HHS Public Access

Author manuscript

Cell Metab. Author manuscript; available in PMC 2020 December 03.

Published in final edited form as:

Cell Metab. 2019 December 03; 30(6): 1091–1106.e8. doi:10.1016/j.cmet.2019.09.013.

The long noncoding RNA *Paupar* modulates PAX6 regulatory activities to promote alpha cell development and function

Ruth A. Singer[1], **Luis Arnes**[2,3,#], **Yi Cui**[4], **Jiguang Wang**[3,&], **Yuqian Gao**[4], **Michelle A. Guney**[5], **Kristin E. Burnum-Johnson**[4], **Raul Rabadan**[3], **Charles Ansong**[4], **Galya Orr**[4], **Lori Sussel**[2,5,6,*]

[1]Integrated Program in Cellular, Molecular and Biomedical Studies, Columbia University Medical Center, New York, NY 10032, USA

[2]Department of Genetics and Development, Columbia University Medical Center, New York, NY 10032, USA

[3]Department of Systems Biology, Columbia University Medical Center, New York, NY 10032, USA

[4]Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA 99354, USA

[5]Barbara Davis Center, University of Colorado Medical Center, Aurora, CO 80045, USA

[6]Lead Contact

SUMMARY

Many studies have highlighted the role of dysregulated glucagon secretion in the etiology of hyperglycemia and diabetes. Accordingly, understanding the mechanisms underlying pancreatic islet α cell development and function has important implications for the discovery of new therapies for diabetes. In this study, comparative transcriptome analyses between embryonic mouse pancreas and adult mouse islets identified several pancreatic lncRNAs that lie in close proximity to essential pancreatic transcription factors, including the *Pax6*-associated lncRNA

*Correspondence: lori.sussel@cuanschutz.edu.

#Current address: Biotech Research and Innovation Centre (BRIC) and The Novo Nordisk Foundation Center for Stem Cell Biology (DanStem), Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

&Current address: Department of Chemical and Biological Engineering, Division of Life Science, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

AUTHOR CONTRIBUTIONS

Conceptualization, R.A.S. and L.S.; Methodology, R.A.S., L.A., Y.J., Y.C., M.A.G., K.E.B.J, R.R., C.A., G.O., and L.S.; Formal Analysis, R.A.S., L.A., Y.J., Y.C., and Y.G.; Investigation, R.A.S., L.A., Y.G., Y.C., M.A.G, and K.E.B.J.; Resources, R.A.S., L.A., and L.S.; Data Curation, R.A.S., L.A., Y.G., Y.C., and K.E.B.J.; Writing – Original Draft, R.A.S. and L.S.; Writing – Review & Editing, R.A.S. and L.S.; Visualization, R.A.S and Y.C.; Supervision, K.E.B.J, R.R., C.A., G.O., and L.S; Funding Acquisition, R.A.S., L.A., and L.S.

DECLARATION OF INTERESTS

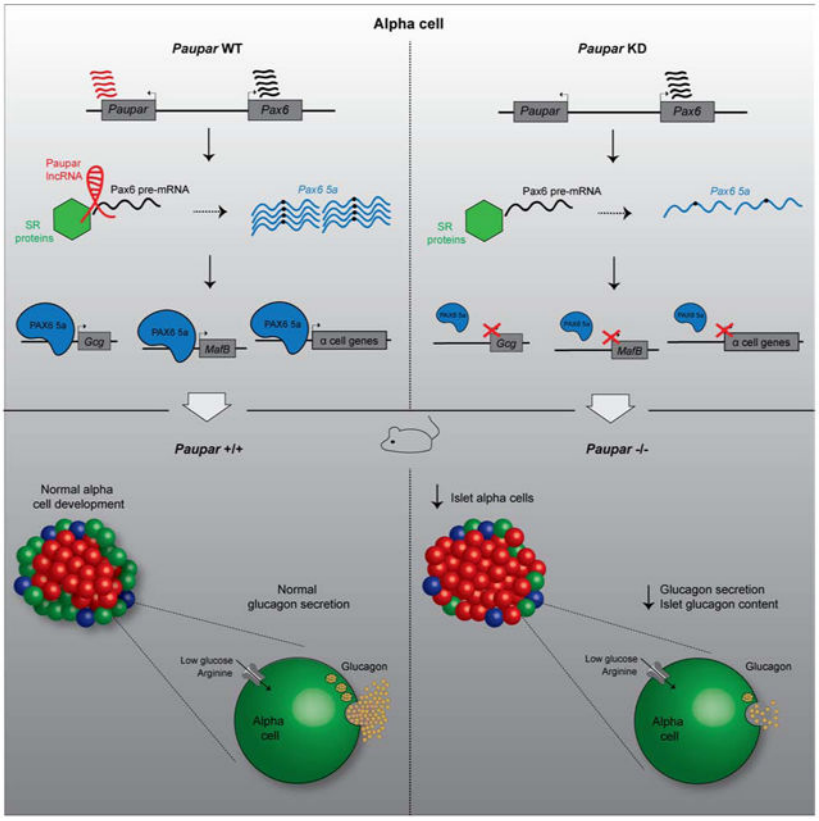
The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Supplemental information includes 7 figures and 5 tables. Tables S1 and S5 could not fit onto three 8.5" × 11" pages, so they have been provided separately as Excel files.

Paupar. We demonstrate that *Paupar* is enriched in glucagon-producing α cells where it promotes the alternative splicing of *Pax6* to an isoform required for activation of essential α cell genes. Consistently, deletion of *Paupar* in mice resulted in dysregulation of PAX6 α cell target genes and corresponding α cell dysfunction, including blunted glucagon secretion. These findings illustrate a distinct mechanism by which a pancreatic lncRNA can coordinate glucose homeostasis by cell-specific regulation of a broadly expressed transcription factor.

Graphical Abstract



IN BRIEF

Singer et al. identify temporally regulated lncRNAs by comparing the transcriptomes of mouse embryonic pancreas to adult islets. One lncRNA, *Paupar*, is enriched in glucagon-producing alpha cells where it confers the cell specific regulatory function of the transcription factor, PAX6, to promote the formation and function of alpha cells.

Keywords

diabetes; pancreatic islets; glucagon; alpha cells; *Paupar*; lncRNAs; long noncoding RNAs; Pax6; transcription factors

INTRODUCTION

Type 1 and type 2 diabetes mellitus (T1D and T2D) are chronic conditions with genetic, immunological, and environmental etiologies that occur due to the failure of the pancreatic islets to maintain glycemic control. Blood glucose homeostasis results from the coordinated, but opposing action of two pancreatic islet-derived hormones, insulin and glucagon. Nutrient ingestion stimulates insulin secretion from islet β cells, which promotes glucose uptake and suppresses liver glucose production. Hypoglycemia stimulates glucagon secretion from islet α cells, which promotes glucose production and release from the liver. While the majority of diabetes treatment options have focused on enhancing β cell function to meet insulin demands, more recent work has shown that α cell dysfunction and hyperglucagonemia also contribute to disease pathophysiology (Unger and Cherrington, 2012; Brissova et al., 2018). Thus, a better understanding of the regulatory mechanisms required for the development and function of these highly-specialized islet endocrine cells will provide important information that could be applied to therapeutic treatments.

The adult endocrine cell populations are all derived from Neurogenin 3 (Neurog3)-expressing endocrine progenitor cells and include insulin-producing β cells, glucagon-producing α cells, somatostatin-producing δ cells, and pancreatic polypeptide-producing PP cells. A large number of transcriptional regulators that are essential for islet cell lineage decisions have been identified and characterized. Many of these factors, including Pdx1 (Offield et al., 1996; Gao et al., 2014), Nkx2-2 (Sussel et al., 1998; Gutiérrez et al., 2017; Churchill et al., 2017), NeuroD1 (Anderson et al., 2009; Naya et al., 1997), and Glis3 (Kang et al., 2016), are expressed in several progenitor and/or islet cell populations and are continuously required in the adult for the maintenance of mature islet cell identity and function (Pan and Wright, 2011; Talchai et al., 2012; Gutiérrez et al., 2017; Ediger et al., 2017). One of the more well-characterized broadly expressed islet transcriptional regulators is the paired and homeodomain transcription factor, PAX6 (St-Onge et al., 1997; Sander et al., 1997). Mice deleted for *Pax6* in pancreatic progenitor cells are born with severely reduced numbers of α, β, and δ cells, suggesting that PAX6 is necessary for the development of multiple islet endocrine cell types (Ashery-Padan et al., 2004; Hart et al., 2013). Conditional deletion of *Pax6* in mature α or β cells demonstrated that PAX6 is also required to maintain the identity and function of mature islet endocrine cells (Ahmad et al., 2015; Gosmain et al., 2012; Mitchell et al., 2017; Swisa et al., 2017). Furthermore, a recent study showed that within mouse β cells, PAX6 directly activates critical β cell genes and represses genes that specify the alternative islet endocrine cell lineages (Swisa et al., 2017). PAX6 also has essential functions in the eye (Shaham et al., 2012) and central nervous system (Manuel et al., 2015). Studies in corneal and epithelial cell lines have shown that splice variants of PAX6 bind distinct DNA motifs (Epstein et al., 1994; Chauhan et al., 2004), yet the molecular mechanism mediating the tissue-, cell-, and gene-specific regulatory functions of PAX6, and other pancreatic transcription factors, remain poorly understood.

While the majority of transcriptional regulatory proteins are expressed in many different cell and tissue types, a class of non-protein coding RNAs (ncRNAs), called long noncoding RNAs (lncRNAs), tend to be highly tissue and cell type specific and temporally regulated

(Carninci et al., 2005; Derrien et al., 2012; Guttman et al., 2011), giving them the potential to confer cell type specificity on broadly expressed transcription factors. In the pancreas, thousands of lncRNAs have been identified in human islets (Moran et al., 2012; Akerman et al., 2017; Fadista et al., 2014; Li et al., 2014), purified human α and β cells (Nica et al., 2013; Bramswig et al., 2013), mouse islets (Ku et al., 2012; Motterle et al., 2017), and purified mouse α and β cells (Ku et al., 2012; Benner et al., 2014), where they exhibit properties consistent with functional genes (Singer and Sussel, 2018). Studies on individual islet lncRNAs have provided evidence that lncRNAs play a role in islet function, primarily through the regulation of essential islet transcription factors, such as *HI-LNC25* on GLIS3 (Moran et al., 2012), β*linc1* on Nkx2-2 (Arnes et al., 2016), and *PLUTO* on PDX1 (Akerman et al., 2017). These findings highlight a role for lncRNAs in islet biology, and their highly restricted expression patterns further suggest that they can influence islet cell specific functions.

To identify developmentally regulated pancreatic lncRNAs, we conducted comparative transcriptome analyses of embryonic mouse pancreas and adult mouse islets and identified 572 dynamically expressed lncRNAs. Our analyses uncovered several pancreatic lncRNAs that lie in close proximity to essential pancreatic transcription factors, including the lncRNA Pax6 Upstream Antisense RNA (*Paupar*), which mapped near the *Pax6* genomic locus in mice and humans (Vance et al., 2014). Interestingly, we found that *Paupar* is enriched in glucagon-producing α cells where it promotes the alternative splicing of *Pax6* to an isoform responsible for PAX6-dependent activation of essential α cell genes. Consistent with a role for *Paupar* in conferring the α-cell specific functions of PAX6, deletion of *Paupar* in mice resulted in dysregulation of PAX6 α cell target genes and α cell dysfunction. Our findings illustrate how lncRNAs can modulate transcription factor activities to achieve cell-specific regulation.

RESULTS

Identification of developmentally regulated lncRNAs in the mouse pancreas

To identify temporally and spatially regulated lncRNAs in the pancreas, we performed RNA-sequencing (RNA-seq) on adult mouse islets and embryonic day 15.5 (e15.5) mice pancreas (EP). We analyzed these datasets using a computational pipeline (Pefanis et al., 2015) designed to identify putatively functional mouse pancreatic lncRNAs (Figure 1A) (Singer and Sussel, 2018). A set of stringent criteria was used to define 2728 pancreatic lncRNAs: (1) > 200 nucleotides (nt) in length; (2) no overlap with protein coding regions; (3) no overlap with pseudogenes (Karro et al., 2006); and (4) low predicted coding probability (Wang et al., 2013) (Figure 1B). We then filtered out genes with a Fragments Per Kilobase of transcript per Million mapped reads (FPKM) < 0.5 to enrich for lncRNAs amenable to molecular analyses (Figure 1C). These parameters yielded 572 high-confidence pancreatic lncRNAs that clustered according to developmental stage (Table S1; Figure 1D). Remarkably, comparative transcriptomics between the two stages showed that approximately half (279 or 48.7%) of all pancreatic lncRNAs were developmentally regulated: 108 lncRNAs were significantly enriched in EPs and 171 lncRNAs were significantly enriched in adult islets (Figure 1E). Given that the embryonic samples came from whole pancreas, we

further verified that the majority of embryonically enriched lncRNAs were expressed in e15.5 Neurog3+ endocrine progenitor cells (Churchill et al., 2017; Table S1).

Previous studies have shown that subsets of lncRNAs regulate, or are co-regulated with, nearby protein-coding genes (Cabili et al., 2011; Moran et al., 2012). To gain insight into the function of these lncRNAs, we utilized Genomic Regions Enrichment of Annotations Tool (GREAT) (McLean et al., 2010) to identify nearby protein-coding genes (Figure 1F). Several lncRNAs mapped within 5 kb of a nearby transcriptional start site (TSS), suggesting they might function as bidirectional lncRNAs (Mercer, Dinger, and Mattick, 2009), while the majority of lncRNAs were located in intergenic regions (> 50 kb from a TSS) (Figure S1A). Gene ontology analyses of all neighboring genes showed a significant enrichment of genes involved in “endocrine pancreas development” and “pancreas development” (Figure 1G), including nine essential pancreatic transcription factors: *Pax6, Nkx6.1, NeuroD1, Hes1, Foxo1, Mnx1, Foxa2, Meis2*, and *Hnf6* (Table S2).

Paupar is a nuclear lncRNA enriched in pancreatic α cells

One of the more differentially regulated lncRNAs (Figure 1E; green square) was the *Pax6* Upstream Antisense RNA (*Paupar*), a previously characterized lncRNA located near the *Pax6* locus (Vance et al. 2014) with an orthologous transcript in human islets (HI-LNC101) (Moran et al., 2012). *Paupar* is a 3482 bp gene transcribed 8 kb upstream and antisense from *Pax6* and contained within the first intron of *Pax6os1*, an antisense RNA with no known function (Figure 2A). *Paupar* lies in a syntenically conserved region on chromosome 2 in mice and chromosome 11 in humans (UCSC LiftOver). *Paupar* is also highly conserved at the nucleotide level across mammals (89% in humans; Table S2), both in its gene body and regions corresponding to a putative promoter (Figure 2A), similar to many functional lncRNAs (Carninci et al. 2005). Comparative sequence analysis of *Paupar* across species did not reveal any conserved ORFs, reinforcing the computational prediction that *Paupar* lacks protein-coding potential (PhyloCSF - 6.5127; CPC 0.2666; CPAT 0.2086). In addition to being developmentally regulated and conserved in humans, we determined *Paupar* expression was reduced 2-fold in islets from *db/db* mice, a model of T2D, suggesting *Paupar* might play a role in islet dysfunction (Figure 2B).

Paupar was previously identified as a single exonic lncRNA in the neuroblastoma N2A cell line, where it appears to have PAX6-dependent and independent functions (Vance et al., 2014). To more globally characterize *Paupar* expression, we performed extensive analysis of published RNA-seq datasets from 25 different mouse tissues and cell lines (Lin et al., 2014; Lawlor et al., 2017) and found that *Paupar* was expressed exclusively in the pancreas, eye, and brain (Figure S2A). These datasets also confirmed the findings of our initial screen, showing enrichment of *Paupar* in adult islets compared to embryonic pancreas (Figure S2A and Figure 1E). Furthermore, expression data from two commonly used immortalized islet cell lines, αTC-1 and βTC-6 cells (Powers et al., 1990; Hamaguchi and Leiter, 1990), that model α and β cells, respectively, suggested *Paupar* expression is enriched in α cells (Figure S2A). Surprisingly, in contrast to what was reported in N2A neuroblastoma cells, in the context of islets, αTCs, βTCs, and the eye, *Paupar* appears to contain three exons and two introns (Figures S2B and S2C).

We confirmed the RNA-Seq data using quantitative reverse transcription PCR (qRT-PCR) on RNA from 11 different mouse tissues to demonstrate that *Paupar* was significantly enriched in adult mouse islets, compared to the eye (~3.5-fold) and brain (~25-fold) (Figure 2C). Expression analyses on whole pancreata at several embryonic and postnatal developmental time points also revealed that onset of *Paupar* expression occurs between postnatal day 7 (P7) and P14 (Figure 2D), which is a critical window of postnatal development associated with the functional maturation of endocrine cells (Nishimura et al., 2006; Hang and Stein, 2011). Given that PAX6 has known regulatory roles in both α and β cells (Gosmain et al., 2010; Swisa et al., 2017), we wanted to confirm the αTC and βTC expression data that suggested *Paupar* expression was enriched in the α cell lineage (Figure S2A). We isolated islets from mice expressing the Glucagon-Venus reporter gene (Reimann et al., 2008) and used FACS to purify α cells from the other islet cell types. Expression analysis of FACS purified mouse α cells and non-α islet cells confirmed *Paupar* was highly enriched in α cells compared to other islet cell types, in contrast to *Pax6*, which is expressed equally in both cell populations (Figure 2E). Expression levels of *Gcg and Ins2*, essential α and β cell genes, respectively, are shown to demonstrate the purity of the samples (Figure 2E).

Given that the *Paupar* locus is highly conserved at the sequence level across placental mammals, we wanted to determine whether *PAUPAR* expression is conserved in human islets. Analyses of published RNA-seq datasets (Blodgett et al., 2015) showed that human *PAUPAR* was significantly enriched in adult human islets compared to fetal pancreas (Figure S2D), similar to *Paupar* expression in mice (Figure 2D). However, unlike its restricted expression in mouse islet α cells, *PAUPAR* was expressed in both human α and β cells (Figure S2E). This may reflect the different gene expression patterns observed between mouse and human islet cells, similar to MafB/MAFB, which is α-cell specific in mice but is expressed in both α and β cells in humans (Hang and Stein, 2011) (Figure S2F). Consistently, we determined that *PAUPAR* expression was positively correlated with both *MAFB* (Figure S2G) and *GCG* (Figure S2H) expression in human islets. Since *Paupar* is downregulated in *db/db* islets, we evaluated human *PAUPAR* expression in islets from T2D patients, as well as islets from healthy patients cultured in low or high glucose concentrations (Figures S2I and S2J). Expression analyses did not reveal a significant difference in *PAUPAR* in T2D islets or in human islets exposed to a glucose challenge (Figures S2I and S2J), which could be due to considerable gene expression variability between human donors or a result of *PAUPAR* having unique regulatory functions in human α versus β cells.

LncRNA localization within a cell can also yield important mechanistic insight; nuclear lncRNAs often regulate transcription and pre-mRNA processing, while cytoplasmic lncRNAs more likely influence mRNA stability and translation (Batista and Chang, 2013). Expression analyses on αTCs following cell fractionation revealed a significant enrichment of *Paupar* in the nucleus compared to the cytoplasm (Figure 2F). Nuclear localization of *Paupar* in αTC cells was confirmed by single-molecule fluorescent in situ hybridization (smFISH) using oligonucleotide probes targeting the full length *Paupar* transcript (Figure 2G). In comparison, glucagon protein is localized to the cytoplasm (Figure 2G). Analysis of *Paupar* smFISH images determined that the average copy number for *Paupar* is 54 ± 31 transcripts per αTC cell (Figure S2K). As a comparison, *Paupar* smFISH in MIN6 beta cell

line showed no specific staining above background levels (Figure S2L). Taken together, these results demonstrate *Paupar* is a conserved lncRNA predominantly localized to the nuclei of mature pancreatic α cells.

Paupar regulates Pax6 α cell target genes

The expression profile of *Paupar* prompted us to investigate its regulatory function in α cells. Three unique sets of antisense oligonucleotides (ASO) successfully downregulated *Paupar* RNA in αTC cells by an average of 57% (Figure 3A). In contrast to what has been previously reported in N2A cells, we observed no change in *Pax6* expression following *Paupar* knockdown (Figure 3A), suggesting that *Paupar* RNA does not influence *Pax6* transcription. We were surprised then, to discover that *Paupar* knockdown led to the downregulation of several canonical *Pax6* α cell target genes, including *Gcg, MafB, Arx*, and *NeuroD1* (Gosmain et al., 2010) (Figure 3A). Furthermore, the amount of *Paupar* KD was positively correlated with the reduction in both *Gcg* (Figure 3B) and *MafB* (Figure 3C) expression. These findings suggest that PAX6-mediated regulation of α cell genes is sensitive to the level of *Paupar* present in αTCs. The expression of *Foxa2*, an α cell gene that is not a target of PAX6, was not altered by loss of *Paupar*, further demonstrating that the activity of *Paupar* is restricted to PAX6 target genes (Figure 3A). Consistently, co-expression analysis of human islets (Fadista et al., 2014) showed *PAUPAR* expression was positively correlated with the expression of both *MAFB* (Figure S2G) and *GCG* (Figure S2H), suggesting *Paupar* may have similar regulatory functions in human islets.

Paupar lncRNA interacts with nuclear proteins involved in alternative splicing

To investigate the molecular mechanism by which *Paupar* regulates α-cell PAX6 target genes, we performed Capture Hybridization Analysis of RNA Targets (CHART) (Simon et al., 2011) to identify the *Paupar* interactome. We used two unique sets of biotinylated capture oligos (COs) to pull down *Paupar*, in addition to control COs (scrambled and *Paupar* sense). Both sets of *Paupar* COs retrieved *Paupar* RNA from αTC nuclear extract (Figures 3D and 3E). Importantly, *Paupar* COs did not pull down *Gapdh* or *TBP*, nor did control COs retrieve *Paupar* (Figures 3D and 3E). The lack of *Paupar* enrichment with sense COs, which cannot directly target the RNA but have sequence identity to the DNA locus, demonstrates that the CHART-based enrichment is RNA mediated (Figure S3A).

To identify *Paupar*-interacting proteins, we performed CHART followed by mass spectrometry (CHART-MS), and identified 56 *Paupar* binding proteins that were enriched > 2-fold over control COs (Table S3). Unlike in N2A cells, we were not able to detect an interaction between *Paupar* and PAX6 protein (Figure S3B). Analysis of *Paupar* interacting proteins using a database of known and predicted protein-protein interaction networks (Search Tool for the Retrieval of Interacting Genes/Proteins (STRING); Szklarczyk et al., 2017) showed that 25 out of the 56 *Paupar*-interacting proteins were predicted to interact with at least one other *Paupar*-interacting protein (Figure 3F). These protein complexes were significantly enriched in several pathways (Figure 3G), including RNA splicing (Figure 3F, blue circles), regulation of gene expression (Figure 3F, red circles), and DNA binding (Figure 3F, green circles). Since the largest complex of interacting proteins contained canonical regulators of alternative splicing (Figure 3F), we compared the set of *Paupar*

interacting proteins with those retrieved by *NEAT1* and *MALAT1*, two nuclear lncRNAs with established roles in alternative splicing (West et al., 2014). This analysis identified 50.6% and 51.7% overlap with *NEAT1* and *MALAT1* interacting proteins, respectively (Figure S3C). In contrast, similar comparison to proteins or lncRNAs that are not involved in alternative splicing mechanisms, MAFA (Scoville et al., 2015), NEUROD1 (Romer et al., 2019) and *Xist* (Chu et al., 2015), showed limited overlap (Figure S3C).

Consistent with a direct interaction between *Paupar* and 8 of the 12 annotated serine and arginine rich splicing factors (SRSFs) (Figure 3F; Table S3), we used a computational tool (RBPmap; Paz et al., 2014) to predict that the full-length *Paupar* transcript had 1685 putative SRSF binding sites above a stringent z-score threshold (> 2). Notably, several of the most significant SRSF sites were located within a region of the *Paupar* locus that has high sequence conservation across placental mammals (Figure 3H, grey rectangle). To address the possibility that SRSF binding occurs because *Paupar* is a spliced transcript, we compared the z-scores for *Paupar* SRSF binding sites to *Nkx2-2*, an mRNA that is spliced but does not regulate RNA splicing, as well as *Malat1*, a lncRNA that is not spliced but regulates SRSF-mediated RNA splicing. We found that *Paupar* had more SRSF binding sites with higher z-scores than both *Nkx2-2* and *Malat1* (Figure 3H). SRSF binding sites for the full ~7 kb *Malat1* locus is shown in Figure S3D. Taken together, these findings suggest that *Paupar* functions in the α cell to regulate alternative splicing.

Paupar promotes the alternative splicing of Pax6 to the isoform required for activation of Pax6 α cell target genes

Previous studies have demonstrated that the regulatory role of *Paupar* in N2A cells is partially mediated through PAX6 (Vance et al., 2014; Pavlaki et al., 2017). The identification of interactions between *Paupar* and the SRSF family of proteins, and the knowledge that *Pax6* has two well-characterized isoforms with distinct regulatory functions (Epstein et al., 1994; Chauhan et al., 2004; Kiselev et al., 2012; Sasamoto et al., 2017) (Figure 4A), led us to examine a possible role for *Paupar*-mediated alternative splicing of *Pax6*. The two major *Pax6* isoforms, termed *Pax6* (Figure 4A, blue lines) and *Pax6 5a* (Figure 4A, red lines), differ from each other by an alternatively spliced exon, “5a”, that adds 14 amino acids to the paired DNA-binding domain of PAX6 protein and alters DNA binding recognition (Kiselev et al., 2012). Since PAX6 regulates distinct target genes in α versus β cells (Gosmain et al., 2010; Gosmain et al., 2012), we hypothesized that its cell-specific regulatory activities could be mediated through its different isoforms. Consistently, computational analysis of alternative splicing events in α versus β cell transcriptomes (DiGruccio et al., 2016) identified the *Pax6 5a* isoform as significantly enriched in α cells ($p < .0001$; data not shown). While non-quantitative RT-PCR analysis detects both isoforms in α cells (Figure 4B), we could only quantify expression of *Pax6 5a*, given that amplification of the shorter *Pax6* isoform produces both isoform products (Figure 4A, black arrows), which confounds qRT-PCR analysis. To directly test whether *Paupar* promotes alternative splicing of *Pax6*, we performed qRT-PCR on RNA from *Paupar*-deficient αTCs. Strikingly, while *Paupar* KD did not induce a change in total *Pax6* mRNA levels, we did observe a significant and specific reduction (45%) of *Pax6 5a* (Figure 4C). We also determined that *Paupar*-mediated regulation of *Pax6* splicing occurs directly though RNA-RNA interaction, rather than

indirectly through SR splicing factors, since the association between *Paupar* and *Pax6* RNA could be partially ablated by RNase treatment (Figure 4D).

A study in non-pancreatic cell lines that stably expressed *Pax6* or *Pax6 5a* showed that the two proteins have different DNA binding specificities (Kiselev et al., 2012). To determine whether *Paupar* mediated-alternative splicing of *Pax6 5a* is necessary for direct PAX6 regulation of α cell target genes, we performed PAX6 ChIP-sequencing (ChIP-seq) in αTC cells with normal or reduced amounts of *Paupar* and assessed changes in PAX6 global occupancy. In control α cells, PAX6 was physically associated with 26244 genomic sites; notably, approximately 10% of the PAX6 sites (2046 peaks) showed differential binding following *Paupar* KD (Figure 4E). Interestingly, regions that exhibited differential binding upon *Paupar* KD were underrepresented in regions < 1 kb from promoters and more predominantly located in intergenic regions (> 3 kb from 5' or 3' UTR) (Figure 4F). Gene ontology (GO) analysis of the genes located within 100kb of the 2046 differentially bound PAX6 sites showed a significant enrichment in biological processes associated with α cell function, including: glucose homeostasis, regulation of ion transport, and cellular response to glucose stimulus (Figure 4G).

To determine whether *Paupar* KD caused a global shift in PAX6-mediated regulation of α cell genes, we compared genes associated with the differentially bound PAX6 sites with a published dataset of α cell enriched genes (DiGruccio et al., 2016). We observed a statistically significant overlap of 60 genes between the datasets (Figure 4H), which likely underrepresents the number genes given that the ChIP-seq peak annotation strategy excludes some distal regulatory elements. Importantly, the overlapping 60 genes contained several genes misregulated by *Paupar* KD, including *Gcg* and *MafB* (Figure 3A). We further confirmed these findings using ChIP-qPCR for PAX6 on the promoters of *Gcg* and *MafB* (Gosmain et al., 2007; Gosmain et al., 2010). As expected, in control αTCs there was a greater than 30-fold enrichment of PAX6 over IgG on *Gcg* and *MafB* regulatory DNA (Figure S3E), while *Paupar* KD induced a 34% and 58% reduction in PAX6 occupancy on the *Gcg* and *MafB* promoters, respectively (Figure S3E). These results, combined with our finding that *Paupar* KD causes specific downregulation of *Pax6 5a* but not total *Pax6* mRNA (Figure 4E), directly demonstrate that *Paupar* confers the α-cell specific regulatory function of *Pax6 5a* via alternative splicing.

Paupar knockout mice have impaired α cell development and function

To determine whether *Paupar*'s regulation of *Pax6* isoform selection affected α cell function in vivo, we generated *Paupar* null (KO) mice by replacing the endogenous *Paupar* locus with the histone-fusion GFP (H2B:GFP) reporter gene (Kanda et al., 1998) (Figures S4A and S4B). Expression analysis of islets from *Paupar* KO mice confirmed the complete loss of *Paupar* RNA (Figure S4C), and immunofluorescence analyses indicated that the GFP reporter recapitulated endogenous *Paupar* expression specifically in the glucagon-producing α cells (Figures S4D and S4E).

Paupar KO mice are viable, fertile, and indistinguishable from their WT littermates with respect to weight (Figure S5A), *ad libitum* blood glucose (Figure S5B), and glucose tolerance (Figure S5C-E). These findings are not surprising given the restricted expression

of *Paupar* in α cells; several studies have shown that mice are highly resistant to perturbations in α cell function (Furuta et al., 2001; Shiota et al., 2003; Heller et al., 2004; Hancock et al., 2010; Wilcox et al., 2013). To overcome this limitation of the mouse model, hypoglycemia, a physiological trigger for glucagon secretion from α cells, can be used to interrogate α cell function in vivo. Consistent with an α cell defect, we showed that 6-week-old *Paupar* KO mice were significantly hypoglycemic compared to WT controls following an overnight fast, suggesting *Paupar* KO mice may exhibit impaired glucagon secretion in response to fasting-induced hypoglycemia (Figure S5F). We further utilized an in vivo insulin tolerance test (ITT) to assess the ability of *Paupar* KO mice to respond to acute hypoglycemia. Strikingly, throughout the ITT, 6-week-old *Paupar* KO mice were significantly hypoglycemic compared to WT mice, suggesting KO mice were less efficient at returning to baseline blood glucose levels (Figures 5A and 5B; Figure S5I). To rule out the possibility that hypoglycemia observed in *Paupar* KO mice during an ITT was due to enhanced insulin sensitivity, we calculated the Homeostatic Model Assessment of Insulin Resistance (HOMA-IR), a well-established measure of insulin resistance (Matthews et al., 1985), in fasted *Paupar* WT and KO mice. As expected, given the absence of *Paupar* expression in peripheral metabolic tissues (Figure S2A), there was no difference in HOMA-IR in *Paupar* KO mice compared to control mice (Figure S5J). Alternatively, the inability of *Paupar* KO mice to respond to hypoglycemia could be due to impaired glucagon secretion. To test this, we measured plasma glucagon before and during an ITT. Thirty minutes after an insulin injection, *Paupar* KO mice had ~2.5-fold decrease in plasma glucagon relative to WT mice (Figure 5C), demonstrating that *Paupar* KO mice secrete significantly less glucagon in response to hypoglycemia.

The physiological response to an insulin challenge is complex and involves several non-pancreatic tissues, including liver, muscle, and fat. Although *Paupar* is not expressed in those tissues (Figure S2A), to eliminate the possibility that the impaired response of *Paupar* KO mice to hypoglycemia was due to increased uptake of glucose by the peripheral tissues, we performed *ex vivo* glucagon secretion assays. Cultured islets from 6-week-old *Paupar* KO mice secreted ~60% less glucagon than WT islets in response to low glucose (2 mM) and ~47% less glucagon than WT islets in response to 10 mM arginine, both potent glucagon secretagogues (Gerich, Charles, and Grodsky, 1974) (Figure 5D). To determine whether this blunted glucagon secretion was due to reduced islet glucagon content, we measured total islet glucagon content and found that *Paupar* KO islets had dramatically less (71%) total glucagon than controls (Figure 5E). Interestingly, when we then normalized *ex vivo* glucagon secretion (Figure 5D) to islet glucagon content (Figure 5E), we discovered that *Paupar* KO islets cells inappropriately hypersecrete glucagon in high glucose conditions (Figure S5K). These findings demonstrate that impaired α cell physiological function in *Paupar* KO islets is at least partially due to decreased glucagon content; however, it does not rule out defects in glucose sensing, membrane depolarization, glucagon production, or exocytosis.

Consistent with the physiological phenotype, morphometric analysis on 6-week-old *Paupar* WT and KO pancreata revealed that while α cells made up ~13% of *Paupar* WT islets (Figure 5F), they comprised only ~6% of *Paupar* KO islets (Figure 5G), corresponding to an average 2.25-fold decrease in the α cell population (Figure 5H). We also observed a modest

increase in β cell area relative to islet area in *Paupar* KO mice (~79%) compared to WT mice (~71%) (Figures 5I-5K). There was no measurable difference in average islet size between *Paupar* WT and KO mice (Figures 5L-5N). Examination of islet morphology in 7-month-old WT and KO mice showed that *Paupar* KO mice had significant α cell (Figures S6A and S6B) and islet hyperplasia (Figure S6C) due to increased α cell proliferation (Figures S6D-S6F). We further showed that these hyperplasic α cells are functional, given that 3- and 6-month-old *Paupar* KO mice were no longer hypoglycemic compared to controls in response to an overnight fast (Figures S5G and S5H). Furthermore, when we conducted ITTs on older (>6 months) mice, we discovered that throughout the assay, *Paupar* KO mice were hyperglycemic compared to control mice (Figures S7A-C), which was the opposite of the phenotype in 6-week-old *Paupar* KO mice (Figures 5A and 5B; Figure S5I). This compensation phenomenon has been well documented in several mouse models of α cell dysfunction (Gelling et al., 2003; Conarello et al., 2007; Hayashi et al., 2009; Courtney et al., 2013; Solloway et al., 2015) and provides further evidence that *Paupar* functions similarly to canonical α cell regulatory proteins. Taken together, these results demonstrate that *Paupar* is required for normal α cell development and function.

Paupar regulates essential α cell genes *in vivo*

To validate our *in vitro* studies which used an RNA knockdown approach to disrupt *Paupar* activity in αTC cells, we performed global transcriptome analyses on 6-week-old isolated islets from *Paupar* WT and KO mice which revealed 3106 differentially expressed genes (DEGs) (p-value < .05) (Figure 6A). Consistent with the α-cell specific phenotype, the DEGs represented > 10% (75/689) of all documented α cell enriched genes (DiGruccio et al., 2016) (Figure 6B). Furthermore, the DEGs included many factors required for α cell function and many known PAX6 α-cell targets, including transcription factors (*Arx, MafB, Irx1*, and *Irx2*) (Table S4), voltage-gated ion channels (*Kcnq2, Kcnip3*), and exocytotic machinery (*Sytl5* and *Sytl2*) (Figure 6C). In addition, the majority (63/75) of the dysregulated α-cell specific genes were downregulated in *Paupar* KO mice, indicating that *Paupar* predominantly regulates gene activation in α cells. Of note, *Paupar* KO islets had normal levels of *Pax6os1*, as well as genes within 100 kb of *Paupar*. These findings, along with experiments showing similar transcriptional changes in *Paupar* KO islets and *Paupar*-deficient αTCs, demonstrate that the *Paupar* KO mouse phenotype is not due to deletion of an important DNA regulatory element.

To specifically assess the regulatory function of *Paupar* in α cells, we isolated islets from *Paupar* WT and KO mice expressing the Glucagon-Venus reporter gene (Reimann et al., 2008) and used FACS to purify α cells from the other islet cell types. Consistent with our *in vitro* studies, α cells from *Paupar* KO mice showed a significant downregulation of *Pax6 5a*, as well as several PAX6 α-cell target genes, while total *Pax6* mRNA levels were unchanged (Figure 6D). These analyses provide further evidence that *Paupar*-mediated regulation of PAX6 α-cell target genes is at the level of *Pax6* splicing rather than transcription.

To expand our understanding of *Paupar*-mediated regulation of PAX6 target genes, we determined the overlap between genes regulated by *Paupar* in islets (3056 genes; p-value < 0.05) and genes associated with altered PAX6 binding following *Paupar* KD in αTCs (2056

genes; p-value < 0.01). This analysis revealed a statistically significant overlap of 328 genes between the datasets (Figure 6E), including many essential α cell genes, such as *MafB, Arx, Irx1* and *Irx2*, whose misregulation could directly explain the *Paupar* KO mouse phenotype. Based on these cumulative studies, we propose a model that *Paupar* promotes α cell function via alternative splicing of the *Pax6 5a* isoform required for regulation of PAX6 α cell target genes (Figure 7).

DISCUSSION

During the past 20 years, a major focus of diabetes research has been directed towards identifying the complex transcription factor networks required for the specification and function of pancreatic islet cells. Yet, how these broadly expressed transcription factors acquire unique regulatory functions at different stages of pancreas development and in different islet cell types remains poorly understood. In this study we uncover a novel mechanism by which the lncRNA, *Paupar*, confers cell specific regulatory function on the essential pancreatic transcription factor, PAX6, by promoting the alternative splicing of *Pax6* to the *Pax6 5a* isoform that is required for the regulation of downstream α cell target genes (Figure 7). We have shown that the loss of *Paupar* blunts the production of *Pax6 5a* isoform, causing misexpression of PAX6 target genes and impaired glucagon-mediated glucose homeostasis (Figure 7). These findings uncover a novel layer of islet gene regulation and provide further evidence that lncRNAs are fundamental players in islet development and function.

Paupar was first identified in neuroblastoma N2A cells as a single exon lncRNA that regulated genes independently and through direct interaction with PAX6. Surprisingly, our expression analyses showed that *Paupar* was most highly enriched in pancreatic islets compared to all other tissues examined, including the eye, brain, and N2A cells. Furthermore, unbiased mapping of the adult islet transcriptome showed evidence of splicing within the *Paupar* locus; cloning and sequencing confirmed that the *Paupar* transcript has three exons and two introns. We also discovered that *Paupar* does not directly interact with PAX6, nor does it regulate *Pax6* expression at the transcript level. These discrepancies between our findings and previous studies in N2A cells possibly reflect tissue-specific regulation and functional activities of *Paupar*. Additionally, although *Paupar* is also expressed in the mouse eye and brain, we did not observe any gross abnormalities or phenotypes associated with these tissues in *Paupar* KO animals. It is also not likely that the loss of *Paupar* in the brain contributes to the α cell phenotype, since we observed impaired glucagon secretion and a significant decrease in glucagon content in isolated islets of *Paupar* KO mice. Furthermore, *Paupar*-mediated regulation of α cell genes *in vitro* was consistent with the observed *in vivo* phenotypes.

Within the pancreas, *Pax6* is expressed in several islet cell types, while *Paupar* expression is restricted to mature α cells. Although extremely low levels of *Paupar* transcript could be detected in β cells, *Paupar* knockdown experiments in the MIN6 beta cell line resulted in minimal gene expression changes (this study- GSE132072), confirming its primary role in α cells. Intriguingly, the onset of *Paupar* pancreatic expression between P7 and P14 corresponds to a critical developmental window during which cells acquire mature

transcriptional profiles. For example, during this postnatal window the transcription factors *MafA* and *MafB* become restricted to mature β and α cells, respectively. These findings, along with the dramatic reduction in α cells seen in 6-week-old *Paupar* KO mice, and experiments showing reduced PAX6-activation of *MafB* in *Paupar*-deficient αTC cells, demonstrate that *Paupar* is required for the differentiation of mature α cells by promoting PAX6-mediated activation of *MafB*.

Previous studies have shown that PAX6 functions as a transcriptional activator and a repressor; however, the mechanism that mediates this dual capability in a single cell type is unknown. The discovery that *Paupar* expression specifically in α cells promotes the alternative splicing of *Pax6* to an isoform with altered DNA binding specificity raises the interesting possibility that the different PAX6 isoforms have unique functions. Since were able to detect both the *Pax6* and *Pax6 5a* isoforms within α cells, it is possible that the ratio of each isoform is crucial for proper gene regulation. This idea is supported by our finding that reduction of the *Pax6 5a* isoform in *Paupar*-deficient α cells corresponded to reduced expression of several canonical PAX6-activiated α cell genes, including *glucagon* and *MafB* (Gosmain et al., 2007; Gosmain et al., 2010), while *ghrelin*, the only gene known to be repressed by PAX6 in α cells (Ahmad et al., 2015), was not upregulated in *Paupar* KO islets or in *Paupar*-deficient αTC cells (Figure 6C and data not shown). Furthermore, the finding that *Paupar* KO mice phenocopy mice deleted for *Pax6* in α cells (Ahmad et al., 2015) suggests the primary mechanism of *Paupar* in the islet is to regulate the PAX6 α cell specific transcriptional program.

Two recent studies examined the transcriptome of purified α cells from T1D donors (Brissova et al., 2018) and single islet cells from T2D donors (Segerstolpe et al., 2016). Although these datasets confirmed *PAUPAR* is expressed in human α cells, *PAUPAR* expression was not altered in the α cells derived from either T1D or T2D patients. A caveat of these studies was the small sample size analyzed; however, consistent with *PAUPAR* regulatory function in mice, *MAFB* and *GCG* expression were also not affected in these datasets. It is therefore possible that *PAUPAR* plays a conserved regulatory role in human α cells, but we still lack the comprehensive understanding of normal human α cell function that is necessary to make conclusions about the role of *PAUPAR* in α cell dysfunction.

While several lncRNAs have been shown to influence alternative splicing through their direct interactions with splicing factors (Romero-Barrios et al., 2018), this study is the first to demonstrate a specific interaction between a cell-restricted lncRNA and its *cis*-related gene that in turn influences the production of a distinct protein isoform. In particular, the presence of *Paupar* in islet α cells skews the alternative splicing of *Pax6* to favor an isoform required for α-cell specific gene regulation. These results highlight an important mechanism through which tissue restricted lncRNAs influence transcription factor target selection to confer cell specific activities. The identification and characterization of additional cell restricted lncRNAs will be instrumental in determining the extent of this gene regulatory mechanism. We predict that additional evidence of *cis* lncRNA-regulation of tissue-specific mRNA splicing will emerge, since many lncRNAs function in the same pathway as their neighboring gene, even though there is often no evidence of regulation at the transcriptional level. In summary, our extensive molecular and functional characterization of an α cell

enriched lncRNA suggest that lncRNAs could represent important tissue and/or cell restricted therapeutic targets to regulate the production and function of cell-specific isoforms of more widely expressed proteins.

LIMITATIONS OF STUDY

Our finding that *Paupar* is predominantly enriched in alpha cells presented several experimental challenges for the molecular analyses since many biochemical assays still require cell numbers that are orders of magnitude beyond what can be reasonably obtained in vivo. Given the relative paucity of alpha cells in mouse islets and their further reduction in the *Paupar* KO mice, it was necessary to rely on an immortalized alpha cell line for several of the mechanistic experiments. When possible, we corroborated our in vitro findings on purified alpha cells from *Paupar* WT and KO mice. Furthermore, our efforts to investigate human *PAUPAR* functions were limited by the absence of a human alpha cell line, as well as our finding that single cell transcriptome data from human islets often lacked the sequencing depth needed to detect low abundance transcripts, such as *PAUPAR* and the majority of tissue-specific lncRNAs. The development of additional experimental models and improved techniques to analyze human diabetic alpha cells would increase the clinical impact of our present findings.

STAR METHODS

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to, and will be fulfilled by, the Lead Contact, Dr. Lori Sussel (lori.sussel@cuanschutz.edu). All unique reagents and mouse lines generated in this study are available from the Lead Contact with a completed Materials Transfer Agreement.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell line and cell culture—αTC1 clone 6 cells (ATCC, CRL-2934) were cultured in DMEM supplemented with 10% FBS, 15 mM Hepes, 0.1 mM non-essential amino acids, 0.02% BSA, 1.5 g/L sodium bicarbonate, 2.0 g/L glucose, and 1% antibiotic-antimycotic. MIN6 cells (Miyazaki et al., 1990) were cultured in DMEM supplemented with 10% FBS and 1% Penicillin-Streptomycin. All cell lines were passaged and maintained following standard techniques at 37°C in 5% CO2 and 95% air.

Mouse studies—We generated the *Paupar* knockout allele using BAC recombineering. First, Gibson assembly was used to clone H2BGFP (Kanda et al., 1998) and a polyA sequence and insert them together into the pL451 plasmid containing flox Neo flox (FNF) (Nam and Benezra, 2009). Short (80 bp) arms homologous to the regions flanking *Paupar* were then added to H2BGFPpA-FNF by PCR. The BAC clone (RP23-465J7; BAC-PAC Resources) was modified by insertion of the H2BGFPpA-FNF into the *Paupar* locus using Cre from SW106 cells, which was then retrieved into pMCS-DTA (a gift from Dr. Kosuke Yusa) with a 2 kb 5’ arm and a 5 kb 3’ rm. Positive clones were validated by PCR and DNA sequencing, and a correctly modified BAC was electroporated into mouse embryonic stem

cells (129SV background) at Columbia University (Herbert Irving Comprehensive Cancer Center Transgenic Shared Resource). Potentially recombined clones were screened by PCR and two positive clones were used to generate chimeric mice that resulted in germline transmission. Chimeras were bred with FLPe transgenic mice (Jackson Laboratories) to excise the neomycin resistance cassette. *Paupar* +/− mice were backcrossed 10 generations into the C57BL/6J (Jackson Laboratories) background. Glucagon-Venus mice, in which cells that express proglucagon are labeled by YFP Venus, have been described previously (Reimann et al., 2008).

All mice were maintained on a mixed C57BL6/129SV genetic background. Mice were group-housed in a 12-hour light/dark cycle (light between 07:00 and 19:00) at 22°C with free access to water and food (LabDiet 5053) and maintained according to protocols (AAAQ3403) approved by the Columbia University Institutional Animal Care and Use Committee. Cages and beddings were changed once per week. Mice were monitored regularly for their health status, and there were no viral and parasitic infections during this study. Euthanasia was performed by CO2 inhalation. Both male and female mice were used for experiments except when noted. Mouse aged ranged from embryonic day 15.5 to 12 months. Specific ages are indicated in the figure legends or methods. Genotyping for the *Paupar* WT or KO allele was performed using primers P1, P2, and P3 (Figure S4A). These primer sequences, as well as the genotyping primers for Glucagon-Venus mice, are listed in Table S5.

Mouse islets from both male and female mice were isolated by perfusion of the pancreas with Collagenase P (Roche) through the common hepatic bile duct at a concentration of 1 mg/ml in M199 medium (Invitrogen). The pancreas was dissected out and dissociated at 37°C for 16 minutes. After several washes in M199 supplemented with 10% FBS (Gemini Bio Products), islets were then filtered (500 μm; Fisher NC0822591), hand-picked to avoid exocrine contamination, and processed for downstream applications.

Human tissue—Human islets from both male and female donors (3 non-diabetic and 3 with T2D) were obtained through the NIH-supported Integrated Islet Distribution Program via the University of Wisconsin, the University of Alberta, and the Southern California Islet Cell Resources Center. The islets were harvested from deceased donors without any identifying information with informed consent and IRB approval at the islet isolation centers. Donors ranged in age from 27 to 71 years old (mean 54.3). Mean BMI was 34.76 (range 27.3–40.1). Islet purity ranged from 55-90% (mean 78%). Upon arrival, islets were cultured in media (CMRL with 5.5 mM glucose, 10% FBS, 1% Glutamax, 1% Pen/Strep) at 37°C in 5% CO2 and 95 % air. For Figure S2J, islets were cultured in either 2.8 mM or 16.7 mM glucose for 5 days prior to RNA extraction.

METHOD DETAILS

LncRNA Identification—To identify novel lncRNAs, we implemented an established pipeline for identifying lncRNAs from RNA-Seq data (Arnes et al., 2018; Pefanis et al., 2015). Briefly, rRNA-depleted total RNA from n=3 e15.5 embryonic mouse pancreata and 12-week-old mouse islets was prepared using the Ribo-Zero rRNA removal kit (Epicentre).

Biological replicates indicate that each RNA-seq data set was generated from individual mice (islets) or 3-4 pooled e15.5 embryonic pancreata. Libraries were prepared with Illumina TruSeq RNA sample preparation kit and then sequenced with 60 million, 2 × 100 paired reads on an Illumina HiSeq 2000 V3 instrument at the Columbia Genome Center. To reconstruct the transcriptomes, we first mapped all reads of total RNAs to the mouse reference genome (mm9) with TopHat v1.3.2 (Trapnell et al., 2012). Cufflinks v1.2.1 (Trapnell et al., 2012) was subsequently applied to assemble the whole transcriptome and to identify all possible transcripts. Then, the six transcriptomes from all samples was merged into a single gene transfer format (GTF) file with quantified gene RPKM values for each biological replicate. LncRNAs were then identified from the GTF file by removing transcripts if any of the following criteria were met: (1) they were overlapped with genes annotated in Ensembl and not annotated as 'lincRNA', 'non_coding', 'anti-sense', '3prime_overlapping_ncrna', 'processed_transcript', 'miRNA', 'misc_RNA', or 'pseudogene'; (2) overlapped with RefSeq genes (mm9) annotated as protein coding, where the RefSeq ID began with 'NM'; (3) overlapped with pseudogenes from Pseudogene.org; (4) less than 200 nt in length; (5) predicted to have protein coding potential by the Coding Potential Assessment Tool (CPAT; coding probability >0.364); (6) gene-level maximum FPKM was less than 0.5. The 572 remaining transcripts are high-confidence pancreatic lncRNAs. See Table S1 for the following information on the 572 pancreatic lncRNAs: transcript ID, gene ID, genomic location (mm9), fold change (islets/EP), transcript length, FPKM values in islets and EP. The last column of Table S1 'Detectable reads in e15.5 Ngn3+ cells' contains manual annotation of expression of each lncRNA an RNAseq dataset from e15.5 Ngn3+ cells (GSE80444; Churchill et al., 2017). This annotation provides readers with additional confirmation that expression of a candidate lncRNA in e15.5 whole pancreas samples corresponds to Ngn3+ endocrine progenitor cells.

RNA extraction and quantitative RT-PCR analysis—Total RNA was extracted from tissue or cells using the RNeasy Plus Micro Kit (Qiagen) or the RNeasy Plus Mini Kit (Qiagen), depending on the sample. Purified RNA was quantified by Nanodrop (Thermo Fisher), which also determined RNA quality. 250 ng of total RNA was used to synthesize cDNA with the SuperScript III First-Strand Synthesis System (Invitrogen) and random hexamer primers. Resultant cDNA was diluted in water and 25 ng was used in each qRT-PCR reaction. Reactions were run on a Bio-Rad CFX96 Real Time System using either gene specific primers with iQ Sybr Green Supermix (Biorad) or TaqMan probes (Applied Biosystems). Expression levels were normalized to TATA-binding protein (*TBP*) and quantified using the ΔΔCT method. Sybr Green primers and AODs used are listed in Table S5.

Cytoplasmic and nuclear RNA fractionation—αTC cells were grown to confluency, detached by trypsinization, and pelleted. Half of the pellet was used for total RNA isolation, and the other half was used for nuclear and cytoplasmic isolation using the PARIS kit (Ambion) following the manufacturer's instructions. Assessment of *Paupar* expression in each compartment was done using qRT-PCR with *Gapdh* and *Neat1* used for cytoplasmic and nuclear controls, respectively.

Single-molecule FISH (smFISH)—Oligonucleotide FISH probes were designed to target the full length of *Paupar* transcript. All probes were designed as 20-nt long, with CG content of 40-60%, no self-repeats and inner loop structures. Probes are labeled with Alexa647 at the 3' end and purchased from Integrated DNA Technologies. The sequence information of all FISH probes is provided in Table S5. The hybridization experiments were performed as previously described protocols (Raj et al., 2008; Lubeck and Cai, 2012; Cui et al., 2018). Cells were seeded onto collagen-coated, glass-bottom petri dishes (#1.5 thickness, MatTek) and grown to 70% confluence. Then cells were fixed with fresh 4% paraformaldehyde and quenched with 0.1% sodium borohydride. Fixed cells were permeabilized and stored with 70% ethanol until final use. FISH probes were diluted to 10 nM in hybridization solution (10% dextran sulfate, 10% formamide, 2X SSC, 0.02% RNase-free BSA, 2 mM ribonucleotide vanadyl complex). Hybridization was performed at 37°C overnight in a humid chamber. For cells co-stained with immunofluorescence, primary antibody was 1:1000 diluted and added to the hybridization solution. The next day, cells were thoroughly rinsed with 10% formamide in 2X SSC, followed by staining with secondary antibody and DAPI. Detailed antibody information is listed in the Key Resources Table. Images were taken from an Olympus IX71-based single-molecule microscope equipped with 405 nm, 488 nm, 542 nm, 594 nm and 640 nm solid lasers. A 100X oil immersion objective lens (NA 1.4) was used and images were captured with an EMCCD camera (Andor iXon Ultra 897). A z-stack scanning was performed to cover the whole cell volume (200 nm scanning step, 30-40 frames). In processing, RNA transcripts were identified at each scanning plane with Gaussian musk fitting algorithm and projected to the final reconstructed images. The processed images were subjected to counting with home-built MATLAB programs. Cell boundaries were determined and segmented based on the auto fluorescence background, in conjunction with DAPI staining. The source code for smFISH analysis is available upon request.

Immunohistochemistry—Tissues were fixed in 4% PFA in PBS overnight, washed in cold PBS, incubated in 30% sucrose, and cryopreserved. Immunofluorescence was performed on 7 mM sections. See Key Resources Table for details on primary and secondary antibodies used. DAPI was applied for 15 min following a 2-hour secondary antibody incubation.

Morphometric analysis—For all morphometric analysis, the entire pancreas was sectioned, and at least six evenly distributed sections were analyzed. To determine percentage hormone positive area over islet area, sections were stained with insulin, glucagon, and DAPI and five islets per section were imaged at 20X on a Zeiss Confocal LSM 710 microscope. Hormone positive area was measured using FIJI with a standardized signal threshold. Islet area was measured manually using FIJI based on morphology and guided by staining. Islets were arranged by size, and the total islet number was used to calculate the percentage represented in each size group. To determine the percentage of proliferating α cells, sections were stained with glucagon, Ki67, and DAPI, and five islets per section were imaged for quantification. Data presented is from n=3 mice per genotype unless otherwise indicated. Scorer was blinded for genotypes during quantification.

Flow Cytometry—Isolated islets from Glucagon-Venus mice were digested in 0.25% trypsin for 7 minutes at 37°C. The enzyme was deactivated by addition of M199 media supplemented with 10% FBS. The resulting single cell suspension was washed in cold 1X dPBS and filtered through a 35 μm filter prior to FACS. DAPI was added to the cell suspension as a viability marker. A BD Aria II Cell Sorter was used to separate populations of Venus-positive or -negative cells. Cells were sorted into Trizol LS and RNA was extracted according to the manufacturer's protocol.

Assessment of islet function

Glucose tolerance test: Mice were fasted overnight (~16 hours), followed by an IP injection of 2 mg D-glucose per gram mouse weight. Tail vein blood samples were collected at 0, 15, 30, 60, 90, and 120 minutes after the injection and glucose concentration was determined using the Accu-Chek Compact Plus Blood Glucose Meter (Roche).

Insulin tolerance test: Mice were fasted for 5 hours starting at 9 am. Insulin (Humalog in PBS) was injected IP at 0.75 units per kg mouse weight. Tail vein blood samples were collected at 0, 15, 30, 60, 90, and 120 minutes after the injection and glucose concentration was determined using the Accu-Chek Compact Plus Blood Glucose Meter (Roche). To measure plasma glucagon during an insulin tolerance test, blood was collected into heparinized tubes immediately before, and 30 minutes after, an insulin injection. Plasma was separated from whole blood by centrifugation and glucagon concentration was measured by ELISA (Mercodia).

Islet glucagon secretion and content: After isolation, duplicate samples of 20 islets per mouse were cultured overnight at 37°C in resting medium (RPMI 1640 with 10% FBS, 1% P/S, and 5.6 mM D-glucose). The next day, islets were transferred into modified Krebs buffer with 0.1% BSA and 20 mM D-glucose for 30 minutes to equilibrate. Islets were then transferred to Krebs buffer with 20 mM glucose for 30 minutes, followed by 2 mM glucose for 30 minutes. Lastly, islets were transferred to either Krebs buffer with 20 mM glucose or 10 mM L-Arginine for 30 minutes. Supernatant was collected following each treatment. Islets were then sonicated in 50 μl of lysis buffer (TE with 0.1% SDS) and lysate was used to determine DNA concentration and measure glucagon content. Glucagon concentration in supernatant and islet lysate was measured by glucagon ELISA (Mercodia).

HOMA-IR determination: Cohorts of mice were fasted overnight for ~16 hours. After the fast, tail vein blood was collected and blood glucose concentration was determined using the Accu-Chek Compact Plus Blood Glucose Meter (Roche). To determine insulin plasma concentration, fasted blood was collected into heparinized tubes. Plasma was separated from whole blood by centrifugation and insulin concentration was measured by ELISA (Mercodia). HOMA-IR was determined by the formula: HOMA-IR = fasting plasma insulin (μU/L) * fasting blood glucose (mmol/L) / 22.5.

RNA sequencing (RNA-seq)—For RNA-seq experiments done on islets from *Paupar* WT and KO mice, as well as MIN6 cells following control or *Paupar* KD, total RNA was extracted and converted into cDNA libraries (TruSeq RNA sample preparation kit, Illumina)

according to manufacturer's instructions. Sequencing was performed to a depth of 30 million, single-end 100 nt reads in three biological replicates per condition. Reads were aligned to the mouse genome (mm10) using HISAT2 v2.1.0 (Kim et al., 2013). Aligned reads were assigned to genes using annotations from Ensembl (Mus_musculus.GRCm38.73.gtf) and HTseq-count v0.10.0 (Anders et al., 2015). Differential expression across cohorts was assessed using DESeq2 v1.18 (Love, Huber, and Anders, 2014). Inclusion of samples required RNA integrity (RIN) values > 8.0 as determined with Agilent Bioanalyzer 2100. Complete RNA-seq data are available through GEO accession numbers: GSE122033 (embryonic pancreas and islet lncRNAs), GSE121884 (*Paupar* WT and KO islets), and GSE132072 (MIN6 cells following control KD or *Paupar* KD).

RNA interference—The day before transfection, 500,000 cells were plated in each well of a 12-well tissue culture plate. Cells were transfected in triplicate with 40 nM *Paupar* or control antisense oligos (ASOs) (IDT) using Lipofectamine 2000, per the manufacturer's instructions (Thermo Fisher). Cells were harvested for RNA 48 hours post transfection. ASO sequences are listed in Table S5.

Capture Hybridization Analysis of RNA Targets (CHART) enrichment and analysis—CHART enrichment experiments were performed as previously described (Simon et al., 2011). Briefly, CHART extract was prepared from approximately 5×10^7 αTC cells per CHART reaction and hybridized with 120 pmol biotinylated capture oligonucleotides (COs) (Table S5) overnight with rotation at room temperature. Complexes were captured using MyOne Streptavidin C1 beads (Invitrogen). Bound material was stringently washed and eluted using RNase H (New England Biolabs) for 15 minutes at room temperature. For RNA analysis, CHART-enriched material was incubated with XLR buffer (final concentrations: 2 mg/mL Proteinase K (Ambion), 33.3 mM Tris pH 7.2, 0.33% SDS, 16.7 mM EDTA) at 55°C for 1 hour and 65°C for 1 hour to reverse cross-links. RNA was purified using Trizol LS according to the manufacturer's instructions. qPCR analysis utilized iTaq Universal SYBR Green One-Step Kit to quantify *Paupar* enrichment compared to control genes. For analysis of CHART enriched proteins, eluate was precipitated using trichloroacetic acid and resuspended in SDS (4.25%), Tris pH 8.8 (529 mM), EDTA (64 mM) and β-mercaptoethanol (1.37 M). Samples were incubated at 98°C for 30 minutes and 65°C for 2 hours to reverse cros s-links. Samples (including 1%, 2.5%, and 5% input) were resolved by SDS-PAGE and transferred to nitrocellulose membranes. Membranes were incubated with antisera to detect PAX6 (Biolegend) at 1 μg/mL and bands visualized with SuperSignal West Pico PLUS Chemiluminescent Substrate.

CHART-MS—Proteins in CHART eluate were precipitated overnight at −20 °C with 4 volumes of cold acetone. The precipitated sample was centrifuged for 20 min at 16,000 × g at 4 °C and the supernatant discarded. The remaining pellet was washed with 1 mL cold acetone, centrifuged for 5 min at 16,000 × g at 4 °C, and the supernatant discarded. The remaining pellet was dried under vacuum. The protein pellet was digested to peptides using the TFE digest protocol (Wang et al., 2005) and peptide samples were analyzed by mass spectrometry. For LC-MS/MS analysis, 5 μl of 0.1 μg/μl of peptides per sample were

analyzed by reverse phase separation (C18) using a Waters nanoEquityTM UPLC system interfaced with a QExactive Plus Orbitrap mass spectrometer (Fort et al., 2018). LC-MS/MS datasets were converted to peak lists (DTA files) using the DeconMSn software (Mayampurath et al., 2008) and searched with MS-GF+ software (Kim, Gupta, and Pevzner, 2008; Kim and Pevzner, 2014) against Uniprot/SwissProt *mus musculus* database. The identified spectra were filtered based on their MSGF+ QValue score, which represents the false discovery rate of peptide-spectrum-matches (PSMs). PSMs with QValue less than 0.01 were retained. Proteins were quantified using the spectral counting quantification approach (Wang et al., 2008). Downstream analyses identified *Paupar* binding proteins that were enriched > 2-fold over controls, excluding proteins found in > 75% of control experiments listed in the Contaminant Repository for Affinity Purification (CRAPome; Mellacheruvu et al., 2013). We used Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) (Szklarczyk et al., 2017) to identify known and predicted protein-protein interaction networks among our Paupar-interacting proteins.

ChIP-qPCR—The ChIP-IT High Sensitivity kit (Active Motif) was used according to the kit instructions. Briefly, 48 hours post ASO transfection (after ensuring efficient *Paupar* KD), αTC cells were crosslinked, lysed, and sonicated to obtain DNA fragments with an average length of 200–1200 bp. Supernatant containing DNA-protein complexes was used for immunoprecipitations reactions consisting of 30 μg chromatin and 4 μg rabbit anti-PAX6 antibody (Biolegend) or rabbit IgG control antibody (included in kit). Immunoprecipitated chromatin was collected using protein G agarose beads, then washed, eluted, and purified according to kit instructions. ChIP-qPCR were performed using the iQ Sybr Green Supermix (Biorad) and primers against negative control (MyoD1), *MafB, Glucagon* promoter regions. Data were normalized to the input signal and IgG values.

ChIP-sequencing (ChIP-seq)—ChIP-seq libraries were prepared according to the instructions of the KAPA Hyper Prep kit for Illumina. Sequencing was performed using the Illumina HiSeq 2500 system according to the manufacturer's instructions. Reads were mapped to the reference genome (mm10) using Bowtie2 (Langmead and Salzberg, 2012) with the '--sensitive-local' setting. Peaks were called using MACS2 v2.1.2 (Zhang et al., 2008) with reads extended by 200 bp while bypassing the standard model using a qvalue cutoff of 0.01. Differential binding between control KD and *Paupar* KD samples was assessed using DiffBind v2.12.0 (Ross-Innes et al., 2012) with default settings. Annotation and visualization of genes nearby PAX6 binding peaks was done using ChIPseeker v.1.20.0 (Yu, Wang, and He, 2015). Complete ChIP-seq data is available through GEO under the accession number GSE132069.

QUANTIFICATION AND STATISTICAL ANALYSIS

Graphs were generated in GraphPad Prism 7 and statistical analyses were performed using Graphpad Prism 7. Statistical parameters including the value of n, statistical test used, and significance (p-value) are reported in the figures and their legends. For studies involving mouse tissues, replicates refer to samples derived from different mice, unless otherwise specified. For studies involving cell culture, replicates refer to transfections performed on different days. No exclusion criteria were employed for either in vivo or in vitro

experiments. Unpaired two-tailed t tests were used to assess significance when comparing qRT-PCR expression values of 2-3 genes between two conditions. Linear regression analysis was used to determine if two variables were significantly correlated to each other. The Fisher's exact test was used to examine the significance of the association (contingency) between two types of classifications. For differential expression of global measurements (RNA-seq), the DESeq2 software (Love et al., 2014) generated adjusted p-values using the Benjamini-Hochberg procedure to correct for multiple-hypothesis testing.

Analysis of publically available RNAseq datasets from GEO was done using the following computational pipeline. Reads were aligned to the mouse genome (mm10) using HISAT2 v2.1.0 (Kim et al., 2013). Aligned reads were assigned to genes using annotations from Ensembl (Mus.musculus.GRCm38.73.gtf) and HTseq-count v0.10.0 (Anders et al., 2015). Differential expression across cohorts was assessed using DESeq2 v1.18 (Love, Huber, and Anders, 2014).

DATA AND CODE AVAILABILITY

The accession numbers for the sequencing data reported in this paper are: GSE122033 (embryonic pancreas and islet lncRNAs), GSE121884 (Paupar WT and KO islets), GSE132072 (MIN6 cells following control KD or *Paupar* KD), and GSE132069 (PAX6 ChIP-seq following control or *Paupar* KD in αTCs). All datasets can be accessed using the SuperSeries accession number GSE132318.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank Fiona M. Gribble and Frank Reimann (University of Cambridge) for the Glucagon-Venus mice. We thank members of the Sussel laboratory for helpful discussions and critical reading of the manuscript. We thank the Columbia University Genome Center for performing the RNA sequencing experiments and the Columbia University Flow Core for assistance with FACS experiments. We are grateful to Chyuan-Sheng (Victor) Lin and the Herbert Irving Comprehensive Cancer Center (HICCC) Transgenic Mouse Shared Resource (TMSR) core for assistance with the generation of *Paupar* KO mice. We thank Martin Machyna and Matthew Simon from Yale University for helpful discussions and technical assistance with the CHART protocol. Additional core facility support was provided by the Columbia Diabetes Endocrinology Research Center (DERC) (NIH P30 DK063608). A portion of the research was performed using the Environmental Molecular Sciences Laboratory (EMSL), a DOE User Facility sponsored by the Office of Biological and Environmental Research located at the Pacific Northwest National Laboratory (PNNL). Funding for the project was provided by the National Institute of Diabetes and Digestive and Kidney Diseases grants F31 DK 107028 (R.A.S.), R01 DK 111406 (L.S), and HIRN consortium grant UC4 DK108101 (L.S., G.O. and C.A.). Support was also provided from the Russell Berrie Foundation (L.A.) and Juvenile Diabetes Foundation APF-2014-197-A-N (L.A.).

REFERENCES

Ahmad Z, Rafeeq M, Collombat P, and Mansouri A (2015). Pax6 Inactivation in the Adult Pancreas Reveals Ghrelin as Endocrine Cell Maturation Marker. PLoS One 10, 1–29.

Akerman I, Tu Z, Beucher A, Rolando DMY, Sauty-Colace C, Benazra M, Nakic N, Yang J, Wang H, Pasquali L, et al. (2017). Human Pancreatic beta Cell lncRNAs Control Cell-Specific Regulatory Networks. Cell Metab 25, 400–411. [PubMed: 28041957]

Anders S, Pyl PT, and Huber W (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169. [PubMed: 25260700]

Anderson KR, Torres CA, Solomon K, Becker TC, Newgard CB, Wright CV, Hagman J, and Sussel L (2009). Cooperative transcriptional regulation of the essential pancreatic islet gene NeuroD1 (beta2) by Nkx2.2 and neurogenin 3. J Biol Chem 284, 31236–31248. [PubMed: 19759004]

Arnes L, Akerman I, Balderes DA, Ferrer J, and Sussel L (2016). βlinc1 encodes a long noncoding RNA that regulates islet β-cell formation and function. Genes & Development 30, 502–507. [PubMed: 26944677]

Arnes L, Liu Z, Wang J, Carlo Maurer H, Sagalovskiy I, Sanchez-Martin M, Bommakanti N, Garofalo DC, Balderes DA, Sussel L, et al. (2018). Comprehensive characterisation of compartment-specific long non-coding RNAs associated with pancreatic ductal adenocarcinoma. Gut 68, 499–511.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, et al. (2000). Gene ontology: tool for the unification of biology. Nat Genet 25, 25–29. [PubMed: 10802651]

Ashery-Padan R, Zhou X, Marquardt T, Herrera P, Toube L, Berry A, and Gruss P (2004). Conditional inactivation of Pax6 in the pancreas causes early onset of diabetes. Dev Biol 269, 479–488. [PubMed: 15110714]

Babicki S, Arndt D, Marcu A, Liang Y, Grant JR, Maciejewski A, and Wishart DS (2016). Heatmapper: web-enabled heat mapping for all. Nucleic Acids Res 44, W147–153. [PubMed: 27190236]

Batista PJ, and Chang HY (2013). Long noncoding RNAs: cellular address codes in development and disease. Cell 152, 1298–1307. [PubMed: 23498938]

Benner C, van der Meulen T, Caceres E, Tigyi K, Donaldson CJ, and Huising MO (2014). The transcriptional landscape of mouse beta cells compared to human beta cells reveals notable species differences in long non-coding RNA and protein-coding gene expression. BMC Genomics 15, 1–16. [PubMed: 24382143]

Blodgett DM, Nowosielska A, Afik S, Pechhold S, Cura AJ, Kennedy NJ, Kim S, Kucukural A, Davis RJ, Kent SC, et al. (2015). Novel Observations From Next-Generation RNA Sequencing of Highly Purified Human Adult and Fetal Islet Cell Subsets. Diabetes 64, 3172–3181. [PubMed: 25931473]

Bramswig NC, Everett LJ, Schug J, Dorrell C, Liu C, Luo Y, Streeter PR, Naji A, Grompe M, and Kaestner KH (2013). Epigenomic plasticity enables human pancreatic alpha to beta cell reprogramming. J Clin Invest 123, 1275–1284. [PubMed: 23434589]

Brissova M, Haliyur R, Saunders D, Shrestha S, Dai C, Blodgett DM, Bottino R, Campbell-Thompson M, Aramandla R, Poffenberger G, et al. (2018). Alpha Cell Function and Gene Expression Are Compromised in Type 1 Diabetes. Cell Rep 22, 2667–2676. [PubMed: 29514095]

Cabili MN, Dunagin MC, McClanahan PD, Biaesch A, Padovan-Merhar O, Regev A, Rinn JL, and Raj A (2015). Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. Genome Biol 16, 1–16. [PubMed: 25583448]

Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C, et al. (2005). The transcriptional landscape of the mammalian genome. Science 309, 1559–1563. [PubMed: 16141072]

Chauhan BK, Yang Y, Cveklova K, and Cvekl A (2004). Functional Properties of Natural Human PAX6 and PAX6(5a) Mutants. Invest Ophthalmol Vis Sci 45, 385–392. [PubMed: 14744876]

Chu C, Zhang QC, da Rocha ST, Flynn RA, Bharadwaj M, Calabrese JM, Magnuson T, Heard E, and Chang HY (2015). Systematic discovery of Xist RNA binding proteins. Cell 161, 404–416. [PubMed: 25843628]

Churchill AJ, Gutierrez GD, Singer RA, Lorberbaum DS, Fischer KA, and Sussel L (2017). Genetic evidence that Nkx2.2 acts primarily downstream of Neurog3 in pancreatic endocrine lineage development. Elife 6 1–18.

Conarello SL, Jiang G, Mu J, Li Z, Woods J, Zycband E, Ronan J, Liu F, Roy RS, Zhu L, et al. (2007). Glucagon receptor knockout mice are resistant to diet-induced obesity and streptozotocin-mediated beta cell loss and hyperglycaemia. Diabetologia 50, 142–150. [PubMed: 17131145]

Courtney M, Gjernes E, Druelle N, Ravaud C, Vieira A, Ben-Othman N, Pfeifer A, Avolio F, Leuckx G, Lacas-Gervais S, et al. (2013). The inactivation of Arx in pancreatic alpha-cells triggers their neogenesis and conversion into functional beta-like cells. PLoS Genet 9, 1–18.

Cui Y, Hu D, Markillie LM, Chrisler WB, Gaffrey MJ, Ansong C, Sussel L, and Orr G (2018). Fluctuation localization imaging-based fluorescence in situ hybridization (fliFISH) for accurate detection and counting of RNA copies in single cells. Nucleic Acids Res 46, 1–12. [PubMed: 29177436]

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res 22, 1775–1789. [PubMed: 22955988]

DiGruccio MR, Mawla AM, Donaldson CJ, Noguchi GM, Vaughan J, Cowing-Zitron C, van der Meulen T, and Huising MO (2016). Comprehensive alpha, beta and delta cell transcriptomes reveal that ghrelin selectively activates delta cells and promotes somatostatin release from pancreatic islets. Mol Metab 5, 449–458. [PubMed: 27408771]

Dorrell C, Schug J, Lin CF, Canaday PS, Fox AJ, Smirnova O, Bonnah R, Street PR, Stoeckert CJ, Kaestner KH, and Grompe M (2011). Transcriptomes of the major human pancreatic cell types 54, 2832–2844.

Ediger BN, Lim HW, Juliana C, Groff DN, Williams LT, Dominguez G, Liu JH, Taylor BL, Walp ER, Kameswaran V, et al. (2017). LIM domain-binding 1 maintains the terminally differentiated state of pancreatic beta cells. J Clin Invest 127, 215–229. [PubMed: 27941246]

ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74. [PubMed: 22955616]

Epstein JA, Glaser T, Cai J, Jepeal L, Walton DS, and Maas RL (1994). Two independent and interactive DNA-binding subdomains of the Pax6 paired domain are regulated by alternative splicing. Genes & Development 8, 2022–2034. [PubMed: 7958875]

Fadista J, Vikman P, Laakso EO, Mollet IG, Esguerra JL, Taneera J, Storm P, Osmark P, Ladenvall C, Prasad RB, et al. (2014). Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. Proc Natl Acad Sci USA 111, 13924–13929. [PubMed: 25201977]

Fort KL, van de Waterbeemd M, Boll D, Reinhardt-Szyba M, Belov ME, Sasaki E, Zschoche R, Hilvert D, Makarov AA, and Heck AJR (2017). Expanding the structural analysis capabilities on an Orbitrap-based mass spectrometer for large macromolecular complexes. Analyst 143, 100–105. [PubMed: 29138777]

Furuta M, Zhou A, Webb G, Carroll R, Ravazzola M, Orci L, and Steiner DF (2001). Severe defect in proglucagon processing in islet alpha-cells of prohormone convertase 2 null mice. J Biol Chem 276, 27197–27202. [PubMed: 11356850]

Gao T, McKenna B, Li C, Reichert M, Nguyen J, Singh T, Yang C, Pannikar A, Doliba N, Zhang T, et al. (2014). Pdx1 maintains beta cell identity and function by repressing an alpha cell program. Cell Metab 19, 259–271. [PubMed: 24506867]

Gelling RW, Du XQ, Dichmann DS, Romer J, Huang H, Cui L, Obici S, Tang B, Holst JJ, Fledelius C, et al. (2003). Lower blood glucose, hyperglucagonemia, and pancreatic alpha cell hyperplasia in glucagon receptor knockout mice. Proc Natl Acad Sci USA 100, 1438–1443. [PubMed: 12552113]

Gerich JE, Charles MA, and Grodsky GM (1974). Characterization of the effects of arginine and glucose on glucagon and insulin release from the perfused rat pancreas. J Clin Invest 54, 833–841. [PubMed: 4430717]

Gosmain Y, Avril I, Mamin A, and Philippe J (2007). Pax-6 and c-Maf functionally interact with the alpha-cell-specific DNA element G1 in vivo to promote glucagon gene expression. J Biol Chem 282, 35024–35034. [PubMed: 17901057]

Gosmain Y, Katz LS, Masson MH, Cheyssac C, Poisson C, and Philippe J (2012). Pax6 is crucial for beta-cell function, insulin biosynthesis, and glucose-induced insulin secretion. Mol Endocrinol 26, 696–709. [PubMed: 22403172]

Gosmain Y, Marthinet E, Cheyssac C, Guerardel A, Mamin A, Katz LS, Bouzakri K, and Philippe J (2010). Pax6 controls the expression of critical genes involved in pancreatic alpha cell differentiation and function. J Biol Chem 285, 33381–33393. [PubMed: 20592023]

Gutierrez GD, Bender AS, Cirulli V, Mastracci TL, Kelly SM, Tsirigos A, Kaestner KH, and Sussel L (2017). Pancreatic beta cell identity requires continual repression of non-beta cell programs. J Clin Invest 127, 244–259. [PubMed: 27941248]

Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L, Yang X, Amit I, Meissner A, Regev A, Rinn JL, Root DE, and Lander ES (2011). LincRNAs act in the circuitry controlling pluripotency and differentiation. Nature 477, 295–299. [PubMed: 21874018]

Hamaguchi K and Leiter E (1990). Comparison of Cytokine Effects on Mouse Pancreatic α-Cell and β-Cell Lines Viability, Secretory Function, and MHC Antigen Expression. Diabetes 39, 415–425. [PubMed: 2108069]

Han A, Stoilov P, Linares AJ, Zhou Y, Fu XD, Black DL (2014). De novo prediction of PTBP1 binding and splicing targets reveals unexpected features of its RNA recognition and function. PLoS_Comput Biol 10, 1–18.

Hancock AS, Du A, Liu J, Miller M, and May CL (2010). Glucagon deficiency reduces hepatic glucose production and improves glucose tolerance in adult mice. Mol Endocrinol 24, 1605–1614. [PubMed: 20592160]

Hang Y, and Stein R (2011). MafA and MafB activity in pancreatic beta cells. Trends Endocrinol Metab 22, 364–373. [PubMed: 21719305]

Hart AW, Mella S, Mendrychowski J, van Heyningen V, and Kleinjan DA (2013). The developmental regulator Pax6 is essential for maintenance of islet cell function in the adult mouse pancreas. PLoS One 8, e54173. [PubMed: 23326594]

Hayashi Y, Yamamoto M, Mizoguchi H, Watanabe C, Ito R, Yamamoto S, Sun XY, and Murata Y (2009). Mice deficient for glucagon gene-derived peptides display normoglycemia and hyperplasia of islet {alpha}-cells but not of intestinal L-cells. Mol Endocrinol 23, 1990–1999. [PubMed: 19819987]

Heller RS, Stoffers DA, Liu A, Schedl A, Crenshaw EB 3rd, Madsen OD, and Serup P (2004). The role of Brn4/Pou3f4 and Pax6 in forming the pancreatic glucagon cell identity. Dev Biol 268, 123–134. [PubMed: 15031110]

Jiang W, Yuting L, Liu R, Zhang K, Zhang Yi. (2015). The lncRNA DEANR1 Facilitates Human Endoderm Differentiation by Activating FOXA2 Expression. Cell Reports 11, 137–148. [PubMed: 25843708]

Kanda T, Sullivan KF, and Wahl GM (1998). Histone–GFP fusion protein enables sensitive analysis of chromosome dynamics in living mammalian cells. Current Biology 8, 377–385. [PubMed: 9545195]

Kang HS, Takeda Y, Jeon K, and Jetten AM (2016). The Spatiotemporal Pattern of Glis3 Expression Indicates a Regulatory Function in Bipotent and Endocrine Progenitors during Early Pancreatic Development and in Beta, PP and Ductal Cells. PLoS One 11, 1–10.

Karro JE, Yan Y, Zheng D, Zhang Z, Carriero N, Cayting P, Harrrison P, and Gerstein M (2007). Pseudogene.org: a comprehensive database and comparison platform for pseudogene annotation. Nucleic Acids Res 35, D55–60. [PubMed: 17099229]

Keller A, Peltzer J, Carpentier G, Horvath I, Olah J, Duchesnay A, Orosz F, and Ovadi J (2007). Interactions of enolase isoforms with tubulin and microtubules during myogenesis. Biochim Biophys Acta 1770, 919–926. [PubMed: 17368730]

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, and Salzberg SL (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology 14, 1–13.

Kim D, Langmead B, and Salzberg SL (2015). HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12, 357–360. [PubMed: 25751142]

Kim S, Gupta N, and Pevzner PA (2008). Spectral Probabilities and Generating Functions of Tandem Mass Spectra: A Strike against Decoy Databases. Journal of Proteome Research 7, 3354–3363. [PubMed: 18597511]

Kim S, and Pevzner PA (2014). MS-GF+ makes progress towards a universal database search tool for proteomics. Nat Commun 5, 1–14.

Kiselev Y, Eriksen TE, Forsdahl S, Nguyen LH, and Mikkola I (2012). 3T3 cell lines stably expressing Pax6 or Pax6(5a)--a new tool used for identification of common and isoform specific target genes. PLoS One 7, 1–16.

Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, and Gao G (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res 35, W345–349. [PubMed: 17631615]

Ku GM, Kim H, Vaughn IW, Hangauer MJ, Myung Oh C, German MS, and McManus MT (2012). Research resource: RNA-Seq reveals unique features of the pancreatic beta-cell transcriptome. Mol Endocrinol 26, 1783–1792. [PubMed: 22915829]

Kuhn RM, Haussler D, and Kent WJ (2013). The UCSC genome browser and associated tools. Brief Bioinform 14, 144–161. [PubMed: 22908213]

Langmead B, and Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357–359. [PubMed: 22388286]

Lawlor N, Youn A, Kursawe R, Ucar D, and Stitzel ML (2017). Alpha TC1 and Beta-TC-6 genomic profiling uncovers both shared and distinct transcriptional regulatory features with their primary islet counterparts. Sci Rep 7, 1–14. [PubMed: 28127051]

Li B, Bi CL, Lang N, Li YZ, Xu C, Zhang YQ, Zhai AX, and Cheng ZF (2014). RNA-seq methods for identifying differentially expressed gene in human pancreatic islet cells treated with pro-inflammatory cytokines. Mol Biol Rep 41, 1917–1925. [PubMed: 24619356]

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. [PubMed: 19505943]

Lin MF, Jungreis I, and Kellis M (2011). PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. Bioinformatics 27, i275–282. [PubMed: 21685081]

Lin S, Lin Y, Nery JR, Urich MA, Breschi A, Davis CA, Dobin A, Zaleski C, Beer MA, Chapman WC, et al. (2014). Comparison of the transcriptional landscapes between human and mouse tissues. Proceedings of the National Academy of Sciences 111, 17224–17229.

Love MI, Huber W, and Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 1–21.

Lubeck E, and Cai L (2012). Single-cell systems biology by super-resolution imaging and combinatorial labeling. Nat Methods 9, 743–748. [PubMed: 22660740]

Manuel MN, Mi D, Mason JO, and Price DJ (2015). Regulation of cerebral cortical neurogenesis by the Pax6 transcription factor. Front Cell Neurosci 9, 1–21. [PubMed: 25667569]

Matthews DR Hosker JP, Rudenski AS, Naylor BA, Treacher DF, and Turner RC(1985). Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia. 28, 412–419. [PubMed: 3899825]

Mayampurath AM, Jaitly N, Purvine SO, Monroe ME, Auberry KJ, Adkins JN, and Smith RD (2008). DeconMSn: a software tool for accurate parent ion monoisotopic mass determination for tandem mass spectra. Bioinformatics 24, 1021–1023. [PubMed: 18304935]

McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, and Bejerano G (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol 28, 495–501. [PubMed: 20436461]

Mellacheruvu D, Wright Z, Couzens AL, Lambert JP, St-Denis NA, Li T, Miteva YV, Hauri S, Sardiu ME, Low TY, et al. (2013). The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. Nat Methods 10, 730–736. [PubMed: 23921808]

Menendez-Gutierrez MP, Roszer T, Fuentes L, Nunez V, Escolano A, Redondo JM, De Clerck N, Metzger D, Valledor AF, and Ricote M (2015). Retinoid X receptors orchestrate osteoclast differentiation and postnatal bone remodeling. J Clin Invest 125, 809–823. [PubMed: 25574839]

Mercer TR, Dinger ME, and Mattick JS (2009). Long non-coding RNAs: insights into functions. Nature Reviews Genetics 10 156–59.

Mitchell RK, Nguyen-Tu MS, Chabosseau P, Callingham RM, Pullen TJ, Cheung R, Leclerc I, Hodson DJ, and Rutter GA (2017). The transcription factor Pax6 is required for pancreatic beta cell identity, glucose-regulated ATP synthesis, and Ca(2+) dynamics in adult mice. J Biol Chem 292, 8892–8906. [PubMed: 28377501]

Miyazaki JI, Araki K, Yamato E, Ikegami H, Asano T, Shibaski Y, Oka Y, and Yamamura KI (1990). Establishment of a pancreatic beta cell line that retains glucose-inducible insulin secretion: special reference to expression of glucose transporter isoforms. Endocrinology 127, 126–132. [PubMed: 2163307]

Moran I, Akerman I, van de Bunt M, Xie R, Benazra M, Nammo T, Arnes L, Nakic N, Garcia-Hurtado J, Rodriguez-Segui S, et al. (2012). Human beta cell transcriptome analysis uncovers lncRNAs that are tissue-specific, dynamically regulated, and abnormally expressed in type 2 diabetes. Cell Metab 16, 435–448. [PubMed: 23040067]

Motterle A, Gattesco S, Peyot ML, Esguerra JLS, Gomez-Ruiz A, Laybutt DR, Gilon P, Burdet F, Ibberson M, Eliasson L, et al. (2017). Identification of islet-enriched long non-coding RNAs contributing to beta-cell failure in type 2 diabetes. Mol Metab 6, 1407–1418. [PubMed: 29107288]

Mustafi D, Kevany BM, Bai X, Golczak M, Adams MD, Wynshaw-Boris A, Palczewski K (2016). Transcriptome analysis reveals rod/cone photoreceptor specific signatures across mammalian retinas. Hum Mol Genet 25, 4376–4388. [PubMed: 28172828]

Nam HS, and Benezra R (2009). High levels of Id1 expression define B1 type adult neural stem cells. Cell Stem Cell 5, 515–526. [PubMed: 19896442]

Naya FJ, Huang HP, Qiu Y, Mutoh H, DeMayo FJ, Leiter AB, and Tsai MJ (1997). Diabetes, defective pancreatic morphogenesis, and abnormal enteroendocrine differentiation in BETA2/NeuroD-deficient mice. Genes & Development 11, 2323–2334. [PubMed: 9308961]

Nica AC, Ongen H, Irminger JC, Bosco D, Berney T, Antonarakis SE, Halban PA, and Dermitzakis ET (2013). Cell-type, allelic, and genetic signatures in the human pancreatic beta cell transcriptome. Genome Res 23, 1554–1562. [PubMed: 23716500]

Nishimura W, Kondo T, Salameh T, El Khattabi I, Dodge R, Bonner-Weir S, and Sharma A (2006). A switch from MafB to MafA expression accompanies differentiation to pancreatic beta-cells. Dev Biol 293, 526–539. [PubMed: 16580660]

Offield MF, Jetton TL, Labosky PA, Ray M, Stein RW, Magnuson MA, Hogan BLM, and Wright CVE (1996). PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. Development 122, 983–995. [PubMed: 8631275]

Pan FC, and Wright C (2011). Pancreas organogenesis: from bud to plexus to gland. Dev Dyn 240, 530–565. [PubMed: 21337462]

Pavlaki I, Alammari F, Sun B, Clark N, Sirey T, Lee S, Woodcock DJ, Ponting CP, Szele FG, and Vance KW (2018). The long non-coding RNA Paupar promotes KAP1-dependent chromatin changes and regulates olfactory bulb neurogenesis. EMBO J 37, 1–16. [PubMed: 29212815]

Paz I, Kosti I, Ares M Jr., Cline M, and Mandel-Gutfreund Y (2014). RBPmap: a web server for mapping binding sites of RNA-binding proteins. Nucleic Acids Res 42, W361–367. [PubMed: 24829458]

Pefanis E, Wang J, Rothschild G, Lim J, Kazadi D, Sun J, Federation A, Chao J, Elliott O, Liu ZP, et al. (2015). RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. Cell 161, 774–789. [PubMed: 25957685]

Powers AC, Efrat S, Mojsov S, Spector D, Habener JF, and Hanahan D (1990). Proglucagon processing similar to normal islets in pancreatic alpha-like cell line derived from transgenic mouse tumor. Diabetes 39, 406–414. [PubMed: 2156740]

Quinlan AR, and Hall IM (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. [PubMed: 20110278]

Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, and Tyagi S (2008). Imaging individual mRNA molecules using multiple singly labeled probes. Nat Methods 5, 877–879. [PubMed: 18806792]

Reimann F, Habib AM, Tolhurst G, Parker HE, Rogers GJ, Gribble FM (2008). Glucose Sensing in L Cells: A Primary Cell Study. Cell Metab 8, 532–539. [PubMed: 19041768]

Romer AI, Singer RA, Egl i D., Sussel L Murine perinatal beta cell proliferation and the differentiation of human stem cell derived insulin expressing cells require NEUROD1. Diabetes. In press.

Romero-Barrios N, Legascue MF, Benhamed M, Ariel F, and Crespi M (2018). Splicing regulation by long noncoding RNAs. Nucleic Acids Res 46, 2169–2184. [PubMed: 29425321]

Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, Brown GD, Gojis O, Ellis IO, et al. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature 481, 389–393. [PubMed: 22217937]

Sander M, Neubuser A, Kalamaras J, Ee. HC, Martin GR, and German MS (1997).Genetic analysis reveals that PAX6 is required for normal transcription of pancreatic hormone genes and islet development. Genes & Development 11, 1662–1673. [PubMed: 9224716]

Sasamoto Y, Hayashi R, Park SJ, Saito-Adachi M, Suzuki Y, Kawasaki S, Quantock AJ, Nakai K, Tsujikawa M, and Nishida K (2016). PAX6 Isoforms, along with Reprogramming Factors, Differentially Regulate the Induction of Cornea-specific Genes. Sci Rep 6, 1–14. [PubMed: 28442746]

Schaffer AE, Taylor BL, Benthuysen JR, Liu J, Thorel F, Yuan W, Jiao Y, Kaestner KH, Herrera PL, Magnuson MA, et al. (2013). Nkx6.1 controls a gene regulatory network required for establishing and maintaining pancreatic Beta cell identity. PLoS Genet 9, 1–15.

Scoville DW, Cyphert HA, Liao L, Xu J, Reynolds A, Guo S, and Stein R (2015). MLL3 and MLL4 Methyltransferases Bind to the MAFA and MAFB Transcription Factors to Regulate Islet b-Cell Function. Diabetes 64, 3772–3783. [PubMed: 26180087]

Segerstolpe Å, Palasantza A, Eliasson P, Andersson EM, Andréasson AC, Sun X, Picelli S, Sabirsh A, Clausen M, et al. (2017). Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes. Cell Metab 24, 593–607.

Shaham O, Menuchin Y, Farhy C, and Ashery-Padan R (2012). Pax6: a multi-level regulator of ocular development. Prog Retin Eye Res 31, 351–376. [PubMed: 22561546]

Shen S, Park JW, Lu ZX, Lin L, Henry MD, Wu YN, Zhou Q, and Xing Y (2014). rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proc Natl Acad Sci U S A 111, E5593–5601. [PubMed: 25480548]

Shiota C, Prasadan K, Guo P, El-Gohary Y, Wiersch J, Xiao X, Esni F, and Gittes GK (2013). Alpha-Cells are dispensable in postnatal morphogenesis and maturation of mouse pancreatic islets. Am J Physiol Endocrinol Metab 305, E1030–1040. [PubMed: 23982158]

Simon MD, Wang CI, Kharchenko PV, West JA, Chapman BA, Alekseyenko AA, Borowsky ML, Kuroda MI, and Kingston RE (2011). The genomic binding sites of a noncoding RNA. Proc Natl Acad Sci U S A 108, 20497–20502. [PubMed: 22143764]

Singer RA, and Sussel L (2018). Islet Long Noncoding RNAs: A Playbook for Discovery and Characterization. Diabetes 67, 1461–1470. [PubMed: 29937433]

Solloway MJ, Madjidi A, Gu C, Eastham-Anderson J, Clarke HJ, Kljavin N, Zavala-Solorio J, Kates L, Friedman B, Brauer M, et al. (2015). Glucagon Couples Hepatic Amino Acid Catabolism to mTOR-Dependent Regulation of alpha-Cell Mass. Cell Rep 12, 495–510. [PubMed: 26166562]

St-Onge L, Sosa-Pineda B, Chowdhury K, Mansouri A, and Gruss P (1997). Pax6 is required for differentiation of glucagon-producing α-cells in mouse pancreas. Nature 387. 406–409.

Sussel L, Kalamaras J, Hartigan-O'Connor DJ, Meneses JJ, R.A. P, J.L.R. R, and German MS (1998). Mice lacking the homeodomain transcription factor Nkx2.2 have diabetes due to arrested differentiation of pancreatic β cells. Development 125, 2213–2221. [PubMed: 9584121]

Swisa A, Avrahami D, Eden N, Zhang J, Feleke E, Dahan T, Cohen-Tayar Y, Stolovich-Rain M, Kaestner KH, Glaser B, et al. (2017). PAX6 maintains beta cell identity by repressing genes of alternative islet cell types. J Clin Invest 127, 230–243. [PubMed: 27941241]

Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, et al. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res 45, D362–D368. [PubMed: 27924014]

Talchai C, Xuan S, Lin HV, Sussel L, and Accili D (2012). Pancreatic beta cell dedifferentiation as a mechanism of diabetic beta cell failure. Cell 150, 1223–1234. [PubMed: 22980982]

Taylor BL, Liu FF, and Sander M (2013). Nkx6.1 is essential for maintaining the functional state of pancreatic beta cells. Cell Rep 4, 1262–1275. [PubMed: 24035389]

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, and Pachter L (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7, 562–578. [PubMed: 22383036]

Unger RH, and Cherrington AD (2012). Glucagonocentric restructuring of diabetes: a pathophysiologic and therapeutic makeover. J Clin Invest 122, 4–12. [PubMed: 22214853]

Vance KW, Sansom SN, Lee S, Chalei V, Kong L, Cooper SE, Oliver PL, and Ponting CP (2014). The long non-coding RNA Paupar regulates the expression of both local and distal genes. EMBO J 33, 296–311. [PubMed: 24488179]

Wang H, Qian WJ, Mottaz HM, Clauss TRW, Anderson DJ, Moore RJ, Camp DG, Pallavicini M, Smith DJ, and Smith RD (2005). Development and Evaluation of a Micro- and Nano-Scale Proteomic Sample Preparation Method. Journal of Proteome Research 4, 2397–2403. [PubMed: 16335993]

Wang L, Park HJ, Dasari S, Wang S, Kocher JP, and Li W (2013). CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res 41, 1–7. [PubMed: 23143271]

Wang M, You J, Bemis KG, Tegeler TJ, and Brown DP (2008). Label-free mass spectrometry-based protein quantification technologies in proteomic analysis. Brief Funct Genomic Proteomic 7, 329–339. [PubMed: 18579615]

Warming S, Costantino N, Court DL, Jenkins NA, and Copeland NG (2005). Simple and highly efficient BAC recombineering using galK selection. Nucleic Acids Res 33, 1–12. [PubMed: 15640442]

West JA, Davis CP, Sunwoo H, Simon MD, Sadreyev RI, Wang PI, Tolstorukov MY, and Kingston RE (2014). The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. Mol Cell 55, 791–802. [PubMed: 25155612]

Wilcox CL, Terry NA, Walp ER, Lee RA, and May CL (2013). Pancreatic alpha-cell specific deletion of mouse Arx leads to alpha-cell identity loss. PLoS One 8, 1–10.

Yu G, Wang LG, and He QY (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics 31, 2382–2383. [PubMed: 25765347]

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, and Liu XS (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol 9, R137. [PubMed: 18798982]

HIGHLIGHTS

- The conserved lncRNA *Paupar* is enriched in postnatal pancreatic islet α cells
- *Paupar* interacts with SR proteins to promote the alternative splicing of *Pax6*
- Loss of *Paupar* alters PAX6 binding specificity to dysregulate essential α cell genes
- *Paupar* KO mice develop fewer islet α cells and have defective glucagon secretion

CONTEXT AND SIGNIFICANCE

Understanding and treating diabetes is limited by insufficient knowledge about the regulatory mechanisms that orchestrate pancreatic islet function. Long noncoding RNAs (lncRNAs) are a relatively unexplored class of molecules that have been implicated in many biological processes and are often misregulated in diseased states. In this study, researchers at the University of Colorado and their collaborators characterize a pancreatic lncRNA, *Paupar*, finding it to be enriched in alpha cells, the class of pancreatic cells that produce the blood-sugar regulating hormone glucagon. *Paupar* enables cell-specific functions of the transcription factor PAX6 by bestowing preference for a PAX6 variant that activates alpha cell genes. Accordingly, mice lacking *Paupar* have fewer alpha cells and secrete insufficient levels of glucagon, leading to impaired glucose homeostasis. These results identify the lncRNA *Paupar* as a potential therapeutic target for the treatment of hyperglycemia and diabetes.

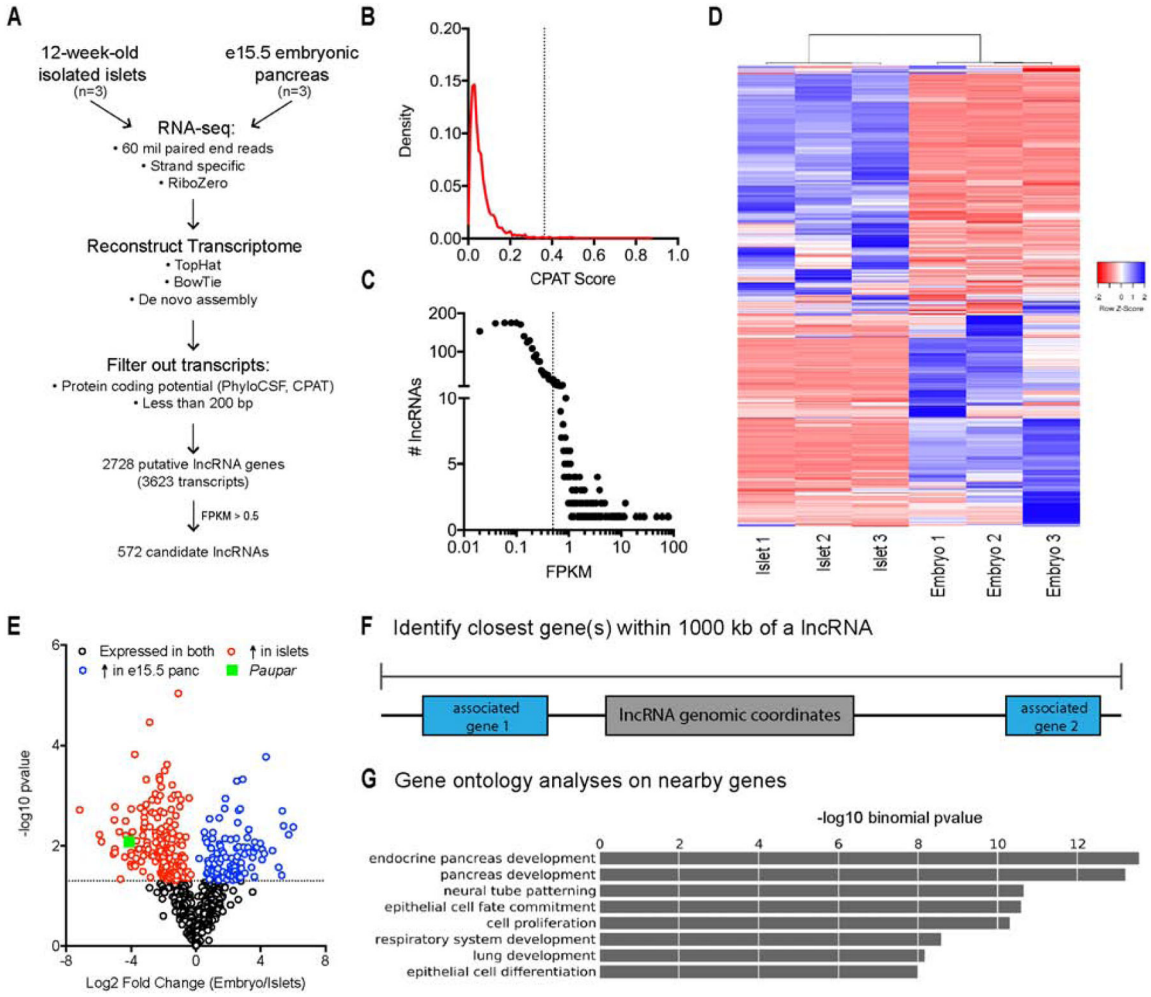


Figure 1. Systematic identification of developmentally regulated lncRNAs in the mouse pancreas.

(A) Overview of the lncRNA discovery pipeline from transcriptome analyses of e15.5 pancreas (EP) and adult isolated islets. n=3.

(B) Histogram of Coding Potential Assessment Tool (CPAT) scores of candidate lncRNAs. Dotted grey line indicates CPAT score < 0.364.

(C) Histogram of FPKM values of candidate lncRNAs. Dotted grey line indicates FPKM > 0.5.

(D) Heat map of 572 candidate lncRNAs with corresponding FPKMs, generated with Heatmapper. Z-scores calculated using Pearson distance measurements.

(E) Volcano plot of differentially expressed genes in EP compared to islets. Blue circles indicate lncRNAs enriched in EP; red circles indicate lncRNAs enriched in islets. Genes with p-value ≥ 0.05 (dotted grey line) are shown as black circles. Green square shows *Paupar* lncRNA.

(F) Overview of GREAT analysis.

(G) Gene ontology analysis on lncRNA-associated genes.

See also Figure S1 and Tables S1 and S2.

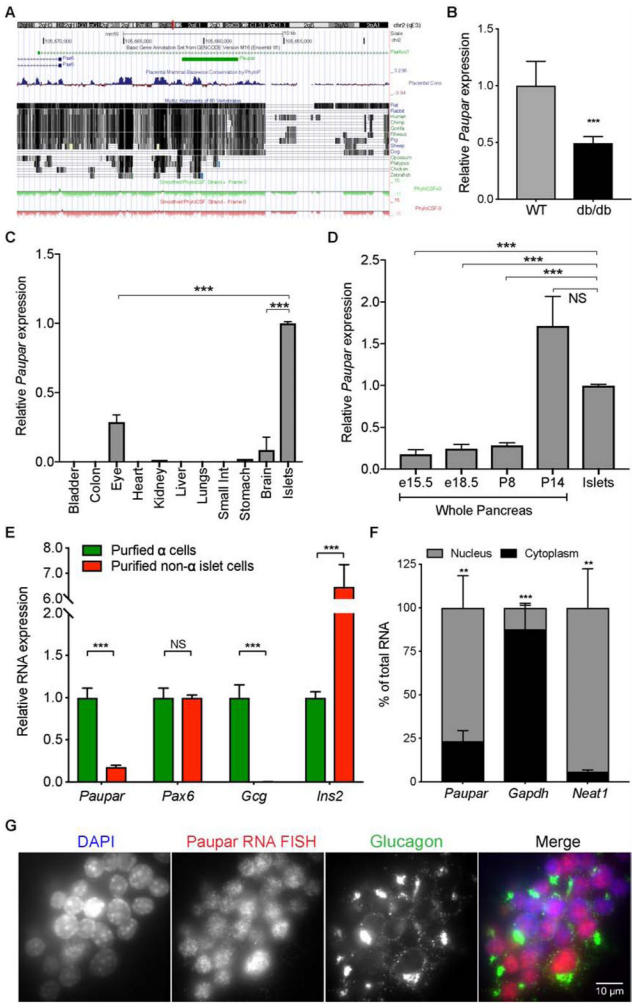


Figure 2. *Paupar* is a nuclear lncRNA enriched in pancreatic α cells.

(A) UCSC mm10 genome browser of the *Paupar* locus.

(B-D) *Paupar* expression by qRT-PCR in, n=3-5 biological replicates, (B) WT and *db/db* mice,

(C) 12-week-old WT mouse tissues, and (D) whole pancreas at indicated ages and 12-week-old isolated islets.

(E) Expression analyses by qRT-PCR of purified α cells (green bars) and non-α islet cells (red bars) cells. n=3 (10-15 pooled mice).

(F) *Paupar* expression by qRT-PCR in αTCs following cell fractionation. *Gapdh* and *Neat1* shown as nuclear (grey) and cytoplasmic (black) markers, respectively.

(G) Single molecule RNA fluorescent in situ hybridization (smFISH) images in αTCs showing *Paupar* (red) co-localized with antibody staining for Glucagon (green) and DAPI (blue). Scale bar indicates 10 μm. Images representative of 3 replicate experiments.

Quantification for (B) and others like it are presented as mean ± SEM, normalized to control values, t test, NS $p > 0.05$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and **** $p < 0.0001$.

See also Figure S2.

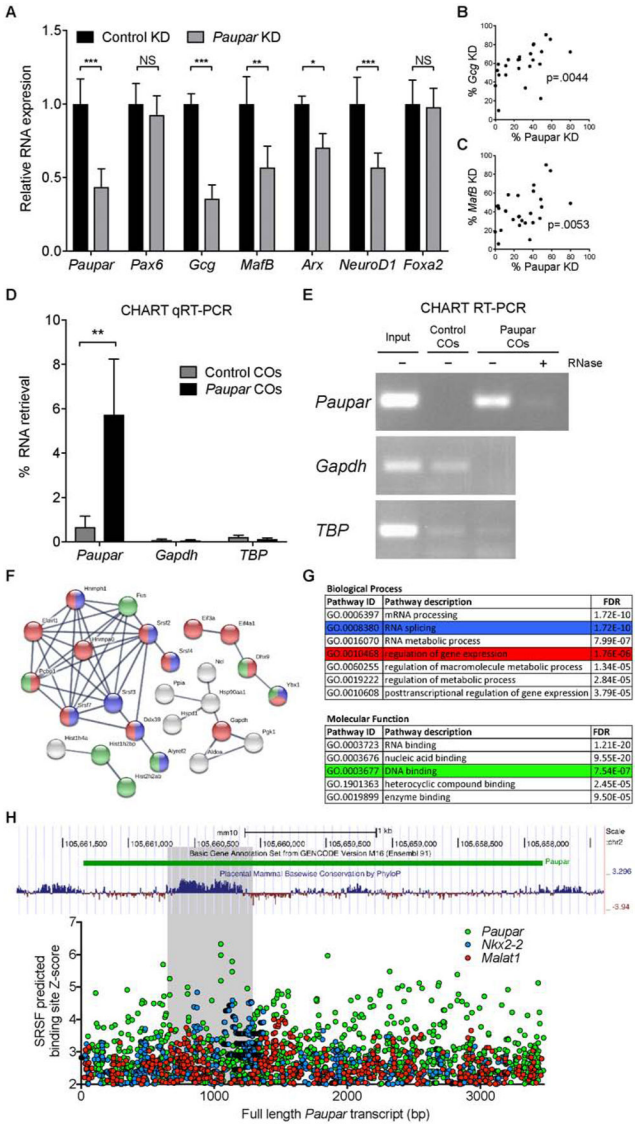


Figure 3. *Paupar* lncRNA regulates PAX6 α–cell target genes and interacts with several nuclear proteins involved in alternative splicing.

(A) Gene expression analysis by qRT-PCR from αTCs following control or *Paupar* knockdown (KD). n=4 for each condition.

(B, C) Correlation plots showing % *Paupar* KD compared to % KD of *Gcg* (B) and MafB (C) KD. P-values calculated using Pearson's correlation.

(D) CHART-qRTPCR analysis on eluate from experiments done with control COs or *Paupar* COs showing percent RNA retrieval relative to input for *Paupar, Gapdh*, and *TBP*.

(E) Gel images corresponding to (D) with an added lane showing *Paupar* COs plus RNase.

(F) Diagram showing output from STRING analysis of *Paupar* interacting proteins found by CHART-MS. Colors correspond to the categories highlighted in (G).

(G) GO analyses of *Paupar* interacting proteins from (F).

(H) Depiction of results from RBPmap showing z-scores for predicted SRSF binding sites mapped to *Paupar* (green dots), Nkx2-2 (blue dots), and Malat1 (red dots) loci. Graph is

aligned to UCSC screenshot showing the *Paupar* locus. Grey rectangle highlights the region of *Paupar* with high sequence conservation in placental mammals.
See also Figure S3 and Table S3.

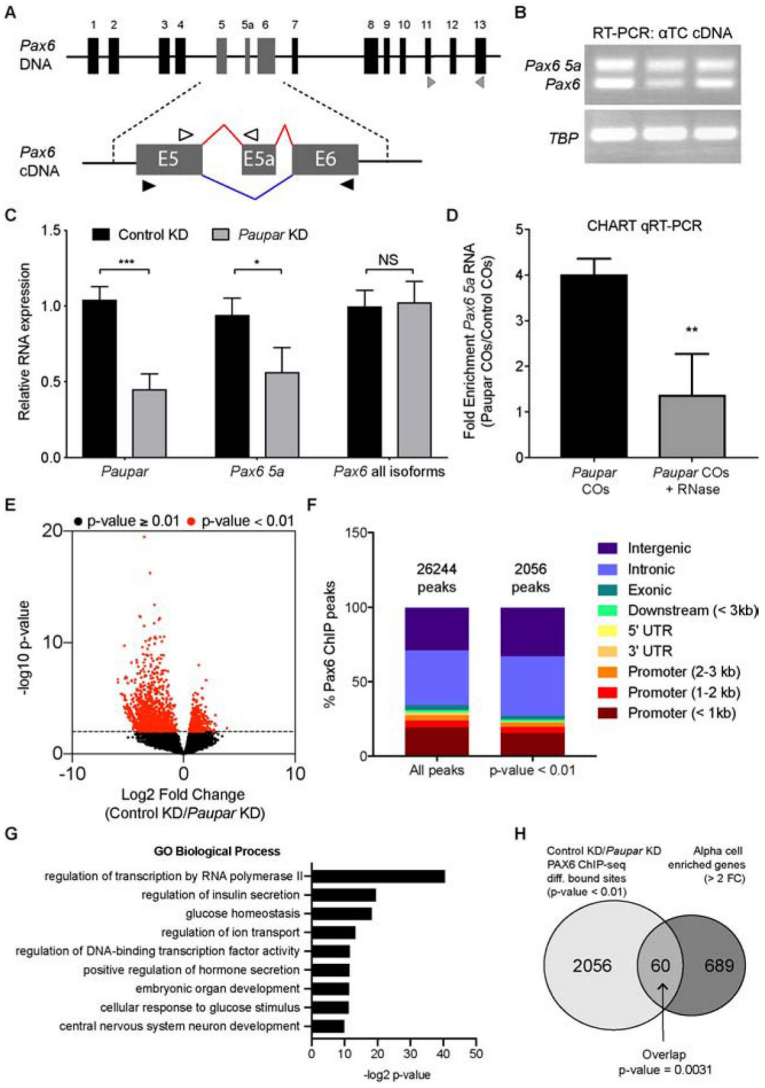


Figure 4. *Paupar* promotes the alternative splicing of *Pax6* to the isoform required for activation of PAX6 α-cell target genes.

(A) Schematic of the *Pax6* genomic locus and *Pax6* pre-mRNA with two isoforms, *Pax6*- red lines and *Pax6 5a*- blue lines. Black filled-in arrows are RT-PCR primers used in (B) to amplify both isoforms. White arrows are qRT-PCR primers used in (C) and (D) to specifically amplify *Pax6 5a*.

(B) RT-PCR analysis of αTC cDNA showing amplification of both *Pax6* isoforms.

(C) qRT-PCR analysis of *Paupar*, total *Pax6* RNA, and *Pax6 5a* from αTCs following control or *Paupar* KD.

(D) CHART qRT-PCR showing the enrichment of *Pax6 5a* in experiments using *Paupar* COs or *Paupar* COs with RNaseA treatment, compared to control COs.

(E) Volcano plot showing PAX6 αTC ChIP-seq peaks. Differentially bound peaks (control vs *Paupar* KD) with p-value < 0.01 are shown as red dots. Black dots correspond to peaks with p-value ≥ 0.01. n=2 biological replicates for each condition.

(F) Genomic distribution of all regions bound by PAX6 and regions differentially bound by PAX6 following *Paupar* KD.

(G) Gene ontology (GO) analyses of genes associated with differentially bound PAX6 peaks. (H) Venn diagram showing genes associated with differentially bound PAX6 peaks compared to genes enriched > 2-fold in α cells compared to other islet cell types (DiGruccio et al., 2016). P-value calculated using Fisher's exact test.

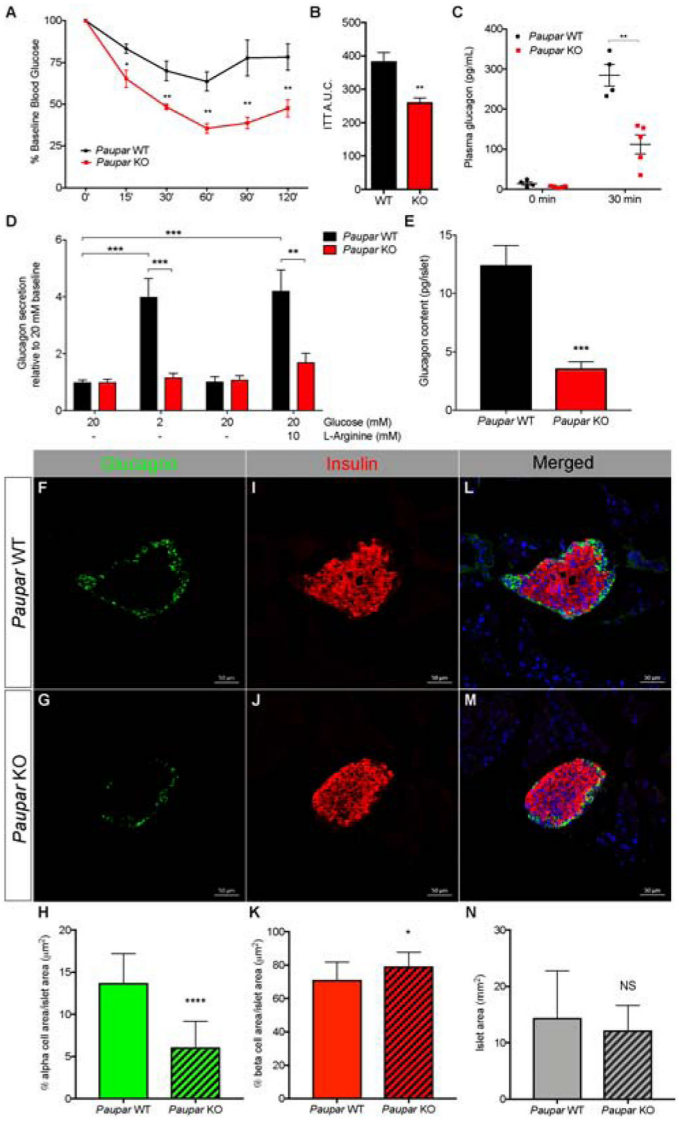


Figure 5. *Paupar* knockout mice have impaired α cell development and function.

(A) Graph of insulin tolerance tests *Paupar* WT (black) and *Paupar* KO (red) mice. n=7-8 for each genotype.

(B) Area under the curve (AUC) calculations for (A).

(C) Plasma glucagon levels (pg/mL) measured by ELISA during ITTs.

(D) The amount of glucagon secreted (pg/islet/hour) by isolated islets in response to indicated stimuli. Values are relative to the average secretion during initial 20 mM glucose incubation. n=4 mice per group, duplicate batches of 20 islets per mouse.

(E) Average glucagon content per islet (pg/islet) measured by ELISA. n=4.

(F-M) Immunofluorescence on *Paupar* WT (F, I, L) and *Paupar* KO (G, J, M) pancreata showing α cells (F, G), β cells, (I, J), and merged channels with DAPI (blue) (L, M). Scale bar indicates 50 μm. Images are representative of 3 replicate experiments

(H) Quantification of (F) and (G). Values shown are average from 15-20 islets per mouse, n=3 mice per genotype.

(K) Same as (H) but showing quantification of (I) and (J).

(N) Quantification of (L) and (M) showing average islet area (mm^2) in *Paupar* WT (plain grey bar) and KO mice (grey striped bar).

See also Figures S4-S7.

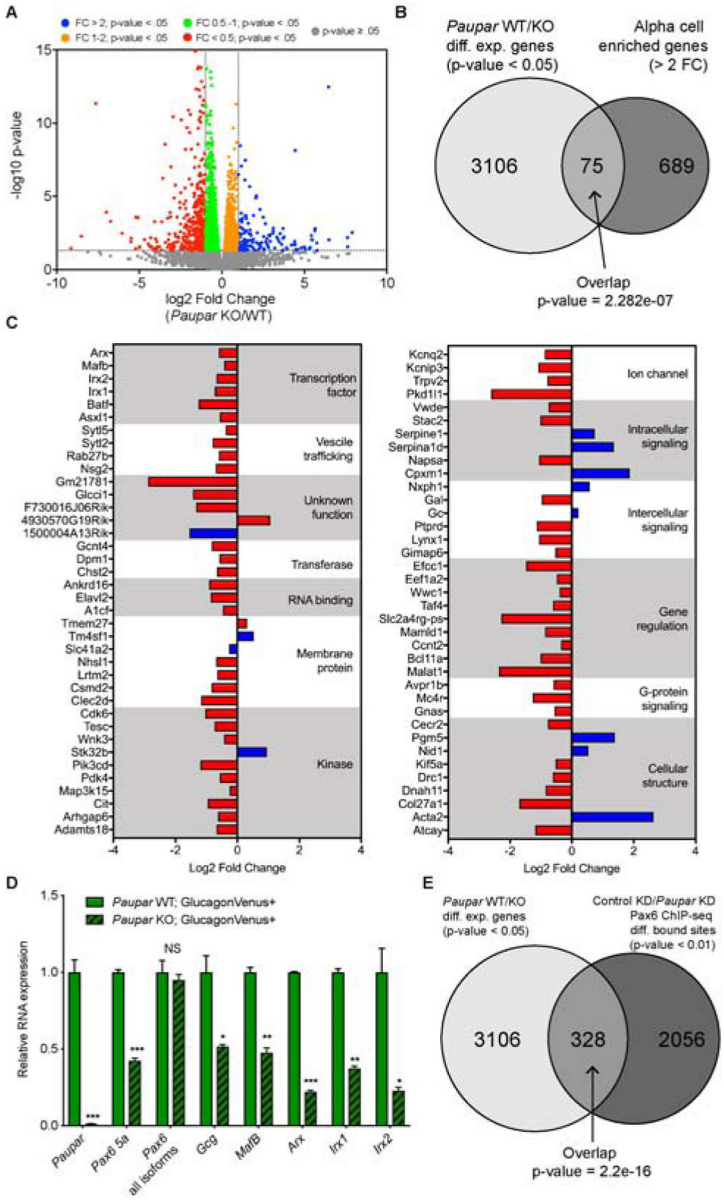


Figure 6. *Paurar* regulates essential α cell genes *in vivo*.

(A) Volcano plot showing differentially expressed genes (DEGs) in *Paurar* KO versus WT mice. Dot colors correspond to p-values and log2 fold changes as indicated in legend above graph.

(B) Venn diagram comparing DEGs from (A) to genes enriched > 2-fold in α cells compared to β cells (DiGruccio et al., 2016). P-value calculated by Fisher's exact test.

(C) Plots showing log2 fold change (*Paurar* KO/WT) values for 75 genes in middle of Venn diagram in (B) grouped by functional category.

(D) qRT-PCR analysis of FACS purified α cells from *Paurar* WT (green bars) and *Paurar* KO (green patterned bars). n=10-15 pooled mice.

(E) Venn diagram showing PAX6 peaks with altered binding following *Paurar* KD (Figure 4E, red dots) compared to DEGs shown in (A).

See also Table S4.

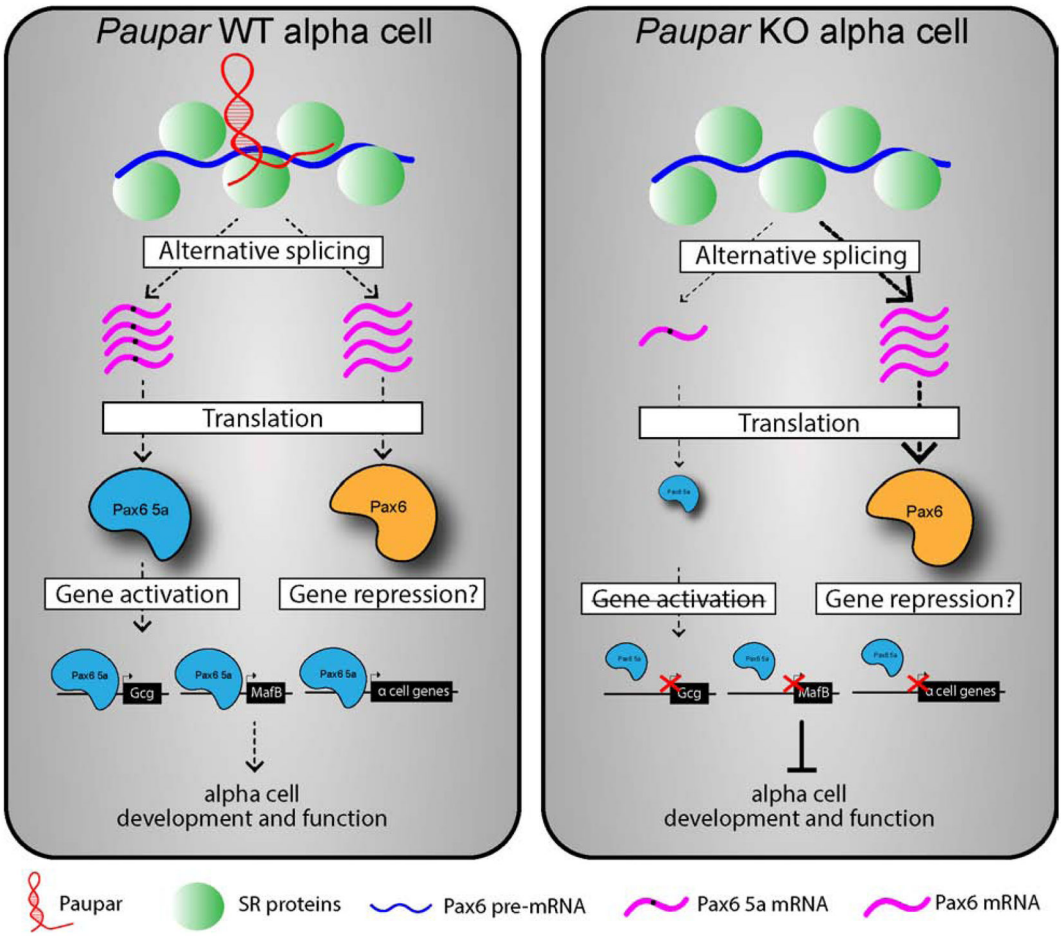


Figure 7. Model of *Paupar* mediated regulation of PAX6 α-cell target genes.

Paupar (red RNA molecule) is enriched in glucagon-producing α cells where it interacts with SR proteins (green circles) to promote the alternative splicing of Pax6 pre-mRNA (blue line) to the *Pax6 5a* (pink lines with black dot) isoform. We demonstrate that *Pax6 5a* is required for activation of the essential α cells genes, including *Gcg* and *MafB*. Deletion of *Paupar* in vivo resulted in dysregulation of PAX6 α-cell target genes and impaired α cell development and function.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
DAPI (4′,6-diamidino-2-phenylindole) (1:1000)	Thermo Fisher	Cat#D1306; RRID: AB_2629482
Rabbit monoclonal anti-Glucagon (1:500)	Santa Cruz	Cat#sc-514592; RRID: AB_2629431
Guinea pig anti-Insulin (1:1000)	DAKO	Cat#A0564; RRID: AB_2617169
Rat monoclonal anti-Somatostatin (1:500)	Abcam	Cat#AB30788; RRID:AB_778010
Rabbit polyclonal anti-Ki67 (1:500)	EMD Millipore	Cat#AB9260; RRID: AB_2142366
Alexa Fluor 488 anti-rabbit (1:500)	Jackson ImmunoResearch	Cat#711-545-152; RRID: AB_2313584
Cy3 anti-guinea pig (1:500)	Jackson ImmunoResearch	Cat#706-165-148; RRID: AB_2340460
Alexa Fluor 647 anti-rabbit (1:500)	Jackson ImmunoResearch	Cat#711-605-152; RRID: AB_2492288
Alexa Fluor 647 anti-rat (1:500)	Jackson ImmunoResarch	Cat#712-605-153; RRID: AB_2340694
Purified rabbit anti-Pax6 for ChIP	Biolegend	Cat#901302; RRID: AB_2749901
Rabbit IgG for ChIP	Millipore Sigma	Cat#I5006; RRID: AB_1163659
Bacterial and Virus Strains		
One Shot TOP10 Chemically Competent *E. coli*	Invitrogen	Cat#3879S
SW106 cells	Warming et al., 2005	N/A
Biological Samples		
Mouse embryonic stem cells (mESCs)	129SV background	N/A
Human cadaveric islets (non-diabetic and T2D donors)	Integrated Islet Distribution Program	https://iidp.coh.org
Chemicals, Peptides, and Recombinant Proteins		
Proteinase K	Promega	Cat#MC5005
Q5 HiFi DNA polymerase	NEB	Cat#M0491S
Gibson assembly kit	NEB	Cat#E2611S
Terrific Broth (TB)	Difco	Cat#243820
LB Agar	Difco	Cat#240110
RNase H	NEB	Cat#M0297S
DpnI	NEB	Cat#R0176S
EcoRI-HF	NEB	Cat#R3101S
HindIII-HF	NEB	Cat#R3104S
Go Taq DNA Polymerase	Promega	Cat#M3001
dNTPs nucleotide mix	Roche	Cat#11581295001
Sodium borohydride	Millipore Sigma	Cat#452882

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Formamide	Roche	Cat#11814320001
Sucrose	Millipore Sigma	Cat#S0389
DMEM media	Thermo Fisher	Cat#11995
Penicillin-Streptomycin	Thermo Fisher	Cat#15140163
HEPES buffer	Gibco	Cat#15630-080
Non-essential amino acids solutions	Sigma	Cat#M7145
Bovine Serum Albumin	Thermo Fisher	Cat#15260037
Sodium bicarbonate	Thermo Fisher	Cat#MT25035CI
Antibiotic-Antimycotic	Thermo Fisher	Cat#15240062
M199 media	Invitrogen	Cat#11150067
Fetal Bovine Serum	Gemini Bio Products	Cat#100106
Precision plus protein kaleidoscope protein standard	Bio-rad	Cat#1610375
Lipofectamine 2000 transfection Reagent	Thermo Fisher	Cat#11668-019
D-Glucose	Millipore Sigma	Cat#G8270
L-Arginine	Millipore Sigma	Cat#A5006
Insulin (NovoLog)	Novo Nordisk	U-100
iQ Sybr Green Supermix	Bio-rad	Cat#1708880
Real Time PCR Mastermix for Taqman	Eurogentec	Cat#RTQP2X0315+
iTaq Universal SYBR Green One-Step Kit	Bio-rad	Cat#1725150
Collagenase P	Millipore Sigma	Cat#11213857001
Donkey Serum	Sigma	Cat#D9663
Dynabeads MyOne Streptavidin C1	Invitrogen	Cat#65001
Paraformaldehyde EM Grade	Polysciences, Inc.	Cat#00380
Trizol LS	Invitrogen	Cat#10296010
Critical Commercial Assays		
RNeasy Plus Mini Kit	Qiagen	Cat#74134
RNeasy Plus Micro Kit	Qiagen	Cat#74034
TruSeq Stranded Total RNA (with Ribo-Zero)	Illumina	Cat#RS-122-2201
Ribo-Zero rRNA removal kit	Epicentre	Cat#MRZH11124
Pierce BCA protein assay kit	Thermo Fisher	Cat#23225
SuperScript III First-Strand Synthesis System	Invitrogen	Cat#18080051
Protein and RNA Isolation System (PARIS) Kit	Ambion	Cat#AM1921
Glucagon ELISA	Mercodia	Cat#10-1281-01
Insulin ELISA	Mercodia	Cat#10-1113-01
ChIP-IT High Sensitivity kit	Active Motif	Cat#53040
Deposited Data		
Raw and processed e15.5 embryonic mouse pancreas RNA-sequencing data	This study	GEO: GSE122033
Raw and processed 12-week-old adult mouse islets RNA-sequencing data	This study	GEO: GSE122033
Raw and processed 6-week-old *Paupar* WT and KO RNA-sequencing data	This study	GEO: GSE121884

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Raw and processed RNA-sequencing data from MIN6 cells following control KD or *Paupar* KD	This study	GEO: GSE132072
Raw and processed Pax6 ChIP-sequencing data from alpha-TC-1 cells following control KD or *Paupar* KD	This study	GEO: GSE132069
RNA-seq data from alpha-TC-1 and beta-TC-6 cells	Lawlor et al., 2017	GEO: GSE99954
RNA-seq data from whole mouse eyes	Mustafi et al., 2016	GEO: GSE38359
RNA-seq data from N2A neuroblastoma cells	Han et al., 2014	GEO: GSE45119
RNA-seq FACS purified mouse alpha and beta cells	DiGruccio et al., 2015	GEO: GSE80673
ENCODE RNA-sequencing data	ENCODE Project Consortium. 2012	GEO: PRJNA66167
RNA-seq FACS purified human alpha and beta cells	Blodgett et al., 2015	GEO: GSE67543
RNA-seq human fetal pancreas	Blodgett et al., 2015	GEO: GSE67543
RNA-seq human islets	Fadista et al., 2014	GEO: GSE50398
RNA-seq FACS purified e15.5 Ngn3+ pancreas cells	Churchill et al., 2017	GEO: GSE80444
Experimental Models: Cell Lines		
alpha TC1 clone 6 (αTC) cells	American Type Culture Collection	Cat#:CRL-2934
Mouse Insulinoma (MIN6) cells	Miyazaki et al., 1990	N/A
Experimental Models: Organisms/Strains		
Paupar$^{tm(H2BGFP)Suss}$ mice	This study	N/A
Glucagon-Venus mice	Reimann et al., 2008	N/A
C57BL/6J mice	Jackson Laboratories	Cat#000664
FLPe transgenic mice	Jackson Laboratories	Cat#003946
db/db mice	Jackson Laboratories	Cat#000697
Oligonucleotides		
Genotyping primers	This study	Table S5
smFISH probes	This study (IDT)	Table S5
Antisense Oligonucleotide (ASO) sequences	This study (IDT)	Table S5
qRT-PCR primers	This study	Table S5
CHART capture oligonucleotides (COs)	This study (IDT)	Table S5
MyoD1 negative control ChIP-qPCR primers	Keller et al 2007	Table S5
MafB ChIP-qPCR primers	Menéndez-Gutiérrez et al., 2015	Table S5
Gcg ChIP-qPCR primers	Schaffer et al., 2013	Table S5
Pax6 TaqMan AOD	Thermo Fisher	Mm00443081_m1
Glucagon TaqMan AOD	Thermo Fisher	Mm00801712_m1
MafB TaqMan AOD	Thermo Fisher	Mm00627481_s1
Arx TaqMan AOD	Thermo Fisher	Mm00545903_m1
NeuroD1 TaqMan AOD	Thermo Fisher	Mm01946604_s1
Foxa2 TaqMan AOD	Thermo Fisher	Mm01976556_s1
Irx1 TaqMan AOD	Thermo Fisher	Mm01352526_m1
Irx2 TaqMan AOD	Thermo Fisher	Mm01340315_m1
Recombinant DNA		

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Paupar BAC clone	BAC-PAC resources	RP23-465J7
H2B:GFP	Kanda et al., 1998	Addgene Cat#11680
pL451	Nam and Benezra, 2009	Addgene Cat#22687
pMCS-DTA	Generous gift from Kosuke Yusa, Osaka University, Japan	N/A
Software and Algorithms		
DESeq2	Love et al., 2014	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
GraphPad Prism 7	GraphPad Software	https://www.graphpad.com/scientific-software/prism
ImageJ	NIH	https://imagej.nih.gov/ij
Photoshop CC 2018	Adobe	N/A
Illustrator CC 2018	Adobe	N/A
R Software Package 3.3.1	The R Foundation	https://www.r-project.org
HISAT2 (v2.1.0)	Kim et al., 2013	https://ccb.jhu.edu/software/hisat2/index.shtml
Samtools (v1.4)	Li et al., 2009	http://samtools.sourceforge.net
HTSeq (v0.10.0)	Anders et al., 2015	https://htseq.readthedocs.io/en/master/install.html
Genomic Regions Enrichment of Annotations Tool (GREAT) (v3.0.0)	McLean et al., 2010	http://great.stanford.edu/public/html
TopHat2 (v2.1.1)	Kim et al., 2013	http://ccb.jhu.edu/software/tophat/index.shtml
Bowtie2 (v2.2.8)	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Bedtools (v2.17.0)	Quinlan and Hall, 2010	http://bedtools.readthedocs.io/en/latest
Heatmapper	Babicki et al., 2016	http://heatmapper.ca
Coding-Potential Assessment Tool (CPAT) (v1.2.4)	Wang et al., 2013	http://rna-cpat.sourceforge.net
PhyloCSF	Lin, Jungreis, and Kellis, 2011	https://github.com/mlin/PhyloCSF/wiki
UCSC Genome Browser	Kuhn, Haussler, and Kent, 2013	https://genome.ucsc.edu
Coding Potential Calculator (CPC)	Kong et al., 2007	http://cpc.cbi.pku.edu.cn
Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) (v10.0)	Szklarczyk et al., 2017	https://string-db.org

REAGENT or RESOURCE	SOURCE	IDENTIFIER
RBPmap (v1.1)	Paz et al., 2014	http://rbpmap.technion.ac.il
Replicate Multivariate Analysis of Transcript Splicing (rMATS) (v4.0.2)	Shen et al., 2014	http://rnaseq-mats.sourceforge.net
Contaminant Repository for Affinity Purification (CRAPome)	Mellacheruvu et al., 2013	https://www.crapome.org/
IGV	The Broad Institute	http://software.broadinstitute.org/software/igv
ApE	M. Wayne Davis	http://jorgensen.biology.utah.edu/wayned/ape
Gene Ontology	Ashburner et al., 2000	http://geneontology.org
MACS2 (v2.1.2)	Zhang et al., 2008	https://github.com/taoliu/MACS
DiffBind (v2.12.0)	Ross-Innes et al., 2012	https://bioconductor.org/packages/release/bioc/html/DiffBind.html
ChIPSeeker (v1.20.0)	Yu, Wang, and He, 2015	https://www.bioconductor.org/packages/release/bioc/html/ChIPseeker.html
Other		