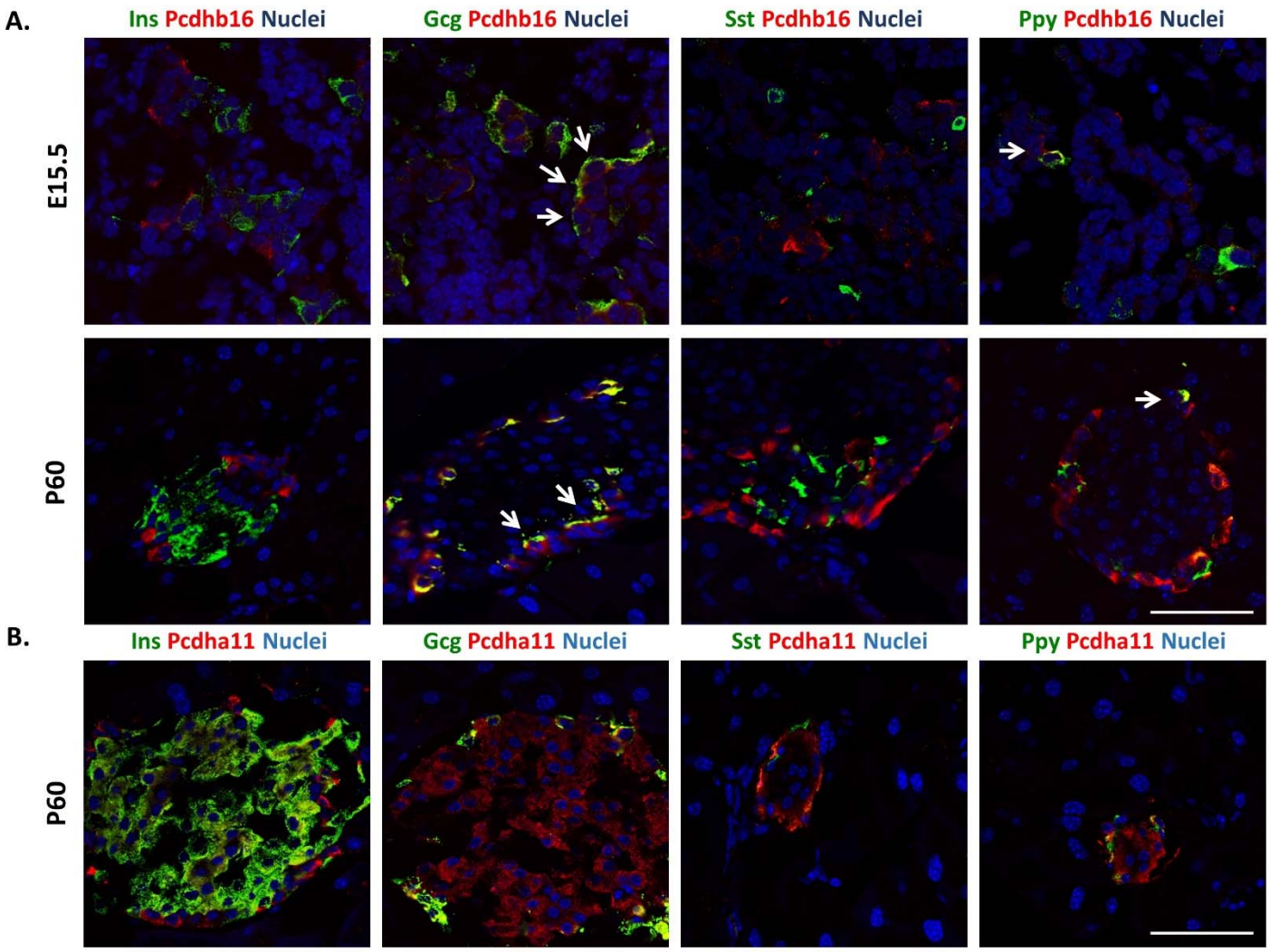


Figure S6. Co-expression of clustered protocadherins with pancreatic hormones. A. Immunofluorescent staining of pancreatic tissue from *Insm1*$^{GFP/+}$ mice at E15.5 and wild type mice at P60. Cluster β protocadherin PCDHB16 (red) co-localizes with Gcg- and Ppy-positive cells but not with Ins- or Sst-positive cells (green) at E15.5 and P60. Arrows show co-expressing cells on merged images. **B.** Cluster α protocadherin PCDHA11 (red) co-localizes with Ins-, Gcg- and some Ppy-positive cells (green) at P60. Nuclei are stained with To-Pro-3 dye (blue). Scale bar 50 μM.

Figure S7.

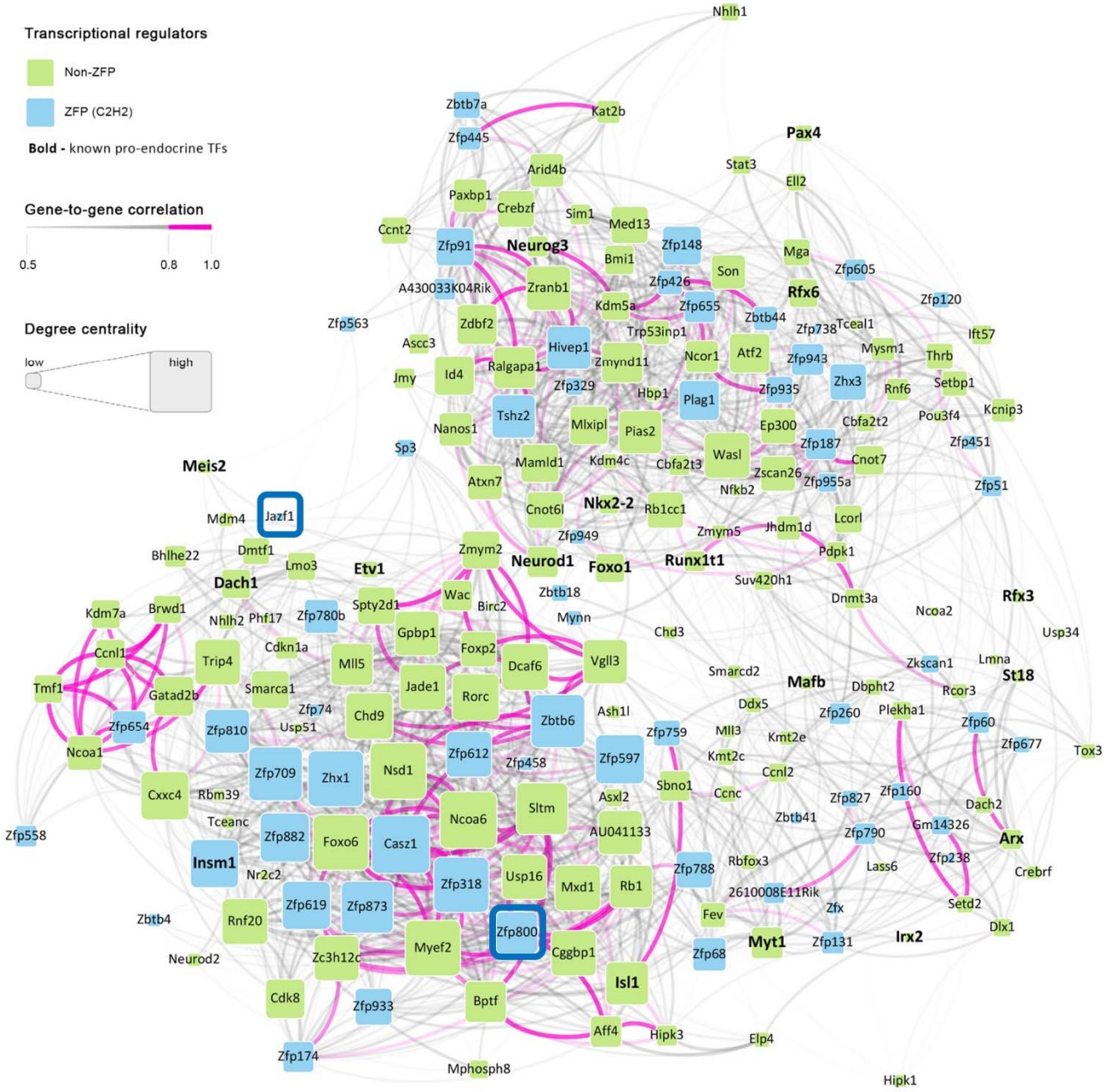


Figure S7. Sub-network of transcription regulators from meta-module J. A sub-network of transcriptional regulators (GOTERM_BP_DIRECT, GO:0006355) from meta-module J shows major interaction between genes predicted to be important for endocrine differentiation. The node size correspond to the scale denoting relative importance of genes for the whole meta-module J network (degree centrality measure). The node colors indicate whether transcriptional regulators belong to C2H2 Zn-finger proteins (ZFPs, light blue) or not (non-ZFPs, light green). Bold text for gene symbols indicates transcription factors (TFs) that has been previously described to be important for endocrine differentiation. Thick blue outline indicates the nodes for *Zfp800* and *Jazf1*, the ZFP genes that were selected for CRISPR/Cas9 knockout analysis.

Figure S8.

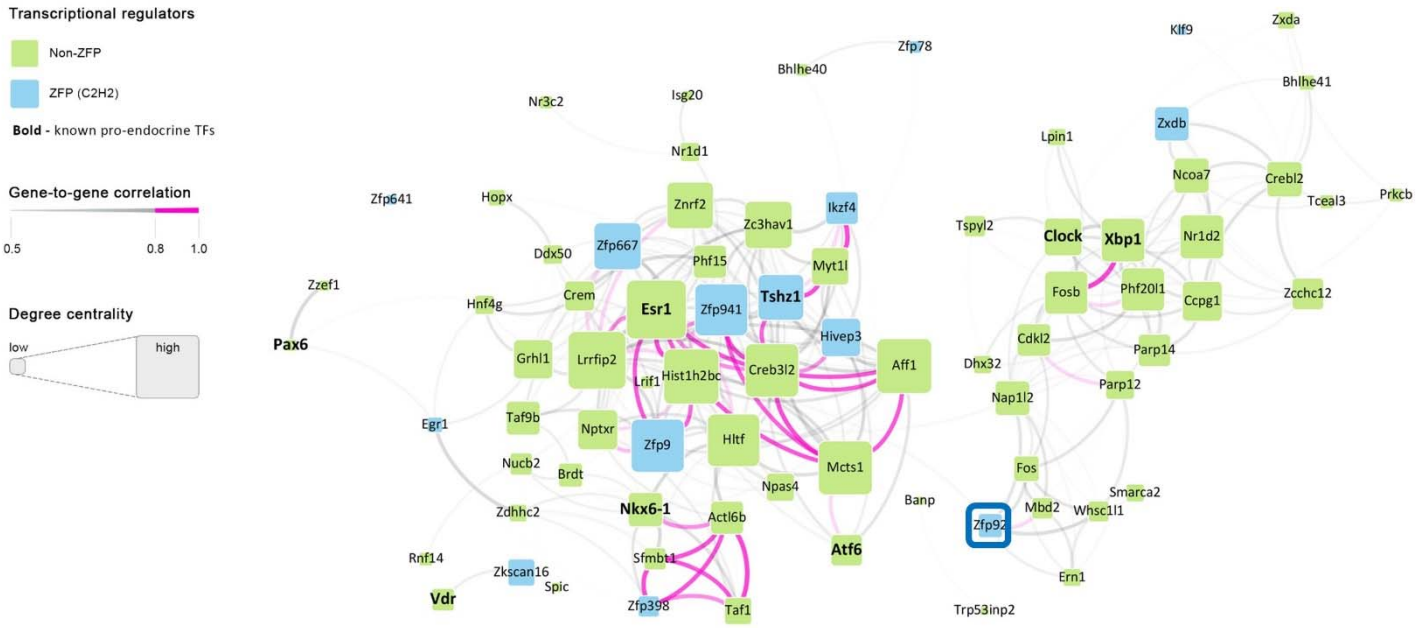


Figure S8. Sub-network of transcription regulators from meta-module D. A sub-network of transcriptional regulators (GOTERM_BP_DIRECT, GO:0006355) from meta-module D shows major interaction between genes predicted to be important for β-cell specification and function. The node size corresponds to the scale denoting relative importance of genes for the whole meta-module D network (degree centrality measure). The node colors indicate whether transcriptional regulators belong to C_2H_2 Zn-finger proteins (ZFPs, light blue) or not (non-ZFPs, light green). Bold text for gene symbols indicates transcription factors (TFs) known to be important for β-cell specification or function. Thick blue outline indicates the node for *Zfp92*, the ZFP gene that was selected for CRISPR/Cas9 knockout analysis.

Figure S9.

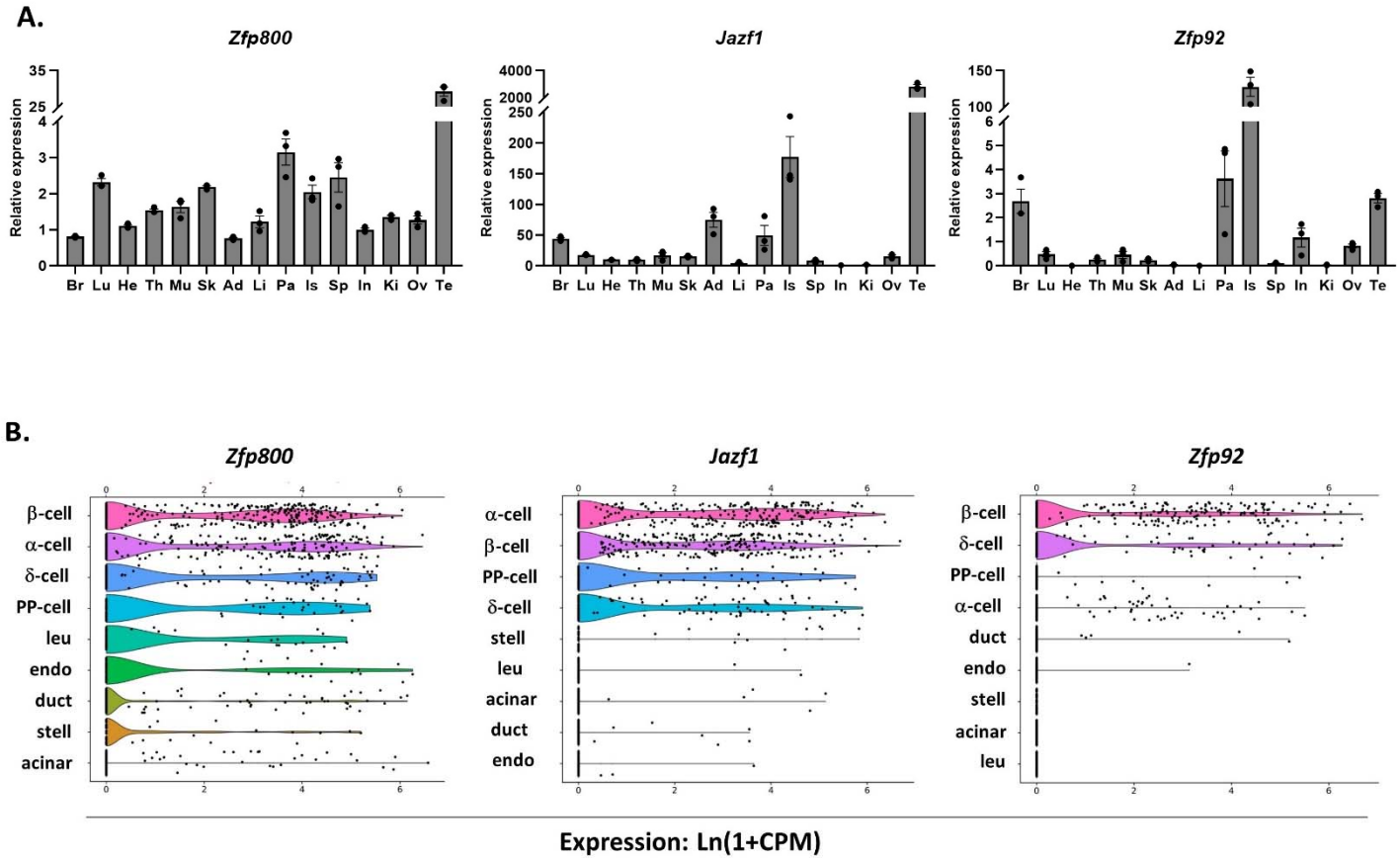


Figure S9. Tissue and islet cell type expression of *Zfp800, Jazf1* and *Zfp92*. **A.** Tissue expression patterns of *Zfp800, Jazf1* and *Zfp92* determined by the RT-qPCR analysis of tissues from adult mice (10 weeks old). All tissues, except ovaries, are from the male mice. N=3. Relative expression is calculated in comparison to the same gene expression in the intestine. Br, brain; Lu, lung; He, heart; Th, thymus; Mu, skeletal muscle; Sk, skin; Ad, adipose; Li, liver; Pa, whole pancreas; Is, pancreatic islets; Sp, spleen; In, small intestine; Ki, kidney; Ov, ovary; Te, testis. **B.** Expression levels of *Zfp800, Jazf1* and *Zfp92* in different islet cell populations as determined by scRNA-seq analysis data available through the Tabula Muris resource (PMID: 30283141). Violin plots display population distribution of expression levels for each of the genes with the most highly expressing population being on top of the graph. α-cell, β-cell, δ-cell, PP-cell are glucagon, insulin, somatostatin and pancreatic polypeptide producing pancreatic endocrine cells. Leu, pancreatic resident leucocytes; endo, endothelial cells; duct, pancreatic ductal cells; stell, pancreatic stellate (mesenchymal) cells; acinar, pancreatic exocrine acinar cells. *Zfp800, Jazf1* and *Zfp92* are all expressed in pancreatic endocrine cells.

Figure S10.

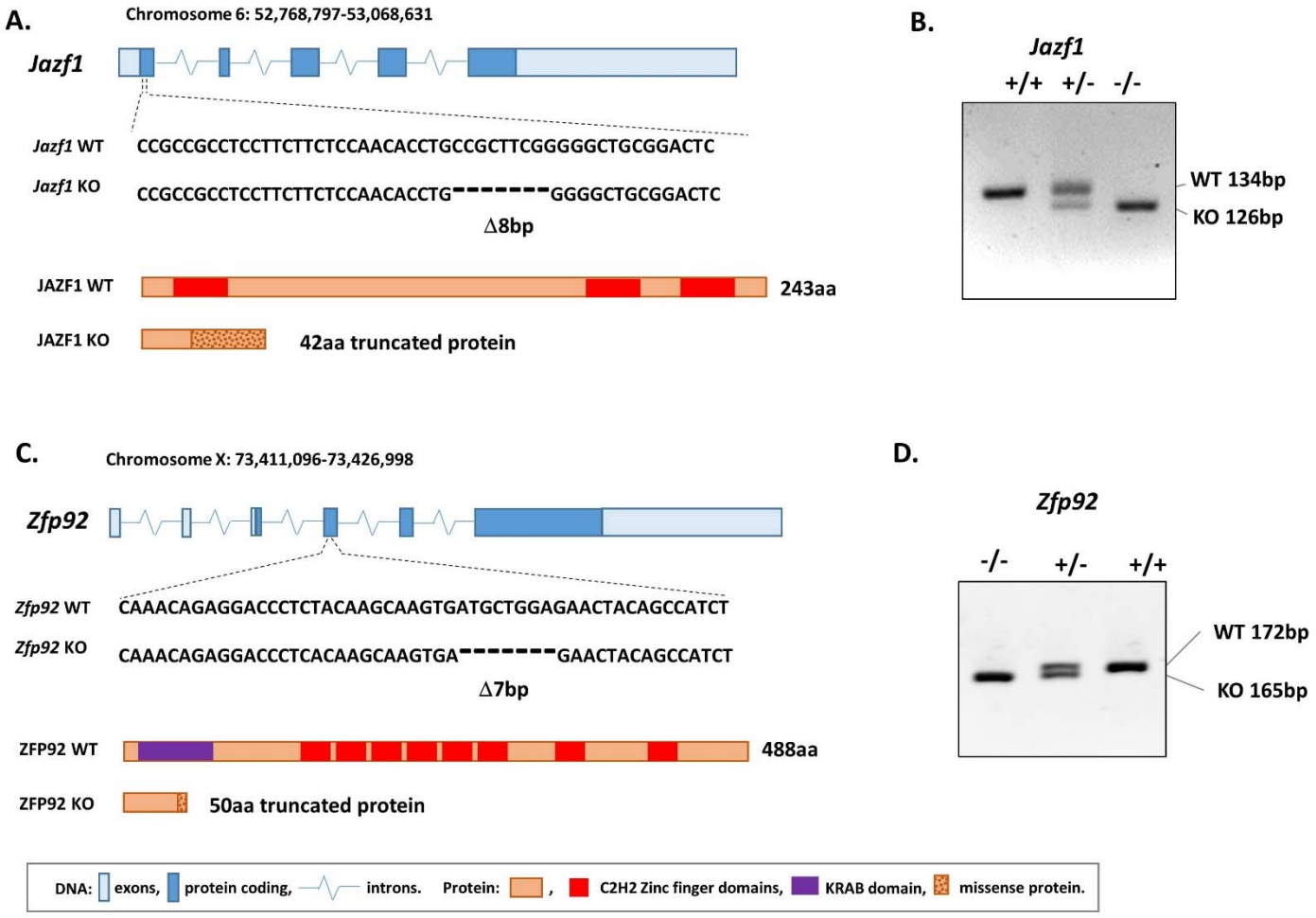


Figure S10. Generation of *Jazf1* and *Zfp92* knockout mice. **A.** Schematic representation of *Jazf1* gene and relative gene location and sequence of CRISPR/Cas9 generated 8bp deletion. The deletion of 8bp generates a frameshift that abrogates normal JAZF1 243aa protein translation and leads to translation of a truncated 42aa peptide with 28 missense amino acids. **B.** Representative image of a PCR genotyping gel for *Jazf1* wild type (WT), heterozygous (Het) and knockout (KO) mice. **C.** Schematic representation of *Zfp92* gene and relative gene location and sequence of CRISPR/Cas9 generated 7bp deletion. The deletion of 7bp leads to a frameshift that abrogates normal ZFP92 488aa protein translation and leads to translation of a truncated 50aa peptide with 8 missense amino acids at the end. **D.** Representative image of a PCR genotyping gel for *Zfp92* wild type (WT), heterozygous (Het) and knockout (KO) mice.

Table S1, supplemental Excel file. **Summary of sequencing results.** Percentage of mapped reads per assay.

Click here to Download Table S1

Table S2, supplemental Excel file. Results from iterative refined WGCNA network analysis. The file includes 2 tables in separate tabs**. 1) Gene Membership:** full listing of the 15,949 expressed genes in the dataset and their module/meta-module membership. This table provides gene names, select gene functional classification annotation (T2D associated expression change, transcription factor, signaling molecule or lincRNA), gene population specific expression, gene meta-module and module assignment in the network, eugengene connectivity (k*Me*), gene expression profile, and gene expression values in each sample (normalized counts). **2) Modules:** details on the 91 module detected by our analysis, including module size, module-density, and module eigengenes.

Click here to Download Table S2

Table S3, supplemental Excel file. Distribution and enrichment of regulatory gene sets across the β-cell GCN modules. Includes 4 tables. **1)** Totals: total count of lincRNAs, transcription factors, and genes in signaling pathways in each module. **2-4)** Enrichment analysis results for lincRNAs, transcription factors, signaling pathways, and T2D genes respectively. Results are sorted by FDR; top-enriched terms are highlighted in gray.

Click here to Download Table S3

Table S4, supplemental Excel file. Zinc-finger proteins associated with endocrine differentiation and β-cell function. The table summarizes transcriptional regulators containing C_2H_2-type Zinc-finger proteins from meta-modules J and D. This table provides mouse gene names, human ortholog gene names, alternative aliases, protein domain composition, status of functional studies, sub-network centrality measures (degree and betweenness centrality calculated within corresponding meta-modules), and a number of published manuscripts based on PubMed search results for mouse and human gene names.

Click here to Download Table S4

Table S5, supplemental Excel file. Results of differential expression analysis of RNAseq datasets from wild type and *Zfp800* knockout pancreata. Differential expression analysis of RNA-seq datasets was done using DEseq2. Expression levels of genes are presented as normalized counts.

Click here to Download Table S5

Table S6, supplemental Excel file. Oligonucleotides used in the study. The table lists oligonucleotide names, sequences, PCR product sizes and applications.

Click here to Download Table S6