

Loss-of-function of ALDH3B2 transdifferentiates human pancreatic duct cells into beta-like cells

Peng Yi

`peng.yi@joslin.harvard.edu`

Joslin Diabetes Center <https://orcid.org/0000-0001-7830-4694>

Jian Li

Joslin Diabetes Center

Yu-chi Lee

Joslin Diabetes Center

Noelle Morrow

UCLA <https://orcid.org/0000-0001-7015-4273>

Jennifer Hollister-Lock

Joslin Diabetes Center

Brooke Sullivan

Joslin Diabetes Center

Taylor Stewart

Joslin Diabetes Center, Harvard Medical School <https://orcid.org/0000-0002-0744-2309>

Siying Wei

Joslin Diabetes Center

Jessica Pereira

Joslin Diabetes Center <https://orcid.org/0000-0002-8503-1732>

Alexander Lee-Papastavros

Joslin Diabetes Center

Susan Bonner-Weir

Joslin Diabetes Center <https://orcid.org/0000-0003-4682-0656>

Kevin Bode

Joslin Diabetes Center

Andy Ma

Joslin Diabetes Center

Mulin He

Joslin Diabetes Center <https://orcid.org/0000-0002-9055-8109>

Hui Pan

Joslin Diabetes Center <https://orcid.org/0000-0001-5748-7217>

Jonathan Dreyfuss

Article

Keywords:

Posted Date: October 8th, 2024

DOI: <https://doi.org/10.21203/rs.3.rs-2222452/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Abstract

Replenishment of pancreatic beta cells is a key to the cure for diabetes. Beta cells regeneration is achieved predominantly by self-replication especially in rodents, but it was also shown that pancreatic duct cells can transdifferentiate into beta cells. How pancreatic duct cells were transdifferentiated and whether we could manipulate the transdifferentiation to replenish beta cell mass is not well understood. Using a genome-wide CRISPR screen, we discovered that loss-of-function of ALDH3B2 is sufficient to transdifferentiate human pancreatic duct cells into functional beta-like cells. The transdifferentiated cells have significant increase in beta cell marker genes expression, secrete insulin in response to glucose, and reduce blood glucose when transplanted into diabetic mice. Our study identifies a novel gene that we could potentially target in human pancreatic duct cells to replenish beta cell mass for diabetes therapy.

INTRODUCTION

Diabetes, no matter the cause or type, is a disease of pancreatic beta cell deficiency¹. Current treatments for diabetes do not provide the same degree of exquisite glycemic control as would a sufficient number of functional beta cells, and thus do not prevent the debilitating correlates of long-term diabetes. To cure diabetes, one has to find a way to stop the recurrent autoimmune attack on beta cells (type 1 diabetes) or resolve persistent peripheral insulin resistance (type 2 diabetes), but restoring a sufficient functional beta cell mass is critical to a cure for both types of diabetes. Beta cells can be replenished by transplanting human cadaveric islets or islet-like cells derived from human embryonic stem cells (hESC) or induced pluripotent stem cells (iPSCs)²⁻⁵. In most cases, transplanted islet cells are HLA-mismatched with the recipient, and immunosuppression is required to prevent graft rejection, creating other complications such as lack of immune defense against pathogens or tumor formation. Alternatively, it may be possible to promote endogenous pancreatic beta cell regeneration. Promoting the regeneration of a patient's own beta cells could be a safer strategy for beta cell mass replenishment; for type 1 diabetes autoimmunity would still need to be tamed but there would be no need for immunosuppression for allo-graft rejection. Beta cell regeneration (reviewed in⁶) can be achieved by self-duplication⁷ or transdifferentiation from other pancreatic cell types such as duct cells⁸, alpha cells^{9,10} or acinar cells¹¹. Pancreatic beta cell replication is the dominant mechanism of beta cell regeneration in adult rodents⁷. However in human, the adult beta cell replication rate is extremely low and it was postulated that human beta cells regeneration is achieved mainly by transdifferentiation from pancreatic duct cells⁸. It has been suggested that pancreatic duct cells may serve as a pool of progenitors for both the islet and acinar tissues after birth and into adulthood¹²⁻¹⁷. Although human duct-to-beta cell transdifferentiation has been evidenced by the existence of insulin-expressing cells in the pancreatic duct epithelium, it is a very rare phenomenon that many doubt would be relevant for sufficient beta cell regeneration. However, if its underlying mechanism can be understood, one could potentially manipulate duct-to-beta cell transdifferentiation to a high enough efficiency to replenish functional beta cell mass for the treatment of diabetes. Here we employed a genome-wide CRISPR

screen to dissect the mechanism of human duct-to-beta cell transdifferentiation and to identify new therapeutic targets for beta cell mass restoration.

RESULTS

A genome-wide CRISPR screen identifies ALDH3B2 as a regulator of human duct-to-beta cell transdifferentiation

Forward genetic screening, the genome-wide CRISPR screening in particular, is a powerful approach to discover novel genes, signaling pathways and the underlying mechanisms of a complex biological phenomenon. Here, we developed a genome-wide CRISPR screening strategy to search for genes that regulate the transdifferentiation of human pancreatic duct cells into beta cells. To ensure efficient gene editing and sufficient cell numbers for the genome-wide screen with sufficient coverage, we chose to use an immortalized pancreatic cell line, PANC-1, a human pancreatic carcinoma cell line of ductal origin that maintains many of the differentiated characteristics of normal mammalian pancreatic ductal epithelium^{18,19}. Our strategy is to use PANC-1 cells solely as a CRISPR screen and target discovery tool, and all the findings from the PANC-1 genetic screen will later be validated and characterized in primary human pancreatic duct cells. We reasoned that insulin expression or insulin promoter activation would be the most direct and the simplest readout for cell transdifferentiation into pancreatic beta-like cells.

Therefore, we introduced a reporter construct, Rat insulin promoter 3.1 (RIP)-EGFP-P2A- Blasticidin-S deaminase (BSD) (referred to as REPB reporter), into the PANC-1 cells via lentiviral transduction to create a REPB-PANC-1 reporter cell line (Fig. 1a). Rat insulin promoter (RIP) is known to be active in human beta cells and RIP 3.1 is a modified rat insulin promoter that is believed to have higher efficiency and beta cell specificity²⁰, which could increase the sensitivity of our CRISPR screen. The P2A peptide ensures the co-expression of the EGFP and BSD reporter genes, the EGFP reporter allows to visually and quantitatively monitor insulin promoter activation, and the expression of the BSD reporter gene confers resistance to blasticidin treatment, making it easy to enrich or select insulin promoter activated cells. To validate the REPB reporter construct, we also transduced the NIT-1 mouse beta cell line with the REPB reporter (REPB-NIT-1 cell line). As shown in Extended Data Fig. 1a and 1b, the EGFP expression can only be observed in the REPB-NIT-1 cells, but not at all in REPB-PANC-1 cells. The REPB-PANC-1 cells, but not the REPB-NIT-1 cells, were sensitive to blasticidin treatment (data not shown).

To execute the CRISPR screen, as illustrated in Fig. 1a, we transduced the REPB-PANC-1 cells with a human lentiviral genome-wide CRISPR knockout library (GeCKO v2)²¹ that comprises approximately 120,000 guide RNAs (gRNAs) targeting a total of 19,050 genes. We used a low multiplicity of infection (MOI) ~ 0.3 to ensure that most of the cells carry only one mutation. Briefly, approximately 10⁸ lentiviral library transduced REPB-PANC-1 cells were treated with low dose of blasticidin (10ug/ml) for 7 days, and then the blasticidin resistant cells were subjected to FACS sorting based on their EGFP intensity (Extended Data Fig. 2). Using next generation sequencing (NGS) and bioinformatic analysis, the gRNA

profile of the highest EGFP-expressing cells (EGFP^{high}, blasticidin-resistant) was generated and compared to that of the cells without blasticidin selection and FACS sorting (Extended Table 1).

We identified candidates whose sgRNAs were cooperatively and significantly positively selected in the EGFP^{high}, blasticidin-resistant REPB-PANC-1 cells vs control cells using MAGeCK^{22,23} and MAGeCKFlute²⁴ at the gene level. As shown in Fig. 1b from MAGeCKFlute where genes that were at least two-fold up are in red and those at least two-fold down are in blue, there is substantial positive selection. In Fig. 1c the half-volcano plot of statistics from MAGeCK-mle illustrates multiple highly significant positively selected hits with robust fold changes, such as ALDH3B2 whose fold change $> 10^5$ (Fig. 1c).

We next considered whether the screening hits were enriched for functional classifications involved in the cell fate change. Top pathways from pathway enrichment analysis of KEGG pathways using MAGeCKFlute are shown in Fig. 1d. Glycolysis/gluconeogenesis, inositol phosphate metabolism, beta-alanine metabolism, and starch and sucrose metabolism represent the highest NES (Normalized enrichment scores) value, which indicates that members of those gene sets tend to participate in the cell transdifferentiation. To understand the protein-protein associations including physical and functional interactions between INS, ACLY, and other GFP-positive screening hits from MAGeCKFlute analysis, the STRING pathway analysis was performed (Fig. 1e). Many CRISPR screening hits, such as ITPKA, GPI, HK2, ALDH3B2, and ALDH3A1 showing independent connection with INS and ACLY, which recognized as an essential gene during cell transdifferentiation²⁵ and metabolic reprogramming²⁶. The screening hits were significantly enriched for shared protein networks, providing additional confidence in the sensitivity to identify interrelated complexes (Fig. 1f). Many hits mapped to the functional categories of metabolic pathway and glycolysis/gluconeogenesis, all of them containing both previously unknown regulators of beta cell transdifferentiation and those with known roles in cell transdifferentiation or metabolic reprogramming.

Get together, we then picked the most enriched gRNA of top 8 candidate gene from the list and generated individual mutant PANC-1 cell lines using the corresponding gRNAs identified in our screen. We used quantitative PCR (qPCR) to analyze the expression of endocrine marker genes including insulin (INS), glucagon (GCG) and somatostatin (SST), as well as pancreatic duct cell marker gene keratin 19 (KRT19 or CK19). Several mutant PANC-1 cell lines showed differential expression of the examined marker genes (Fig. 1c, 1d and Extended Data Fig. 3a and 3b). In particular, PANC-1 cells transduced with a gRNA targeting ALDH3B2 showed the highest INS expression and the lowest KRT19 levels compared to non-targeting control (NTC) gRNA transduced PANC-1 cells. ALDH3B2, also known as ALDH8, is one of 19 members of the human aldehyde dehydrogenase (ALDH) superfamily that converts various types of aldehydes to carboxylic acids²⁷. It is well documented that ALDH genes are important regulators of stem cells and cell fate determination²⁸. Another close member in the ALDH family, ALDH1A3, was recently shown to contribute to pancreatic beta cell failure and de-differentiation in type 2 diabetes²⁹. We reasoned that ALDH3B2, as an enzyme, could potentially be an easier therapeutic target for small

molecules targeting. Therefore, we prioritized ALDH3B2 for further in-depth validation and characterization.

Loss-of-function of ALDH3B2 trans-differentiates PANC-1 cells into pancreatic beta-like cells

We generated an ALDH3B2 mutant PANC-1 cell line by lentiviral transduction of SpCas9 and ALDH3B2 gRNA into PANC-1 cells. Genomic sequencing of the targeted region in ALDH3B2^{mut} PANC-1 cells revealed that more than 75% of the sequenced ALDH3B2 alleles carried indel mutations (Extended Data Fig. 4). Western blot confirmed that ALDH3B2 protein level in the ALDH3B2^{mut} PANC-1 cells was reduced by ~ 40% compared to the non-targeting-control (NTC) gRNA lentivirus transduced PANC-1 cells (Extended Data Fig. 5a and 5b). To ensure that loss-of-function of ALDH3B2 induces *bona fide* cell transdifferentiation and did not only just activate the insulin promoter, we conducted a series of qPCR experiments to examine additional genes characteristic for pancreatic beta cells. We found that the expression of pancreatic endocrine hormones, insulin (INS) and somatostatin (SST), but not glucagon (GCG) or pancreatic polypeptide (PP), were significantly increased in the ALDH3B2 mutant PANC-1 cells (Fig. 2a). In addition, the expression of several of key beta cell transcription factors, including *PDX1*, *MAFA*, *NGN3* and *PAX6*, were also significantly upregulated in the ALDH3B2 mutant PANC-1 cells (Fig. 2b). We examined additional genes that are critical for beta cell function and found that *GLUT1* (*SLC2A1*), *GLUT2* (*SLC2A2*), *GLUCOKINASE* (*GCK*), subunits of the ATP-sensitive potassium (K-ATP) (*KCNJ11* and *ABCC8*), *CPE* and *IA2* were all significantly upregulated in the ALDH3B2 mutant PANC-1 cells (Fig. 2c). ALDH3B2 mutant PANC-1 cells had slightly decreased expression of pancreatic duct markers KRT19 and CA2 but not of HNF1B or Sox9 (Fig. 2d). Immunofluorescence imaging showed that clusters of ALDH3B2 mutant PANC-1 cells expressed human Insulin (INS) and C-peptide (CPEP), whereas neither insulin and C-peptide were detected in control PANC-1 cells (Fig. 2e). The insulin content of the ALDH3B2 mutant PANC-1 was significantly increased compared to control cells (Fig. 2f). In addition, using electron microscopy (EM), we found that many of the ALDH3B2 mutant PANC-1 cells had insulin granules (vesicles with halo, arrows in Fig. 2g), while no insulin granules were detectable in control -PANC-1 cells (Fig. 2g). Intriguingly, the ALDH3B2 mutant PANC-1 cells had intensive endoplasmic reticulum (ER) network (Fig. 2g), a characteristic found in pancreatic beta cells but not pancreatic duct cells. Collectively, these studies indicate that loss-of-function mutations in ALDH3B2 in PANC-1 cells not only trigger insulin promoter activation but also precipitate a significant cell fate transformation, shifting from a pancreatic ductal phenotype to a beta-like profile.

To evaluate whether these trans-differentiated pancreatic beta-like cells were functional, we performed *in vitro* glucose stimulated insulin secretion (GSIS) assay comparing control NTC-PANC-1 cells and transdifferentiated ALDH3B2 mutant PANC-1 cells, and we found that the ALDH3B2 mutant PANC-1 cells can secrete significantly more insulin at baseline (2.8 mM glucose) compared to the control NTC-PANC-1 cells, and also mildly respond to higher glucose (16.7 mM glucose) (Extended Data Fig. 6a). We then examined whether the ALDH3B2 mutant PANC-1 cells is also functional *in vivo* in diabetic mouse model.

We used streptozotocin (STZ) to induce diabetes in NSG mice, where majority of the beta cells were killed by STZ injection (Extended Data Fig. 7), and then transplanted the ALDH3B2 mutant or control PANC-1 cells (NTC) subcutaneously into these diabetic mice (Fig. 2h, upper panel). Mice transplanted with the ALDH3B2 mutant PANC-1 cells showed significantly decreased daily random blood glucose (Fig. 2h, lower panel) and improved glucose tolerance (Fig. 2i) compared to mice transplanted with NTC-PANC-1 cells. Notably, when the ALDH3B2 mutant PANC-1 graft was removed at the end of the study, blood glucose increased to the same level as in the control mice, confirming that the blood-glucose-lowering was indeed caused by the transplanted ALDH3B2 mutant PANC-1 cells (Fig. 2h, lower panel). Human insulin serum levels were also significantly higher in mice transplanted with ALDH3B2 mutant PANC-1 cells (Fig. 2j). Immunofluorescent imaging showed that transplanted ALDH3B2 mutant PANC-1 cells co-expressed PDX1, Insulin, NKX6.1 and C-peptide (Fig. 2k, 2l, Extended Data Fig. 8a and 8b). We observed that a few cells co-express somatostatin (SST) and Insulin, but no cell expresses glucagon (GCG), or the exocrine cell marker gene amylase (AMY) (Extended Data Fig. 8c and 8d). Of all the transplanted ALDH3B2 mutant PANC-1 cells, ~ 20% were PDX1⁺ (Fig. 2m) and ~ 8% were INS⁺ or NKX6.1⁺ (Fig. 2n and 2o). Almost all the INS⁺ cells were also NKX6.1⁺, suggesting that trans-differentiated beta-like cells adopted a true beta cell phenotype. It should also be noted that only ~ 45% of the PDX1⁺ cells co-expressed insulin (Fig. 2p), and we speculate that the PDX1⁺/INS⁻ cells may represent pancreatic progenitor-like cells that have yet committed to beta cell fate.

We employed an inducible shRNA system to ensure that the transdifferentiation of PANC-1 cells into beta-like cells by ALDH3B2 CRISPR knockout was indeed due to the loss-of-function of ALDH3B2 and not caused by off-target effects of the ALDH3B2 gRNA. We generated PANC-1 cell lines carrying a Tet-On inducible ALDH3B2 shRNA or a scrambled control shRNA (Fig. 3a, left panel). The ALDH3B2 shRNA PANC-1 cells with Doxycycline treatment showed significantly reduced ALDH3B2 mRNA expression (Fig. 3a right panel) and protein level (Extended Data Fig. 5c and 5d) after doxycycline (dox) treatment. Similar to the ALDH3B2 CRISPR mutant PANC-1 cells, knock-down of ALDH3B2 by shRNA also trans-differentiated PANC-1 cells into beta-like cells. A series of qPCR experiments showed that the expression of key beta cell transcription factors including *PDX1*, *MAFA*, *NGN3*, *NEUROD* and *PAX6* (Fig. 3b), endocrine hormone insulin (*INS*) and somatostatin (*SST*) (Fig. 3c), and beta cell function related genes (*SLC2A2*, *GCK*, *KCNJ11* and *ABCC8*) were significantly increased (Fig. 3d), whereas the expression of pancreatic duct cell marker genes (*KRT19*, *CA2* and *SOX9*) were reduced (Fig. 3e). Human insulin could also be detected in shALDH3B2 PANC-1 cells (+ Dox) but not in shControl PANC-1 cells or in shALDH3B2 PANC-1 cells (+ Dox) by immunofluorescence (Fig. 3f). These analyses confirmed that loss-of-function of ALDH3B2 by CRISPR targeting or shRNA silencing allowed PANC-1 cells to transdifferentiate and adopt a beta-like cell fate.

Loss-of-function of ALDH3B2 transdifferentiates human primary pancreatic duct cells into beta-like cells.

Next, we tested whether loss-of-function of ALDH3B2 was also able to transdifferentiate human primary pancreatic duct (HPPD) cells into beta-like cells. HPPD cells were isolated and affinity-purified from human donor islet-depleted pancreatic acinar tissue from Integrated Islet Distribution Program (IIDP) ³⁰.

qPCR analyses confirmed lack of insulin expression (Fig. 4a) and high expression of the pancreatic duct markers *KRT19* (Fig. 4b) in the purified HPPD cells compared to primary human islets. Interestingly, we observed that *ALDH3B2* expression levels are markedly lower in human islets compared to pancreatic duct cells (Fig. 4c). This differential expression pattern aligns with our results where the mutation of *ALDH3B2* in human pancreatic duct cells promotes their transdifferentiation into beta-like cells. These findings suggest that the reduction of *ALDH3B2* could be a critical step in the cellular reprogramming process leading to a beta-cell phenotype. Purified HPPD cells were transduced with lentiviruses carrying SpCas9 and either *ALDH3B2* gRNA (*ALDH3B2*^{mut}-HPPD) or a non-targeting control gRNA (NTC-HPPD). qPCR analysis demonstrated that *ALDH3B2*^{mut}-HPPD cells had significantly higher expression of key beta cell transcription factors (*PDX1* and *MAFA*) (Fig. 4d), endocrine hormone insulin (*INS*) and somatostatin (*SST*) (Fig. 4e) and beta cell function-related genes (*GCK*, *SLC2A1*, *SLC2A2*, *KCNJ11* and *CPE*) (Fig. 4f). The expression of several pancreatic duct cell marker genes was either unchanged (*KRT19* and *HNF1B*) or slightly reduced (*SOX9*) (Fig. 4g). Using immunofluorescent imaging, we found that a fraction of HPPD-*ALDH3B2*^{mut} cells expressed Insulin while still retaining CK19 expression, a possible signature of newly transdifferentiated beta cells from pancreatic duct cells, whereas no insulin expression could be detected in HPPD-NTC cells (Fig. 4h). Furthermore, we also examined whether the *ALDH3B2*^{mut}-HPPD cells have insulin granules using electron microscopy (EM). Although not as many insulin granules as in primary human beta cells, the *ALDH3B2*^{mut}-HPPD cells do have significant amount of mature insulin granules, while no insulin granules were detectable in the control NTC-HPPD cells (Fig. 4i).

We performed *in vitro* GSIS assay to evaluate the function of the *ALDH3B2*^{mut}-HPPD cells, and found that compared to control NTC-HPPD cells, the *ALDH3B2*^{mut}-HPPD cells can secrete significantly more insulin and mildly respond to high glucose (Extended Data Fig. 6b). Although the level of insulin secretion from the *ALDH3B2*^{mut}-HPPD cells is still much lower than human islets (Extended Data Fig. 6b), we suspected that it is due to relatively low efficacy of the transdifferentiation and immaturity of the transdifferentiated beta-like cells, especially in the *in vitro* experiment setting. We then transplanted *ALDH3B2*^{mut}-HPPD or NTC-HPPD cells under the kidney capsule of STZ induced diabetic NSG mice, and monitored their blood glucose over time (Fig. 5a, upper panel). Mice transplanted with HPPD-*ALDH3B2*^{mut} cells had significantly lower blood glucose compared to NTC-HPPD cells transplanted mice (Fig. 5a, lower panel). When the *ALDH3B2* mutant HPPD grafts were removed at 56 days post-transplantation, blood glucose increased to similar level as in the control HPPD transplanted mice, suggesting that the blood-glucose-lowering effect was indeed conferred by the transplanted *ALDH3B2* mutant HPPD cells (Fig. 5a, lower panel). Importantly, transplanted *ALDH3B2*^{mut}-HPPD cells secreted human insulin in response to glucose challenge. We detected a significantly increase in human serum insulin 5 minutes after glucose injection (both at 1 week and 3 weeks post-transplantation). No such response was observed in mice transplanted with NTC-HPPD cells or in non-transplanted control NSG mice (Fig. 5b and 5c). Immunofluorescent analysis revealed that Insulin⁺ cells, C-Peptide⁺ cells, PDX1⁺ cells, NKX6.1⁺ cells were only observed in mice transplanted with *ALDH3B2*^{mut}-HPPD cells but not with

NTC-HPPD cells (Fig. 5d, 5e, Extended Data Fig. 9a-d). Almost all of the transplanted ALDH3B2^{mut}-HPPD cells expressed pancreatic duct marker genes CK19 and SOX9 but not glucagon (GCG), Somatostatin (SST) or Amylase (AMY) (Extended Data Fig. 9e and 9f). Approximately 40% of the transplanted cells were successfully transduced with the NTC or ALDH3B2 gRNA lentivirus (shown by quantification of the percentage of Cas9 (Flag-tagged)⁺/CK19⁺ cells, Fig. 5d and 5f), and among all the gRNA lentivirus infected pancreatic duct cells, ~ 15% of ALDH3B2^{mut}-HPPD cells expressed insulin (Fig. 5g). Interestingly, we found that the majority of the pancreatic duct cells infected with ALDH3B2 gRNA lentivirus co-expressed PDX1 and CK19 (Fig. 5e and 5h), and approximately 12% of the PDX1⁺ cells co-expressed insulin (Fig. 5i). Co-expression of PDX1 and CK19 is a signature of pancreatic progenitor cells³¹, and we postulate that ALDH3B2 loss-of-function may cause the de-differentiation of mature duct cells into pancreatic progenitor-like cells, a portion of which then subsequently differentiate into beta-like cells.

Loss of ALDH3B2 function in pancreatic duct cells causes epigenetic changes

We found that ALDH3B2 loss-of-function allowed pancreatic duct cells to adopt a beta-like cell profile, and we next asked if transdifferentiation was associated with epigenetic changes. To this end, we analyzed DNA methylation in the human insulin gene region by bisulfite conversion assay. DNA methylation was significantly reduced in ALDH3B2 mutant PANC-1 cells compared to control NTC-PANC-1 cells at the + 63, +127 and + 139 positions of the human insulin locus, which are three well-characterized DNA methylation sites in the insulin locus³² (Fig. 6a and 6b). For DNA methylation analysis of primary human pancreatic duct cells, we included primary human islets for comparison. Again, ALDH3B2 mutation significantly reduced DNA methylation at the same three sites in the insulin gene locus (Fig. 6c and 6d). ALDH3B2 mutation did not reduce the DNA methylation to the level observed in primary islets. This might be due to the fact that only a fraction (8–15%) of pancreatic duct cells were transdifferentiated into beta-like cells with ALDH3B2 mutation. Overall, DNA methylation analyses suggest that loss-of-function of ALDH3B2 caused epigenetic changes in the pancreatic duct cells to induce a stable cell fate change into pancreatic beta-like cells.

ALDH3B2 loss-of-function in human pancreatic duct cells induces heterogeneous beta-like cell populations with overlapping endocrine and duct cell identity

To investigate the characteristics of transdifferentiated beta-like cells in more detail, we performed 3' gene expression single cell RNA sequencing of control or ALDH3B2 mutant HPPD cells and identified 13 unique cell cluster (Fig. 7a). Whereas the control NTC-HPPD condition only shows a few insulin positive cells in cluster 10, which could represent rare spontaneous transdifferentiated beta-like cells from duct cells, ALDH3B2 mutant HPPD cells develop insulin-expressing cells in various cell clusters and most prominent in cluster 5, 10, 11 and 12 (Fig. 7b-d). The relative proportion of beta-like-cell-containing cluster 5, 10, 11 and 12 is largely increased in the ALDH3B2 mutant HPPD cells compared to control

NTD-HPPD cells (Extended Data Fig. 10a). The percentage of insulin-expressing cells in ALDH3B2 mutant HPPD cells is about 18.1% but only 0.6% in NTC HPPD cells, and about 93% of all cells in both conditions still retain the expression of duct cell marker KRT19 (Extended Data Fig. 10b). The majority of insulin-positive beta-like cells also show significantly higher expression of other beta cell marker genes such as CHGA and TTR but also duct cell identity marker genes KRT17, 19, and 23 (Fig. 7e). Differential gene expression analysis of insulin high-expressing cells (relative intensity > 1) compared to insulin low-expressing cells (relative intensity < 1) within the ALDH3B2 mutant HPPD cell condition showed significant upregulation of key beta cell marker genes such as CHGA, IAPP, SCGN, SCG3 and SCG5 (Fig. 7f). Of note, gene set enrichment analysis (GSEA) confirmed the upregulation of key beta cell-specific gene sets related to peptide hormone metabolism, regulation of insulin secretion, and insulin processing (Fig. 7g). Re-analysis of insulin high-expressing cells in the ALDH3B2 mutant HPPD cells also identified heterogeneous cell populations with high beta cell identity (high INS/CHGA/IAPP co-expression), and polyhormonal cells (co-expression of INS/GCG/PPY), and endocrine progenitor-like cells (co-expression of INS and PAX6) (Extended Data Fig. 11). To further characterize the cell identity of insulin-expressing cells we compared the average gene expression of several marker genes specific for beta cells and other endocrine cells, endocrine progenitor cells, and duct cells³³ within all 13 cell clusters (Fig. 7h and 7i). In cluster 10 of the control condition, we detected insulin expression in about 40% of all cells, and the cells in this cluster show a low ductal cell-specific expression profile but high expression of endocrine progenitor cell marker such as PAX6 and INSM1 in 60–80% of all cells, indicating that the majority of cells in cluster 10 may represent endocrine progenitor-like cells. Interestingly, ALDH3B2 loss-of-function in cluster 10 shifts cell identity towards beta-like cells, shown by downregulation of almost all endocrine progenitor marker genes and sustained expression of INS and CHGA. Cluster 11 and 12 show altered duct cell identity compared to cluster 0 to 9 even in the control condition and seem to have a higher potential of beta-like cell transdifferentiation following loss-of-function of ALDH3B2. In comparison, cells in cluster 5 demonstrates strong duct cell identity but admit high potential for beta-like cell transdifferentiation as well.

To better understand why cells in cluster 5 and 12 have a higher chance to transdifferentiate into beta-like cells we performed trajectory inference analysis to identify the potential starting point of transdifferentiation (from low to high insulin expression, Extended Data Fig. 10c). Beta-like cells in cluster 5 may originated from cluster 2 and cluster 5 itself. Re-analysis of cluster 5 revealed several insulin low-expressing cell cluster as possible starting points. However, insulin positive cells in cluster 5 show significant upregulation of genes involved in energetic processes such as oxidative phosphorylation, aerobic respiration, and ATP synthesis may representing important prerequisites for beta-like cell transdifferentiation (Extended Data Fig. 10d and 10e). Cluster 3 may represent the originating cluster for beta-like cells in cluster 12. Differential gene expression analysis comparing cluster 2 and 3 (high potential for beta-like cell transdifferentiation) to cluster 4 (low potential of transdifferentiation) reveals that upregulation of genes important for translation and oxidative phosphorylation (cluster 2), enrichment of small GTPases RAC1/RHO (cluster 3), and elevated glycolysis (cluster 2 and 3) may favor beta-like cell transdifferentiation mediated by loss of function of ALDH3B2

(Extended Data Fig. 10f-k). In summary, ALDH3B2 loss-of-function may drive transdifferentiation of pancreatic duct cells partially through duct cell-derived endocrine progenitor-like stage, and then into beta-like cells that still keep partial duct cell identity. Elevation of energy metabolism such as oxidative phosphorylation and glycolysis may be a key step for pancreatic duct cells to transdifferentiate into beta-like cells.

Dynamic change in ALDH3B2 mutant human primary pancreatic duct cells transdifferentiation is revealed by RNA Velocity and PAGA Trajectory Analysis.

We next analyzed RNA velocity of our single cell data using scVelo³⁴ to investigate a possible transition among duct- and β -cell subclusters. scVelo identified 10 unique cell clusters, which are shown in a UMAP plot with streamlined velocities in Fig. 8a. The plot demonstrated a branching pattern emerging from cluster 2 towards other subclusters, although the specific paths were challenging to enumerate. So we next applied PAGA graph abstraction, which has been benchmarked as a top performing method for trajectory inference (Fig. 8b). It provides a graph-like map of the data topology with weighted edges corresponding to the connectivity between two clusters. We could see that the INS-expressing cells in cluster 5 were developed from cells in cluster 2, and then further developed into INS-expressing cells in cluster 9. The trajectory of INS-negative cells in the cluster also exhibited clear directionality, such as clusters (2_4), (2_1_6), and (2_8_0, 7, or 3).

The transcriptional dynamic model enabled the recovery of latent time associated with cellular processes, representing an internal clock for cells undergoing differentiation based on transcriptional dynamics. The velocity latent time highlighted a consistent developmental order, with all clusters progressing from cluster 2 to subsequent subclusters (Fig. 8c). Using velocity length to characterize the speed of transition or differentiation, we observed that clusters 2 and 1 exhibited significantly higher lengths, indicative of robust splicing activity and overall velocity confidence across ductal and β -cell subpopulations. (Fig. 8d). Furthermore, we evaluated the velocity of various representative genes associated with β - and duct-cells, revealing that the expression levels and velocity did not consistently align. Notably, INS expression peaked in the transitional clusters (2_5_9), while RNA velocity was predominantly higher across most clusters, excluding clusters 6, 7, and parts of 9. Selectively elevated INS expression was observed in β -cells, whereas a modest increase in RNA velocity was noted in the ductal subclusters. Subsequent analysis of the β -cell gene TRPM3 demonstrated a concordance between its expression pattern and RNA velocity in cluster 9. Similarly, examination of the duct cell-specific gene KRT19 revealed heightened expression in cluster 1 along with high RNA velocity, contrasting with the observation that the expression pattern of the duct cell marker CFTR did not align with RNA velocity in cluster 4 (Fig. 8e). Collectively, these findings lend support to the hypothesis that cells within cluster 5 may undergo a transition towards a more β -cell-like phenotype in response to ALDH3B2 loss-of-function.

Examining the top 40 driver genes within the relevant lineages (clusters 2_5_9), we identified clusters of genes exhibiting specific temporal abundance in distinct cell types. Heatmaps illustrated the primary

occurrences of spliced counts for the top-ranked dynamic genes in clusters 2, 5, and 9, organized by the latent time of cells (Fig. 8f). Notably, genes such as *ITGB1*³⁵, *NEDD9*³⁶, *HIF1A*³⁷, and *SERPINE1*³⁸, known for their involvement in transdifferentiation processes linked to TGF-beta signaling, were prominently featured. Furthermore, our analysis revealed a significant upsurge in the expression levels of genes like *ATP2A3*, *CACNA2D1*, *C2CD4A*, and *SCGN* during the later phase of latent time within a brief duration. This observation suggests a pivotal role for Ca^{2+} signaling in the transdifferentiation of duct cells into functional beta cells. Intriguingly, we also noted a dynamic reduction in *ALDH1A3* expression during this transdifferentiation process. Additionally, glycolysis-related genes such as *LDHA* and *HIF1A* were implicated in the transdifferentiation process, further underscoring the complexity and multifaceted nature of the cellular transformations taking place.

DISCUSSION

In this study, we have described the first unbiased and genome-wide CRISPR screen in search for genes that regulate the transdifferentiation of human pancreatic duct cells into insulin-producing beta-like cells. We show that loss-of-function of a single gene, *ALDH3B2*, in human pancreatic duct cells is sufficient to drive them towards a beta-like cell fate. Although the pancreatic duct-to-beta cell transdifferentiation in human has been observed, as evidenced by the existence of *INS*⁺/*CK19*⁺ cells in pancreatic ductal epithelium, it is still a relatively rare event with the percentage of *INS*⁺ pancreatic duct cells estimated at ~ 1%³⁹. Here we show that disruption of *ALDH3B2* can drive transdifferentiation of primary human pancreatic duct cells into beta-like cells with an efficiency of ~ 15%. This significant result carries potential for the development of a therapeutic intervention that could promote pancreatic duct-to-beta cell transdifferentiation for human beta cell mass replenishment.

We demonstrated that the trans-differentiated human pancreatic beta-like cells are functional, responsive to glucose challenge *in vivo*, and able to significantly lower blood glucose in diabetic animal models. Notably, neither *ALDH3B2*^{mut} PANC-1 cells nor HPPD cells were able to lower blood glucose to euglycemic levels in our studies, and this could be due to two reasons: (1) The transplanted beta-like cells number was not enough due to experimental limitation, or (2) the glucose sensitivity or set-point of transdifferentiated beta-like cells may be different from that of true pancreatic beta cells. In a separate experiment of transplantation of *ALDH3B2*^{mut}-HPPD cells into euglycemic NSG mice, we performed glucose tolerance (GTT) measurements. The peak glucose level in this experiment was approximately 250mg/dL, but transplanted *ALDH3B2*^{mut}-HPPD cells were still able to improve the glucose tolerance (Extended Data Fig. 12a) and secrete human insulin in response to glucose challenge (Extended Data Fig. 12b and 7c). Based on these data, it seems unlikely that transdifferentiated beta-like cells had a higher glucose set-point than primary pancreatic beta cells. We may be able to lower the blood glucose further in diabetic mice if a larger number of transdifferentiated cells were transplanted. To this end, we will need to improve lentiviral transduction efficiency and perhaps identify more effective gRNA sequences that target *ALDH3B2* gene.

At first glance in comparison with primary human islets, the ALDH3B2 mutation transdifferentiated beta-like cells seem to have only modest expression of key beta cell markers (Fig. 4d-g), ability of lowering blood glucose and secrete human insulin in diabetic mice (Fig. 5a-c) and changes in DNA methylation status on insulin promoter (Fig. 6c and 6d). However, all these measurements and characterization were done on a mixed cell population with only less than 15% of the cells being transdifferentiated beta-like cells. In theory, the actual changes in each transdifferentiated beta-like cell should be more dramatic than what our experimental data showed. Our findings of immunofluorescent staining for insulin and key beta cell transcription factors and insulin granules with abundant ER ultrastructurally in the transdifferentiated beta-like cells strongly support the conclusion that loss-of-function of ALDH3B2 does induce *bona fide* cell transdifferentiation from human pancreatic duct cells into functional beta-like cells.

It is interesting that the ALDH family is often considered as a stem cell or progenitor marker. High aldehyde dehydrogenase activity (using Aldefluor assay^{40,41}) has been widely used to identify adult stem cells or progenitor cells in various tissues/organs, including hematopoietic stem cells (HSC)⁴¹, neuronal progenitor cells (NPC)^{42,43} or potential pancreatic progenitor cells⁴⁴. However, the Aldefluor assay cannot distinguish between the activity of different aldehyde dehydrogenases, so different ALDH members may be marking stem/progenitor cells in different tissues/organs. The function of ALDH3B2 is not yet well understood. One rodent study reported that ALDH3B2 localizes to lipid droplets in cells and catalyzes the conversion of long-chain fatty aldehydes into long-chain fatty acids⁴⁵. How this function would impact pancreatic duct cell transdifferentiation into beta-like cells is unclear and warrants further studies. Notwithstanding, the discovery of ALDH3B2 as a regulator of pancreatic duct-to-beta cell transdifferentiation provides a novel therapeutic target for pancreatic beta cell mass restoration. Endogenous pancreatic duct cells could potentially be targeted by gene-editing to mutate ALDH3B2 and induce transdifferentiation. Alternatively, ALDH3B2 enzymatic activity may be targeted by small molecules inhibitors to achieve the effects similar to those we observed using genetic disruption of the ALDH3B2 gene. The ALDH3B2 enzymatic assay established by Kitamura and colleagues⁴⁵ using long chain fatty aldehyde as substrate could be a useful chemical screen platform for potential new drug discovery. In support of this approach, the broad Aldehyde dehydrogenase (ALDH) inhibitors diethylaminobenzaldehyde (DEAB) and Disulfiram (DSF) have been shown to promote beta cell differentiation in zebrafish and in PANC-1 cells⁴⁶, where it is possible that DEAB and DSF elicit their effect through inhibiting ALDH3B2. It will be of great importance to find out which aldehydes are unique substrates for ALDH3B2 and search for ALDH3B2-specific inhibitors, since mutation of the close member ALDH3A1 does not have the same effect (Fig. 1c and 1d) and another close member, ALDH1A3, was shown to be involved in the de-differentiation of pancreatic beta cells in Type 2 diabetes patients²⁹.

Given the interest in understanding the mechanisms driving beta cell transdifferentiation induced by ALDH3B2 loss-of-function, our primary focus has been on the progression of the INS⁺ cluster. By further examining the subcluster of INS⁺ cells that emerge when incorporating ALDH3B2 mutant datasets, we can visually depict the pathways in a coherent manner (refer to Fig. 8b and 8e). This analysis sheds light on a specific set of connections along the 2_5_9 differentiation axis, highlighting the transdifferentiation

from pancreatic duct cells to beta cells. This dynamic model enhances our ability to interpret velocity and identify driver genes more effectively (see Fig. 8f) by exploring correlations between gene kinetics and expression patterns across cells. Notably, we observe that the key potential driver genes are associated with TGF-beta signaling^{47,48}, glycolysis²⁵, and Ca²⁺ signaling pathways⁴⁹. These findings are consistent with previous data from other cell types and may serve as valuable insights for future investigations focused on pancreatic duct-to-beta cell transdifferentiation.

Our study identifying ALDH3B2 as a regulator of pancreatic duct-to-beta cell transdifferentiation was conducted using a human pancreatic duct cell line and primary human pancreatic duct cells. Given the difficulties of translating rodent studies into human in the beta cell regeneration research field, the findings presented here in human cells are directly relevant and have clear potential for the development of human diabetes therapeutics.

MATERIALS AND METHODS

Mice

NSG (NOD.Cg-Prkdc^{scid} Il2rg^{tm1Wjl}/SzJ) mice were purchased from the Jackson Laboratory (Bar Harbor, ME). Animals were housed in pathogen-free facilities at the Joslin Diabetes Center and all experimental procedures were approved and performed in accordance with institutional guidelines and regulations.

REPB reporter construction and REPB PANC-1 cell generation

The REPB reporter lentivirus vector was constructed by assembling Rat insulin promoter, RIP3.1 promoter²⁰, EGFP and Blasticidin-S deaminase (BSD). The EGFP and BSD genes are fused together with P2A peptide. The REPB reporter lentivirus was used to infect PANC-1 cells (ATCC #CRL-1469), and the infected PANC-1 cells are then single-cell sorted by FACS. PANC-1 clones with confirmed REPB reporter genome incorporation were used in the genome-wide CRISPR screen.

CRISPR GeCKO library screen

The human GeCKO-v2 (Genome-Scale CRISPR Knock-Out) lentiviral pooled library was obtained from Addgene (Addgene, # 1000000048) and was prepared as previously described⁵⁰. 100 million REPB PANC-1 cells were infected with human GeCKO CRISPR lentiviral library at MOI of 0.3, and then subsequently selected with puromycin (2.5 µg/ml) at day 3 post lentiviral infection. After cells recover from puromycin selection, 20 million cells were collected as baseline control (CON-1 and CON-2), the rest of the cells were further selected with blasticidin (10 µg/ml) for 7 days. The blasticidin-resistant cells were allowed to grow back to full confluence and then subjected to FACS sorting on their EGFP intensity. Cell population with the highest EGFP intensity were collected as experiment group (EXP-1 and EXP-2). Genomic DNA was extracted from the cells (Quick-gDNA midiprep kit, Zymo Research), the NGS

(Next Generation Sequencing) libraries were prepared as previously described⁵¹, and then subjected to NGS sequencing analysis (Novogene, CA). The gRNA sequences from the NGS sequencing data were extracted using standard bioinformatics methods, and the read count of gRNAs were calculated as Count Per Million (CPM).

Analysis of pooled CRISPR screen

To identify enriched genes based on sgRNAs in this CRISPR screen, we used the MAGeCK algorithm^{22,23} in R. We used MAGeCK's 'MLE' subcommand for maximum likelihood estimation of gene essentialities using the TMM normalized counts⁵² at the gene level. This was followed by applying MAGeCKFlute's²⁴ function FluteMLE with the argument to incorporate DepMap genes⁵³ in normalization, as recommended for human data, which produced pathway enrichment analysis. Among screening hits from MAGeCKFlute, protein connectivity networks based on physical and functional interactions were identified using STRING v12⁵⁴, where only interactions with a medium confidence score of ≥ 0.4 were selected.

Cell lines

PANC-1 (#CRL-1469) and 293FT (#R7007) cell lines were obtained from ATCC and Thermo Fisher Scientific, respectively. Cells were maintained in DMEM (Gibco, 10313039), supplemented with 10% fetal bovine serum (FBS, Gibco), L-alanyl-L-glutamine (Gibco) and penicillin/streptomycin (Corning), in a 37°C incubator with 5% CO₂. To generate non-targeting control (NTC) and ALDH3B2^{mut} PANC-1 cells, non-targeting (NT) gRNA (5'- GCTTTCACGGAGGTTTCGACG-3') or ALDH3B2 gRNA (5'- GCCCTCCTCACCTGCGGCGA-3', HGLibA_01571) oligos were cloned into LentiCRISPR-v2 vector. Wild type PANC-1 cells were then transduced by NTC or ALDH3B2 gRNA- containing lentivirus, and subsequently selected by puromycin treatment. Indel mutation in ALDH3B2^{mut} cells was confirmed by deep sequencing analysis (MGH DNA Core Facility, Cambridge, MA). All plasmid sequences were verified by Sanger sequencing before transduction and transfection. To generate shControl and shALDH3B2 PANC-1 cells, a scrambled or ALDH3B2-targeting shRNA was cloned into FH1t(INSR)UTG-GFP vector (A gift from Dr. Stephan Kissler). shControl and shALDH3B2 lentiviruses were then used to infect PANC-1 cells and the cells were selected by puromycin treatment.

PANC-1 cells transplantation studies

Experimental diabetes was induced in 8-week-old NSG male mice by intraperitoneal injection of streptozotocin (STZ) (40 mg STZ/kg body weight for five consecutive days). Animals were considered diabetic only if morning-fed blood glucose exceeded 350 mg/dl. Three days after STZ injection, $\sim 10^7$ PANC-1 cells (carry NT and ALDH3B2^{mut}) were transplanted subcutaneously into each diabetic NSG mouse. Blood glucose was monitored every 3–4 days. Six weeks post cells transplantation, Intraperitoneal glucose tolerance test (IPGTT) was performed. Mice were fasted for 16 hours. The plasma glucose levels of the mice before (baseline) or 15, 30, 60, 90, and 120 minutes after

intraperitoneal injection of 2 mg/g body weight glucose were recorded by a Glucose Meter. At day 52 post cells transplantation, grafts were surgically removed from NSG mice, and the blood glucose was measured 4 days later.

Human pancreatic ductal cell isolation and purification

Human primary pancreatic ductal cells isolation was performed as previously described^{30,55}. In brief, “Human acinar tissue” from Integrated Islet Distribution Program (IIDP) (Donor information is summarized in Extended Table 2) was washed 2 times with PBS, and then incubated with Trypsin solution (1.5ml 0.25% Trypsin in 20ml PBS) shaking at 37°C for 15 min. Dispersed cells were centrifuged at 1,000 rpm for 5 min, the supernatant was aspirated, and then the pellets were resuspended with mouse antihuman CA19-9 antibody (Invitrogen; clone 116-NS-19-9) in 2 ml PBS solution. After 15 min incubation at 4°C, the cell suspension was mixed with 10ml PBS solution (375mg EDTA and 2.5g BSA in 500ml PBS) gently. Tubes were centrifuged at 1,000 rpm for 5 min, supernatant was aspirated, 250 µl/tube goat anti mouse IgG microbeads (Miltenyi Biotec) in PBS solution were added, and pellets were mixed. After 20 min incubation at 4°C, wash with PBS solution 2 times. Tubes were centrifuged at 1,000 rpm for 5 min, and pellets were resuspended in 20 ml cold PBS solution and passed through 40 µm cell strainers to remove newly formed clumps of cells. MACS magnetic LS separation columns (Miltenyi Biotec) were prepared according to the manufacturer’s instructions.

Preparation and transplantation of primary human pancreatic duct cells

Human primary pancreatic ductal cells isolation was performed as above described. Purified human primary pancreatic ductal cells (HPPD) were immediately cultured in a low-attachment plate in RMPI DMEM/F12 medium (Gibco), supplemented with 10% FBS and penicillin/streptomycin. Lentivirus encoding a NT or ALDH3B2 gRNA together with Cas9 endonuclease was added to the culture media for overnight infection. The next day, HPPD cells were washed with culture media twice and $\sim 10^7$ cells were transplanted under the left kidney capsule of 8-week-old of STZ induced diabetic male NSG mice. Graft recipients were left to recover from surgery for three weeks. At day 56 post-transplantation, the kidney transplanted with grafts were surgically removed for gene expression analysis by immunofluorescence. The final blood glucose measurement of NSG mice was done 4 days later.

Quantitative real-time PCR (qPCR)

Cells or grafts were treated with TRIzol (Thermo Fisher Scientific) for RNA extraction following the manufacturer’s protocol. Purified RNA was reverse-transcribed into cDNA using the SuperScript IV first-strand synthesis kit (Invitrogen). INS (Hs00355773_m1), GCG (Hs01031536_m1), SST (Hs00356144_m1) PDX1(Hs00236830_m1), NKX6-1 (Hs00232355_m1), GCK (Hs01564555_m1), SLC2A2 (Hs00165775_m1), SLC2A1 (Hs00892681_m1), CPE (Hs00960598_m1), CA2 (Hs01070108_m1), KRT19 (Hs00761767_s1), SOX9 (Hs00165814_m1), ALDH3B2 (Hs02511514_s1) and ACTB (Hs01060665_g1) probes for TaqMan assays were purchased from Thermo Fisher Scientific. All Gene

expression levels were analyzed by SYBR green PowerUp qPCR assays (Applied Biosystems). Primer sequences used is shown in the following table:

Gene	Forward primer sequence	Reverse primer sequence
ACTB	5' CACCATTGGCAATGAGCGGTTC 3'	5'AGGTCTTTGCGGATGTCCACGT 3'
INS	5' ACGAGGCTTCTTCTACACACCC 3'	5' TCCACAATGCCACGCTTCTGCA 3'
PDX1	5'GAAGTCTACCAAAGCTCACGCG 3'	5' GGAACTCCTTCTCCAGCTCTAG 3'
NKX6.1	5' CCTATTCGTTGGGGATGACAGAG 3'	5' TCTGTCTCCGAGTCCTGCTTCT 3'
PAX6	5' CTGAGGAATCAGAGAAGACAGGC 3'	5' ATGGAGCCAGATGTGAAGGAGG 3'
MAFA	5' GCTTCAGCAAGGAGGAGGTCAT 3'	5' TCTGGAGTTGGCACTTCTCGCT 3'
NEUROD	5' GGTGCCTTGCTATTCTAAGACGC 3'	5' GCAAAGCGTCTGAACGAAGGAG 3'
NGN3	5' CCTAAGAGCGAGTTGGCACTGA 3'	5' AGTGCCGAGTTGAGGTTGTGCA 3'
GCG	5' CGTTCCCTTCAAGACACAGAGG 3'	5' ACGCCTGGAGTCCAGATACTTG 3'
SST	5' CCAGACTCCGTCAGTTTCTGCA 3'	5' TTCCAGGGCATCATTCTCCGTC 3'
PP	5' AGACACAAAGAGGACACGCTGG 3'	5' GAGTCGTAGGAGACAGAAGGTG 3'
SLC2A2	5' ATGTCAGTGGGACTTGTGCTGC 3'	5' AACTCAGCCACCATGAACCAGG 3'
GCK	5' CATCTCCGACTTCCTGGACAAG 3'	5' TGGTCCAGTTGAGAAGGATGCC 3'
ABCC8	5' GACGACAAGAGGACAGTGGTCT 3'	5' GCATTCAGACCTCTGGAAGTCC 3'
KCNJ11	5' TGTGTCACCAGCATCCACTCCT 3'	5' GTTCTGCACGATGAGGATCAGG 3'
IA2	5' TGGAGATCCTGGCTGAGCATGT 3'	5' GGTCACATCAGCCAAAGACAGG 3'
KRT19	5' AGCTAGAGGTGAAGATCCGCGA 3'	5' GCAGGACAATCCTGGAGTTCTC 3'
CA2	5' GTGACCTGGATTGTGCTCAAGG 3'	5' GTTGTCCACCATCAGTTCTTCGG 3'
HNF1B	5' CCCAGCAAATCTTGTACCAGGC 3'	5' ACCTCAGTGACCAAGTTGGAGC 3'
SOX9	5' AGGAAGCTCGCGGACCAGTAC 3'	5' GGTGGTCCTTCTTGTGCTGCAC 3'

All qPCR assays were performed using a QuantStudio 6 Flex Real-Time PCR system (Applied Biosystems).

Immunofluorescence staining and confocal microscopy

The subcutaneously transplanted PANC-1 cells or the kidney with HPPD graft transplantation were surgically removed from the mice, fixed 1 hour in 4% paraformaldehyde at 4°C, and dehydrated using 30% sucrose solution overnight. The tissues were embedded in disposable base molds (Thermo Fisher Scientific) and 10 mm sections were cut. For staining, slides were blocked with PBS + 0.1% Triton X-100

(Thermo Fisher Scientific) + 5% donkey serum (Sigma-Aldrich) for 1 hr at room temperature (RT), incubated with primary antibodies overnight at 4°C, washed, incubated with secondary antibody incubation for 1 hr at RT, incubated with Hoechst 33342 (Invitrogen) for 10 min at RT, and washed. For imaging, samples were mounted in fluorescence mounting medium (Dako), covered with coverslips, and sealed with nail polish. Representative images were taken using a Zeiss LSM 710 confocal microscope. Primary antibody: Insulin (A0564, Dako), Pdx1 (5679S, Cell Signaling Technology),

Dykdddk Tag (14793S, Cell Signaling Technology), C-peptide (GN-ID4, Developmental Studies Hybridoma Bank) and Cytokeratin 19 (Abcam, ab7754).

Cells were seeded into 4-well culture slide (Falcon) at density of 10^5 cells/well. After another 24 hours, the cells were fixed, stained and subjected to fluorescence microscopic analysis as above described.

Serum insulin measurement

Mouse blood was collected from the tail tip and allowed to clot. Serum was separated by brief centrifugation according to standard protocol. The serum insulin level was measured using STELLUX® Chemi Human Insulin ELISA kits (Alpco).

Electron microscopy

NTC-PANC-1 or ALDH3B2^{mut}-PANC-1 were fixed at RT for 2 hr with a mixture containing 1.25% PFA, 2.5% glutaraldehyde, and 0.03% picric acid in 0.1M sodium cacodylate buffer (pH 7.4). Samples were then sent to the Advanced Microscopy Core of Joslin for further processing and transmission electron microscope imaging.

DNA methylation analyses

Bisulfite conversion (Zymo Research, EZ DNA Methylation-Direct Kits) of DNA from PANC-1, human primary pancreatic ductal cells carry NT or ALDH3B2^{mut} and human islets were performed as described previously⁵⁶. Bisulfite-treated DNA was PCR amplified, using primers (human INS promoter forward primer – 5' AGGATAGGTTGTATTAGAAGAGGTTATTAAG 3'; human INS promoter reverse primer- 5' CCCCTAAACTCACCCCCACATACTTC 3') specific for bisulfite treated DNA but independent of methylation status at monitored CpG sites. Reaction conditions for the first round of PCR were 5 cycles of 95°C 1 min, 52°C 3 min, 72°C 3 min followed by 40 cycles of 95°C 30 s, 55°C 45 s, 72°C 45 s followed by 7 min at 72°C. PCR products were gel purified and used for deep sequencing analysis (MGH DNA Core Facility, Cambridge, MA).

Single cell RNA sequencing

Isolation of human primary pancreatic ductal cells was performed as described above. Purified human primary pancreatic ductal cells (HPPD) were immediately cultured in a low-attachment plate in RMPI DMEM/F12 medium (Gibco), supplemented with 10% FBS and penicillin/streptomycin. Lentivirus encoding a NT or *ALDH3B2* gRNA together with a Cas9-mCherry reporter construct was added to the

culture media. 6 days later cells were harvested and alive mCherry positive cells were isolated by FACS sorting. 40,000 cells per sample were used for scRNAseq using the Chromium Next GEM Single Cell 3' GEM, Library & Gel Bead Kit v3.1 (cat # PN-1000213; 10 x Genomics) according to manufactures' instructions. Illumina NovaSeq 6000 with about 1.3 billion reads total was used for sequencing the purified 3' gene expression library. The single cell RNA-seq dataset was processed, explored and visualized using Cellenics® community instance (<https://scp.biomage.net/>) that's hosted by Biomage (<https://biomage.net/>) using preset standard filtering thresholds and integration methods. Pre-filtered count matrices were uploaded to Cellenics. Dead or dying cells were removed by filtering droplets with high mitochondrial content (3% cut-off). Outliers in the distribution of number of genes vs number of UMIs were removed by fitting a linear regression model (p-values between 1.97E-04–2.34E-04). Cells with a high probability of being doublets were filtered out using the scDbtFinder method (threshold range: 0.50–0.56). Overall filtering rates after processing are in the range of 10.75% – 11.35% of cells. Data normalization, principal component analysis (PCA) and data integration using Harmony were performed on data from high-quality cells. Clusters were identified using the Louvain method, and a Uniform Manifold Approximation and Projection (UMAP) embedding was calculated to visualize the results. Cluster-specific marker genes were identified by comparing cells of each cluster to all other cells using the presto package implementation of the Wilcoxon rank-sum test.

Modeling transcriptional dynamics of scRNA-seq database by seVelo

To recover the directed dynamic information of newly transcribed, unspliced pre-mRNAs and mature, spliced mRNAs, we used *velocyto*⁵⁷ to count both the unspliced and spliced RNA reads from the counts that were generated by 10X Cellranger⁵⁸. We filtered out genes that had less than 10 counts and were expressed in less than 10 cells (both unspliced and spliced), normalized every cell by its total counts, and logarithmize the normalized counts. We then filtered out cells that had less than 1000 total counts. Finally, we kept the top 3000 highly variable genes. We computed the first and second-order moments (means and uncentered variances) among nearest neighbors in PCA space. First-order is needed for deterministic velocity estimation, while stochastic estimation also requires second-order moments. Transcriptional induction for a particular gene increases newly transcribed precursor unspliced mRNAs while, conversely, repression or absence of transcription decreases unspliced mRNAs. We used *scVelo*³⁴ to solve the full transcriptional dynamics of splicing kinetics using a likelihood-based dynamical model.

Statistical analyses

Statistical analyses were performed by unpaired or paired tests as indicated using the Prism 8 software. All data are presented as mean ± SEM. Significance was defined as *p < 0.05, **p < 0.01 and ***p < 0.001. No samples were excluded from the analysis. Data analysis was not blinded. All data are representative of two or more similar experiments.

Declarations

ACKNOWLEDGEMENTS

This research was supported by funds from the Beatson Foundation to P.Y and S.B-W and by postdoctoral fellowship to J. L. from the Mary K. Iacocca Foundation and JDRF. We acknowledge the support from core facilities funded both by the NIDDK Diabetes Research Center (P30DK036836) and the Joslin Diabetes Center, especially our CRISPR screen core facility and bioinformatic core (Dr. Hui Pan and Dr. Jonathan Dreyfuss). We also acknowledge the generous comments and editing from Dr. Gordon Weir and Dr. Stephan Kissler of Joslin Diabetes Center.

References

1. Weir, G.C., Bonner-Weir, S. & Leahy, J.L. Islet mass and function in diabetes and transplantation. *Diabetes* 39, 401–5 (1990).
2. Nair, G. & Hebrok, M. Islet formation in mice and men: lessons for the generation of functional insulin-producing beta-cells from human pluripotent stem cells. *Curr Opin Genet Dev* 32, 171–80 (2015).
3. Nostro, M.C. *et al.* Efficient generation of NKX6-1 + pancreatic progenitors from multiple human pluripotent stem cell lines. *Stem Cell Reports* 4, 591–604 (2015).
4. Pagliuca, F.W. *et al.* Generation of functional human pancreatic beta cells in vitro. *Cell* 159, 428 – 39 (2014).
5. Shapiro, A.M. *et al.* Islet transplantation in seven patients with type 1 diabetes mellitus using a glucocorticoid-free immunosuppressive regimen. *N Engl J Med* 343, 230–8 (2000).
6. Aguayo-Mazzucato, C. & Bonner-Weir, S. Pancreatic beta Cell Regeneration as a Possible Therapy for Diabetes. *Cell Metab* 27, 57–67 (2018).
7. Dor, Y., Brown, J., Martinez, O.I. & Melton, D.A. Adult pancreatic beta-cells are formed by self-duplication rather than stem-cell differentiation. *Nature* 429, 41 – 6 (2004).
8. Butler, A.E. *et al.* Adaptive changes in pancreatic beta cell fractional area and beta cell turnover in human pregnancy. *Diabetologia* 53, 2167–76 (2010).
9. Ben-Othman, N. *et al.* Long-Term GABA Administration Induces Alpha Cell-Mediated Beta-like Cell Neogenesis. *Cell* 168, 73–85 e11 (2017).
10. Collombat, P. *et al.* The ectopic expression of Pax4 in the mouse pancreas converts progenitor cells into alpha and subsequently beta cells. *Cell* 138, 449–62 (2009).
11. Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J. & Melton, D.A. In vivo reprogramming of adult pancreatic exocrine cells to beta-cells. *Nature* 455, 627 – 32 (2008).
12. Bonner-Weir, S., Aguayo-Mazzucato, C. & Weir, G.C. Dynamic development of the pancreas from birth to adulthood. *Ups J Med Sci* 121, 155–8 (2016).

13. Bonner-Weir, S. *et al.* The pancreatic ductal epithelium serves as a potential pool of progenitor cells. *Pediatr Diabetes* 5 Suppl 2, 16–22 (2004).
14. Inada, A. *et al.* Carbonic anhydrase II-positive pancreatic cells are progenitors for both endocrine and exocrine pancreas after birth. *Proc Natl Acad Sci U S A* 105, 19915–9 (2008).
15. Qadir, M.M.F. *et al.* P2RY1/ALK3-Expressing Cells within the Adult Human Exocrine Pancreas Are BMP-7 Expandable and Exhibit Progenitor-like Characteristics. *Cell Rep* 22, 2408–2420 (2018).
16. Qadir, M.M.F. *et al.* Single-cell resolution analysis of the human pancreatic ductal progenitor cell niche. *Proc Natl Acad Sci U S A* 117, 10876–10887 (2020).
17. Wang, D. *et al.* Long-Term Expansion of Pancreatic Islet Organoids from Resident Procr(+) Progenitors. *Cell* 180, 1198–1211 e19 (2020).
18. Lieber, M., Mazzetta, J., Nelson-Rees, W., Kaplan, M. & Todaro, G. Establishment of a continuous tumor-cell line (panc-1) from a human carcinoma of the exocrine pancreas. *Int J Cancer* 15, 741–7 (1975).
19. Madden, M.E. & Sarraz, M.P., Jr. Morphological and biochemical characterization of a human pancreatic ductal cell line (PANC-1). *Pancreas* 3, 512 – 28 (1988).
20. Chai, R., Chen, S., Ding, J. & Grayburn, P.A. Efficient, glucose responsive and islet-specific transgene expression by a modified rat insulin promoter. *Gene Ther* 16, 1202–9 (2009).
21. Sanjana, N.E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* 11, 783–784 (2014).
22. Li, W. *et al.* MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol* 15, 554 (2014).
23. Li, W. *et al.* Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. *Genome Biology* 16, 281 (2015).
24. Wang, B. *et al.* Integrative analysis of pooled CRISPR genetic screens using MAGeCKFlute. *Nat Protoc* 14, 756–780 (2019).
25. Lai, L., Reineke, E., Hamilton, D.J. & Cooke, J.P. Glycolytic Switch Is Required for Transdifferentiation to Endothelial Lineage. *Circulation* 139, 119–133 (2019).
26. Vaughn, N. & Haviland, D.L. Acly promotes metabolic reprogramming and induction of IRF4 during early CD8(+) T cell activation. *Cytometry A* 99, 825–831 (2021).
27. Marchitti, S.A., Brocker, C., Stagos, D. & Vasiliou, V. Non-P450 aldehyde oxidizing enzymes: the aldehyde dehydrogenase superfamily. *Expert Opin Drug Metab Toxicol* 4, 697–720 (2008).
28. Vassalli, G. Aldehyde Dehydrogenases: Not Just Markers, but Functional Regulators of Stem Cells. *Stem Cells Int* 2019, 3904645 (2019).
29. Kim-Muller, J.Y. *et al.* Aldehyde dehydrogenase 1a3 defines a subset of failing pancreatic beta cells in diabetic mice. *Nat Commun* 7, 12631 (2016).
30. Yatoh, S. *et al.* Differentiation of affinity-purified human pancreatic duct cells to beta-cells. *Diabetes* 56, 1802–9 (2007).

31. Wescott, M.P. *et al.* Pancreatic ductal morphogenesis and the Pdx1 homeodomain transcription factor. *Mol Biol Cell* 20, 4838–44 (2009).
32. Neiman, D. *et al.* Islet cells share promoter hypomethylation independently of expression, but exhibit cell-type-specific methylation in enhancers. *Proc Natl Acad Sci U S A* 114, 13525–13530 (2017).
33. Ma, Z. *et al.* Deciphering early human pancreas development at the single-cell level. *Nat Commun* 14, 5354 (2023).
34. Bergen, V., Lange, M., Peidli, S., Wolf, F.A. & Theis, F.J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol* 38, 1408–1414 (2020).
35. Sarnobat, D. *et al.* Taurine rescues pancreatic beta-cell stress by stimulating alpha-cell transdifferentiation. *Biofactors* 49, 646–662 (2023).
36. Kong, C. *et al.* NEDD9 is a positive regulator of epithelial-mesenchymal transition and promotes invasion in aggressive breast cancer. *PLoS One* 6, e22666 (2011).
37. Liu, C. *et al.* HIF1alpha Regulates Early Metabolic Changes due to Activation of Innate Immunity in Nuclear Reprogramming. *Stem Cell Reports* 14, 192–200 (2020).
38. Kong, H.J. *et al.* Crosstalk between YAP and TGFbeta regulates SERPINE1 expression in mesenchymal lung cancer cells. *Int J Oncol* 58, 111–121 (2021).
39. Mezza, T. *et al.* Insulin resistance alters islet morphology in nondiabetic humans. *Diabetes* 63, 994–1007 (2014).
40. Jones, R.J. *et al.* Assessment of aldehyde dehydrogenase in viable cells. *Blood* 85, 2742–6 (1995).
41. Storms, R.W. *et al.* Isolation of primitive human hematopoietic progenitors on the basis of aldehyde dehydrogenase activity. *Proc Natl Acad Sci U S A* 96, 9118–23 (1999).
42. Cai, J. *et al.* Membrane properties of rat embryonic multipotent neural stem cells. *J Neurochem* 88, 212 – 26 (2004).
43. Obermair, F.J. *et al.* A novel classification of quiescent and transit amplifying adult neural stem cells by surface and metabolic markers permits a defined simultaneous isolation. *Stem Cell Res* 5, 131–43 (2010).
44. Rovira, M. *et al.* Isolation and characterization of centroacinar/terminal ductal progenitor cells in adult mouse pancreas. *Proc Natl Acad Sci U S A* 107, 75–80 (2010).
45. Kitamura, T., Takagi, S., Naganuma, T. & Kihara, A. Mouse aldehyde dehydrogenase ALDH3B2 is localized to lipid droplets via two C-terminal tryptophan residues and lipid modification. *Biochem J* 465, 79–87 (2015).
46. Rovira, M. *et al.* Chemical screen identifies FDA-approved drugs and target pathways that induce precocious pancreatic endocrine differentiation. *Proc Natl Acad Sci U S A* 108, 19264–9 (2011).
47. Ninomiya, K., Takahashi, A., Fujioka, Y., Ishikawa, Y. & Yokoyama, M. Transforming growth factor-beta signaling enhances transdifferentiation of macrophages into smooth muscle-like cells. *Hypertens Res* 29, 269–76 (2006).

48. Miettinen, P.J., Ebner, R., Lopez, A.R. & Derynck, R. TGF-beta induced transdifferentiation of mammary epithelial cells to mesenchymal cells: involvement of type I receptors. *J Cell Biol* 127, 2021–36 (1994).
49. Sabatini, P.V., Speckmann, T. & Lynn, F.C. Friend and foe: beta-cell Ca(2+) signaling and the development of diabetes. *Mol Metab* 21, 1–12 (2019).
50. Joung, J. *et al.* Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat Protoc* 12, 828–863 (2017).
51. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 343, 84–87 (2014).
52. Robinson, M. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology* 11, 1–9 (2010).
53. Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell* 170, 564–576 e16 (2017).
54. Szklarczyk, D. *et al.* STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 47, D607–D613 (2019).
55. Corritore, E. *et al.* beta-Cell differentiation of human pancreatic duct-derived cells after in vitro expansion. *Cell Reprogram* 16, 456–66 (2014).
56. Millar, D.S., Warnecke, P.M., Melki, J.R. & Clark, S.J. Methylation sequencing from limiting DNA: embryonic, fixed, and microdissected cells. *Methods* 27, 108 – 13 (2002).
57. La Manno, G. *et al.* RNA velocity of single cells. *Nature* 560, 494–498 (2018).
58. Zheng, G.X. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat Commun* 8, 14049 (2017).

Figures

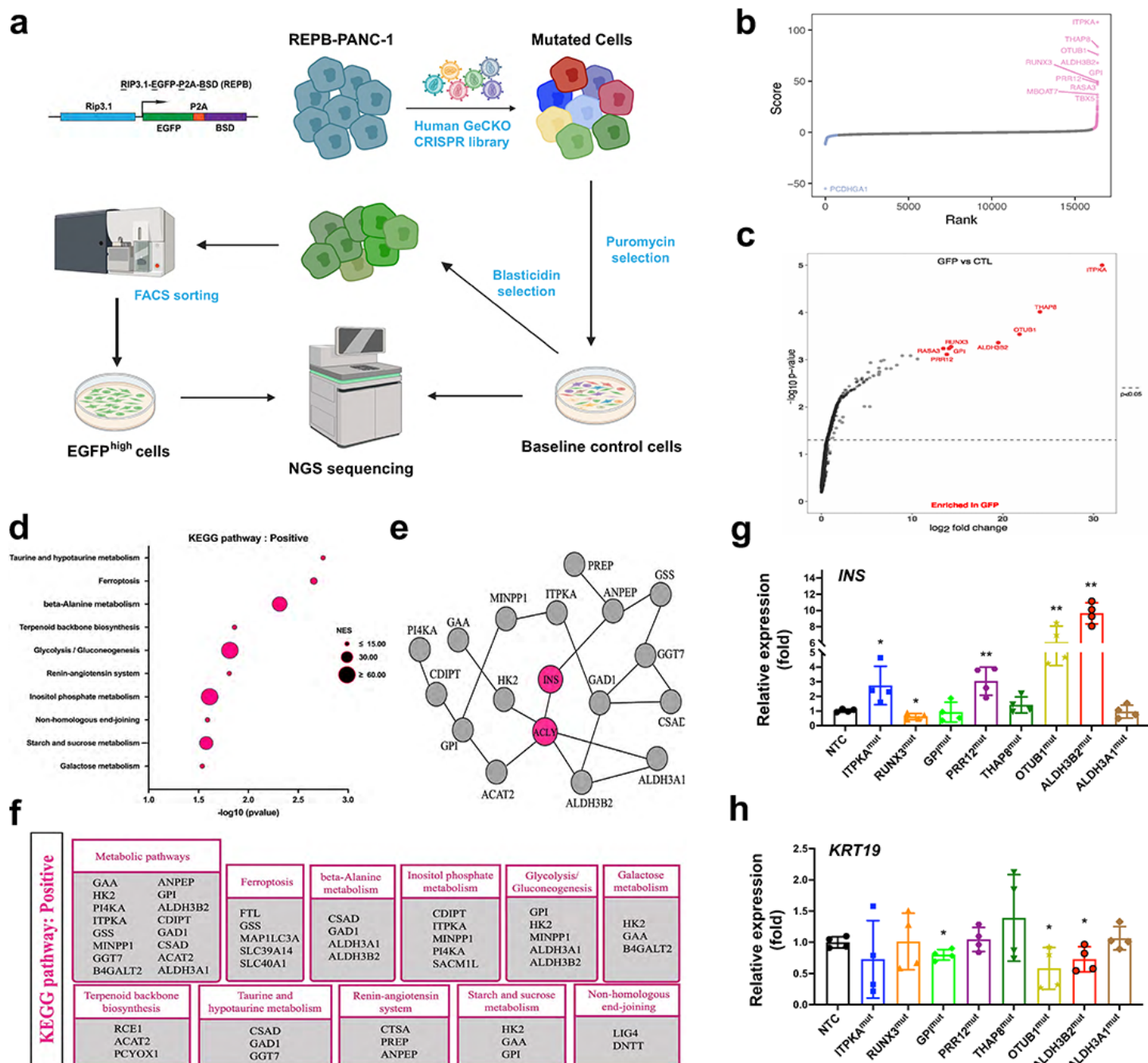


Fig. 1

Figure 1

Genome-wide CRISPR screen identifies ALDH3B2 as a regulator of pancreatic duct-to-beta cell transdifferentiation.

a, The structure of REPB reporter and an illustration of genome-wide CRISPR screen workflow. The REPB reporter contains Rat insulin promoter (RIP3.1) driving the expression of EGFP and Blasticidin-S deaminase (BSD), the EGFP and BSD genes are fused together with P2A peptide. **b**, Rank all the genes based on their scores and label genes in the rank plot. Y-axis represents difference in scores between EGFP^{high}, blasticidin-resistant REPB-PANC-1 cells vs control cells from MAGECKFlute at the gene level,

where positive scores indicate positive enrichment and negative scores indicate negative enrichment; x-axis represents the gene rank. Large magnitude Score are colored, where the most extreme genes are labeled. **c**, Half-volcano plot at the gene level showing positive enrichment where x-axis is $\log_2$ fold change of EGFP^{high}, blasticidin-resistant REPB-PANC-1 cells vs control cells from MAGECK-mle and y-axis is significance that was winsorized to 10^{-5} . The strongest hits are colored in red and labeled. The horizontal dashed line represents a p-value of 0.05. **d**, KEGG pathways enriched for positively selected genes from MAGECKFlute where dot size indicates Normalized Enrichment Score (NES). **e and f**, STRING pathway analysis showing protein-protein associations including physical and functional interactions between INS, ACLY, and other GFP-positive screening hits from MAGECKFlute (**e**) and for GFP-positive screening hits from MAGECKFlute clustered into functionally associated groups (**f**). **g and h**: qPCR analysis of INS (**g**) and KRT19 (**h**) expression level for top 8 candidate gene mutant PANC-1 cells.

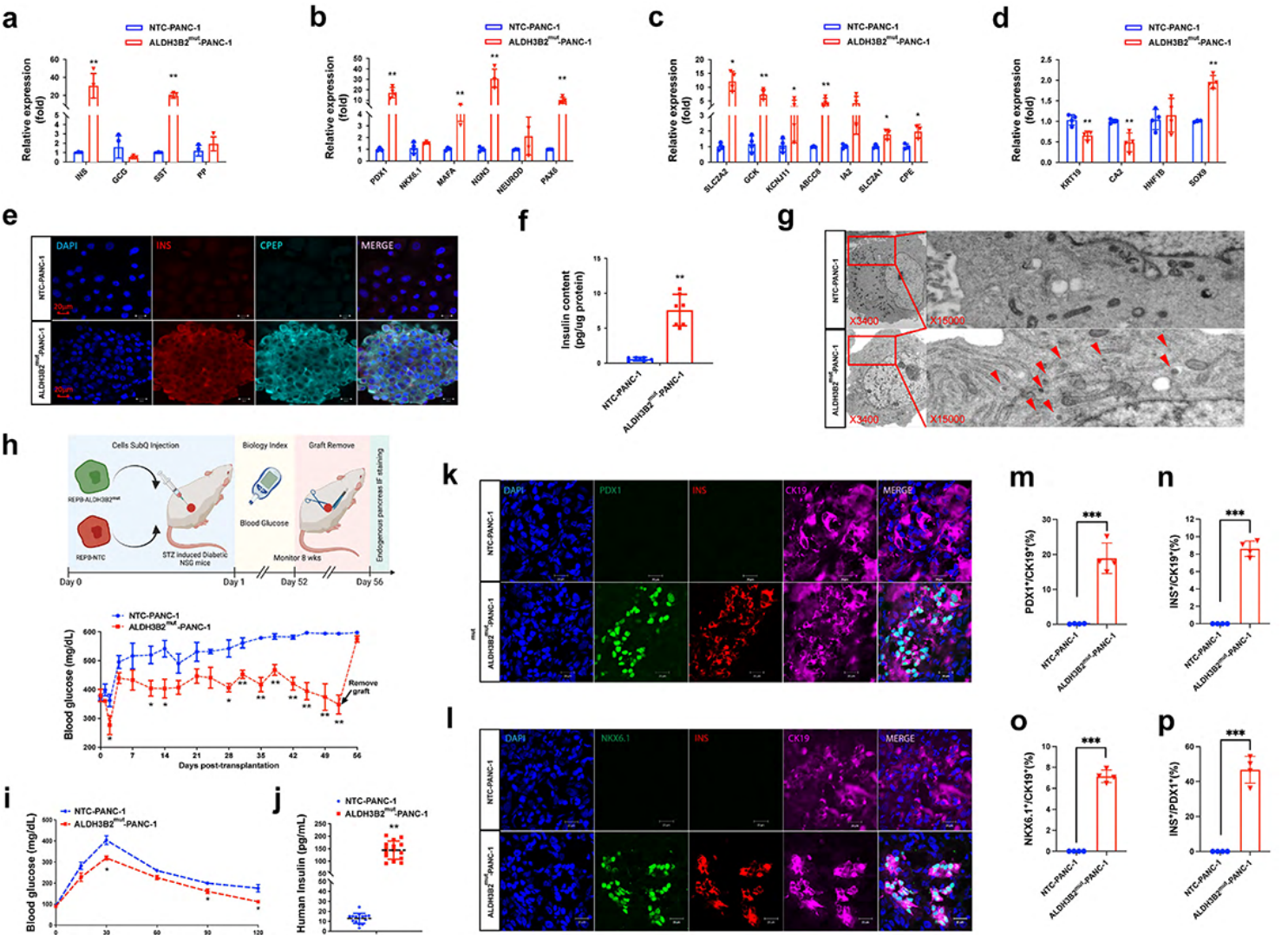


Fig. 2

Figure 2

Loss-of-function of ALDH3B2 trans-differentiates PANC-1 cells into insulin-producing beta-like cells.

a-d: qPCR analysis of human beta cell and duct signature genes. Data show mean $\pm$ SEM of n=4 biological replicates per condition and are representative of 2–3 independent experiments. *p < 0.05, **p < 0.01, calculated by two-way ANOVA with Sidak's multiple comparisons test. **a**, Genes related to pancreatic endocrine hormones. **b**, Genes related to beta cell key transcription factors. **c**, Genes related to beta cell function genes. **d**, Genes related to pancreatic duct cell markers. **e**, Insulin and C-peptide immunofluorescence in NTC cells and ALDH3B2^{mut}-PANC-1 cells. **f**: Insulin content in NTC cells and ALDH3B2^{mut}-PANC-1 cells were measured by insulin ELISA, normalized by cells genomic DNA content. n=7 biological replicates per condition and genotype. Data show mean $\pm$ SEM, **p < 0.01. **g**, Electron microscopy (EM) analysis of NTC-PANC-1 cells and ALDH3B2^{mut}-PANC-1 cells. Insulin secretion vesicles are indicated by arrow heads. **h**, An illustration of the in vivo function evaluation of trans-differentiated PANC-1 cells, and random blood glucose levels in the NTC and ALDH3B2^{mut}-PANC-1 cells transplanted diabetic NSG mice over 8 weeks of monitoring. Data show mean $\pm$ SEM of n=5 mice (NTC-PANC-1) and n=6 mice (ALDH3B2^{mut}-PANC-1) per group and are representative of three experiments, calculated by two-way ANOVA with Tukey's multiple comparisons test. **i**, An Intraperitoneal glucose tolerance test (IPGTT) in NTC or ALDH3B2^{mut}-PANC-1 cells transplanted NSG mice 6 weeks post-transplantation. **j**, Serum human insulin levels measurement in the NTC and ALDH3B2^{mut}-PANC-1 cells transplanted diabetic NSG mice. **k**, Pdx1 and Insulin immunofluorescence in transplanted NTC cells and ALDH3B2^{mut}-PANC-1 cells. **l**, Nkx6.1 and Insulin immunofluorescence in transplanted NTC cells and ALDH3B2^{mut}-PANC-1 cells. **m-o**, Percentage of Pdx1⁺ (**m**), Insulin⁺ (**n**), Nkx6.1⁺ (**o**) cells in all CK19⁺ cells. **p**, Percentage of Insulin⁺ cells in all Pdx1⁺ cells.

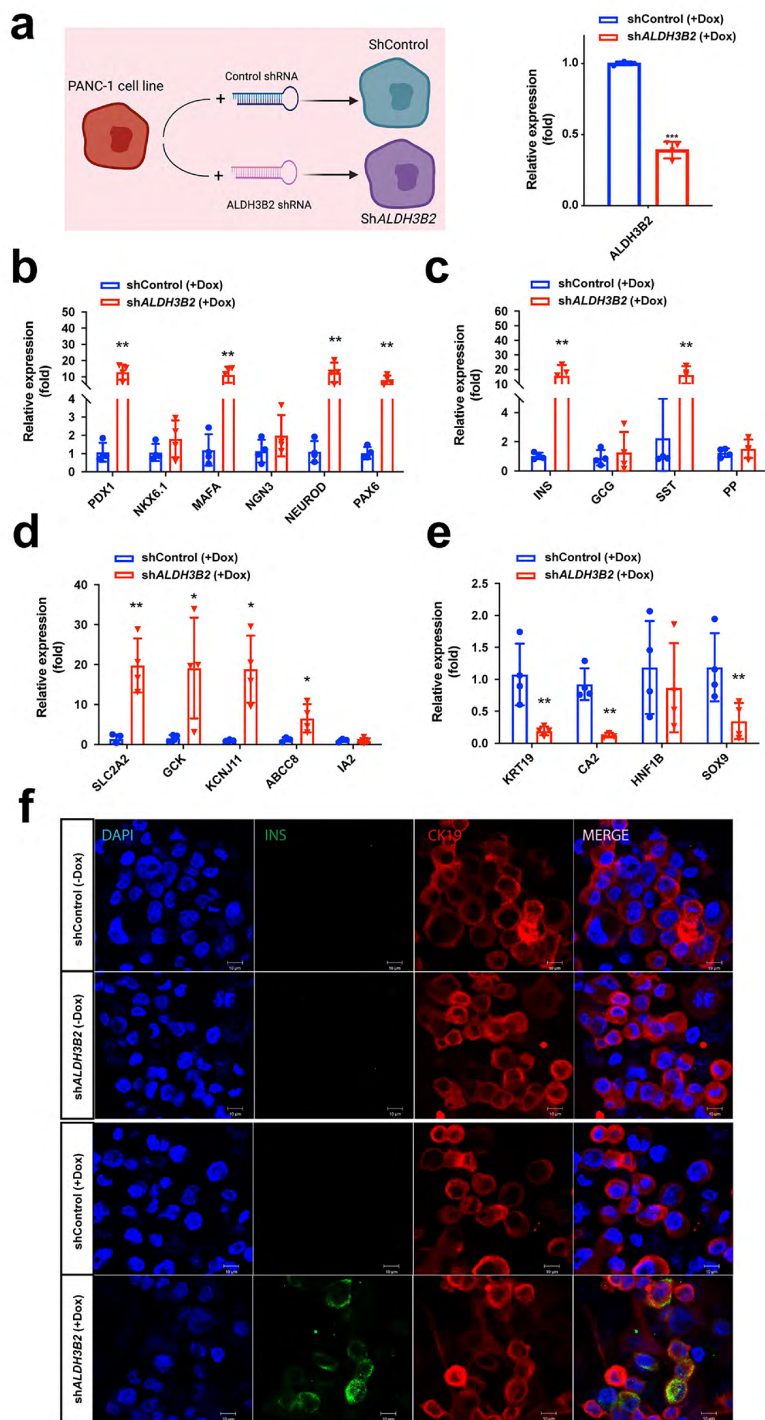


Fig. 3

Figure 3

Knock-down of ALDH3B2 by shRNA trans-differentiates PANC-1 cells into beta-like cells.

a, Generation of inducible non-targeting (shControl) cells and shALDH3B2-PANC-1 cells, and ALDH3B2 expression is significantly reduced in shALDH3B2 PANC-1 cells with Doxycycline treatment shown by qPCR analysis. **b-e**: qPCR analysis of human beta cell and duct signature genes. Data show mean $\pm$ SEM

of n=4 biological replicates per condition and are representative of 2–3 independent experiments. *p < 0.05, **p < 0.01, calculated by two-way ANOVA with Sidak's multiple comparisons test. **b**, Genes related to beta cell key transcription factors. **c**, Genes related to pancreatic endocrine hormones. **d**, Genes related to beta cell function genes. **e**, Genes related to pancreatic duct cell markers. **f**, Insulin and C-peptide immunofluorescence imaging in shControl and shALDH3B2 PANC-1 cells with or without Doxycycline treatment.

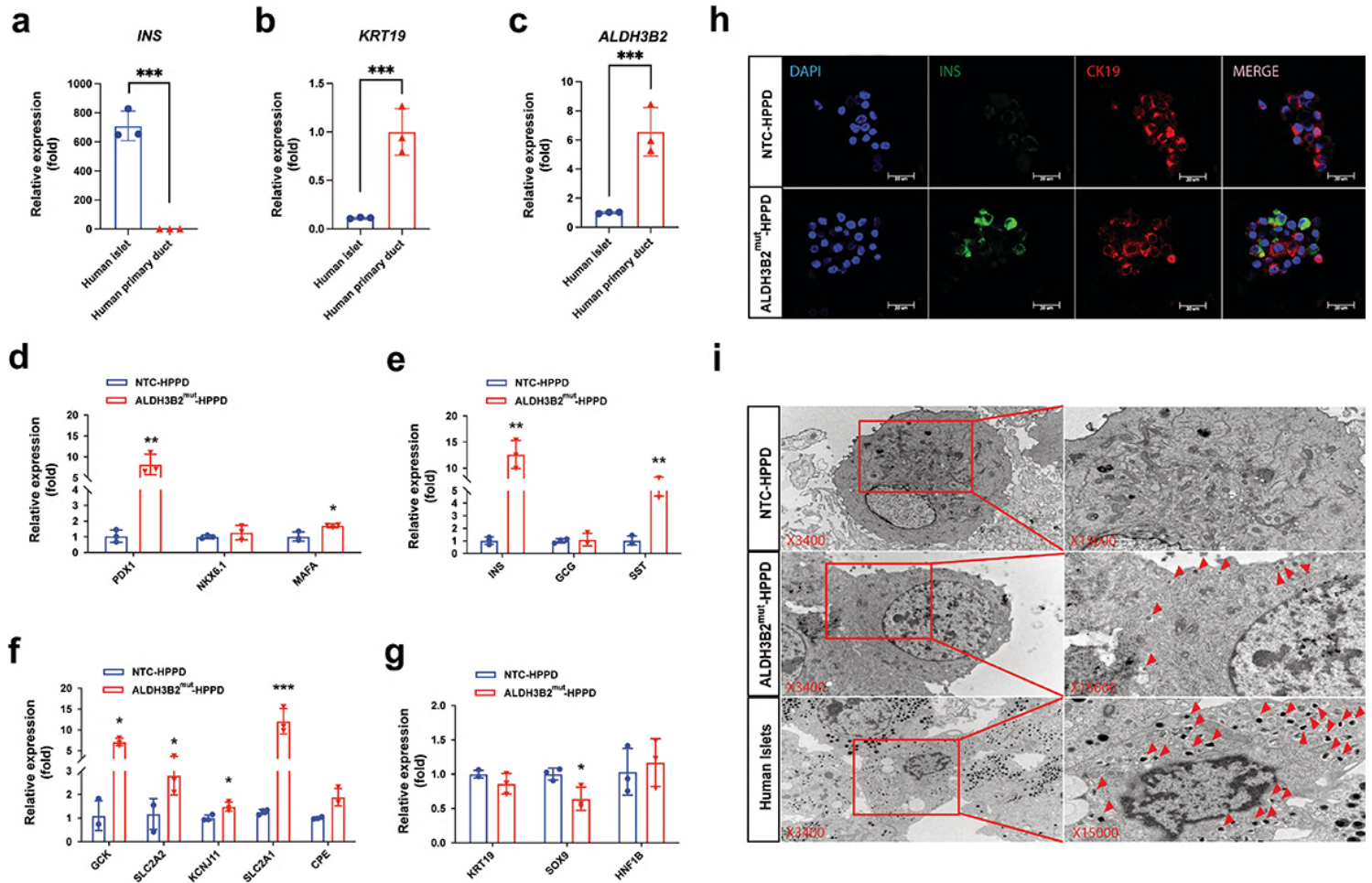


Fig. 4

Figure 4

Loss-of-function of ALDH3B2 trans-differentiates human primary pancreatic ductal cells into beta-like cells.

a-c: qPCR analysis of human Insulin (**a**), KRT19 (**b**) and ALDH3B2 (**c**) expression in purified human primary pancreatic duct cells (HPPD) and human primary islets. Data show mean $\pm$ SEM of n=3. ***p < 0.005. **d-g**, qPCR analysis of human beta cell and duct signature genes. Data show mean $\pm$ SEM of n=3 biological replicates per condition and are representative of 2 independent experiments. *P < 0.05, **P < 0.01, calculated by two-way ANOVA with Sidak's multiple comparisons test. **d**, Genes related to beta cell key transcription factors. **e**, Genes related to pancreatic endocrine hormones. **f**, Genes related to beta

cell function genes. **g**, Genes related to pancreatic duct cell markers. **h**, Insulin and CK19 immunofluorescence in NTC-HPPD and ALDH3B2^{mut}-HPPD cells. **i**, Electron microscopy (EM) analysis of NTC-HPPD cells, ALDH3B2^{mut}-HPPD cells and human primary islet cells. Arrow heads points to insulin secretion vesicles.

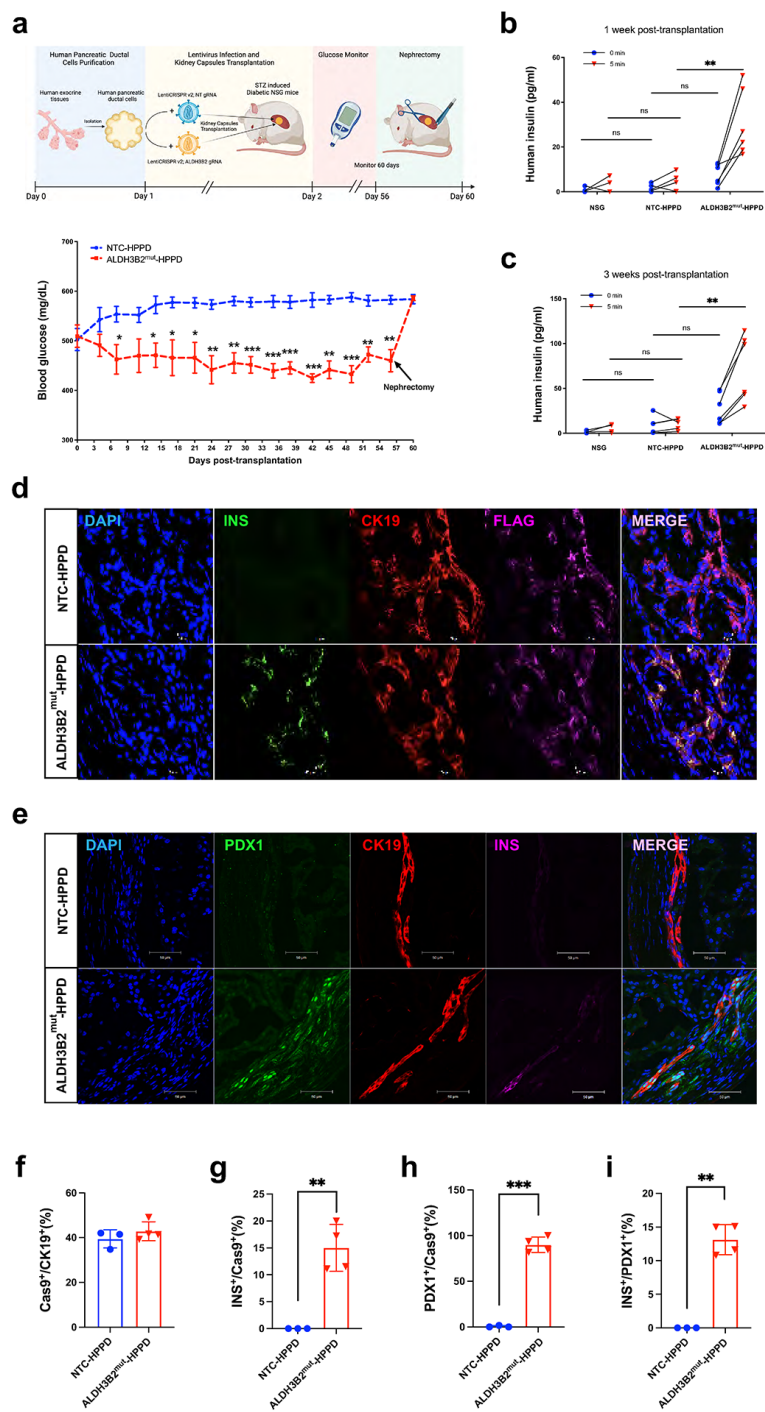


Fig. 5

Figure 5

ALDH3B2 mutation trans-differentiated human primary pancreatic duct cells are functional *in vivo*.

a, An illustration of the in vivo function evaluation of NTC-HPPD cells and ALDH3B2^{mut}-HPPD cells, and random blood glucose levels in the NTC-HPPD cells and ALDH3B2^{mut}-HPPD cells transplanted diabetic NSG mice for 60 days of monitoring. Data show mean $\pm$ SEM of n=5 mice per group, calculated by two-way ANOVA with Tukey's multiple comparisons test. **b and c**, Serum human insulin levels in non-transplanted NSG mice (Normal NSG), NTC-HPPD cells and ALDH3B2^{mut}-HPPD cells transplanted NSG mice 5 minutes after intraperitoneal (IP) glucose injection at 1 week post-transplantation (**b**) and 3 weeks post-transplantation (**c**). Calculated by two-way ANOVA with Sidak's multiple comparisons test. * NSG vs. NTC-HPPD; # NSG vs. ALDH3B2^{mut}-HPPD. **d**, Insulin, CK19 and Cas9 (Flag) immunofluorescence of transplanted NTC-HPPD and ALDH3B2^{mut}-HPPD cells. **e**, Pdx1, CK19 and Cas9 (Flag) immunofluorescence of transplanted NTC-HPPD and ALDH3B2^{mut}-HPPD cells. **f-i**, percentage of Cas9 lentivirus transduced cells in all transplanted duct cells (**f**), Insulin⁺ cells in all Cas9 lentivirus transduced cells (**g**), Pdx1⁺ cells in all Cas9 lentivirus transduced cells (**h**) and Insulin⁺ cells in all Pdx1⁺ cells (**i**).

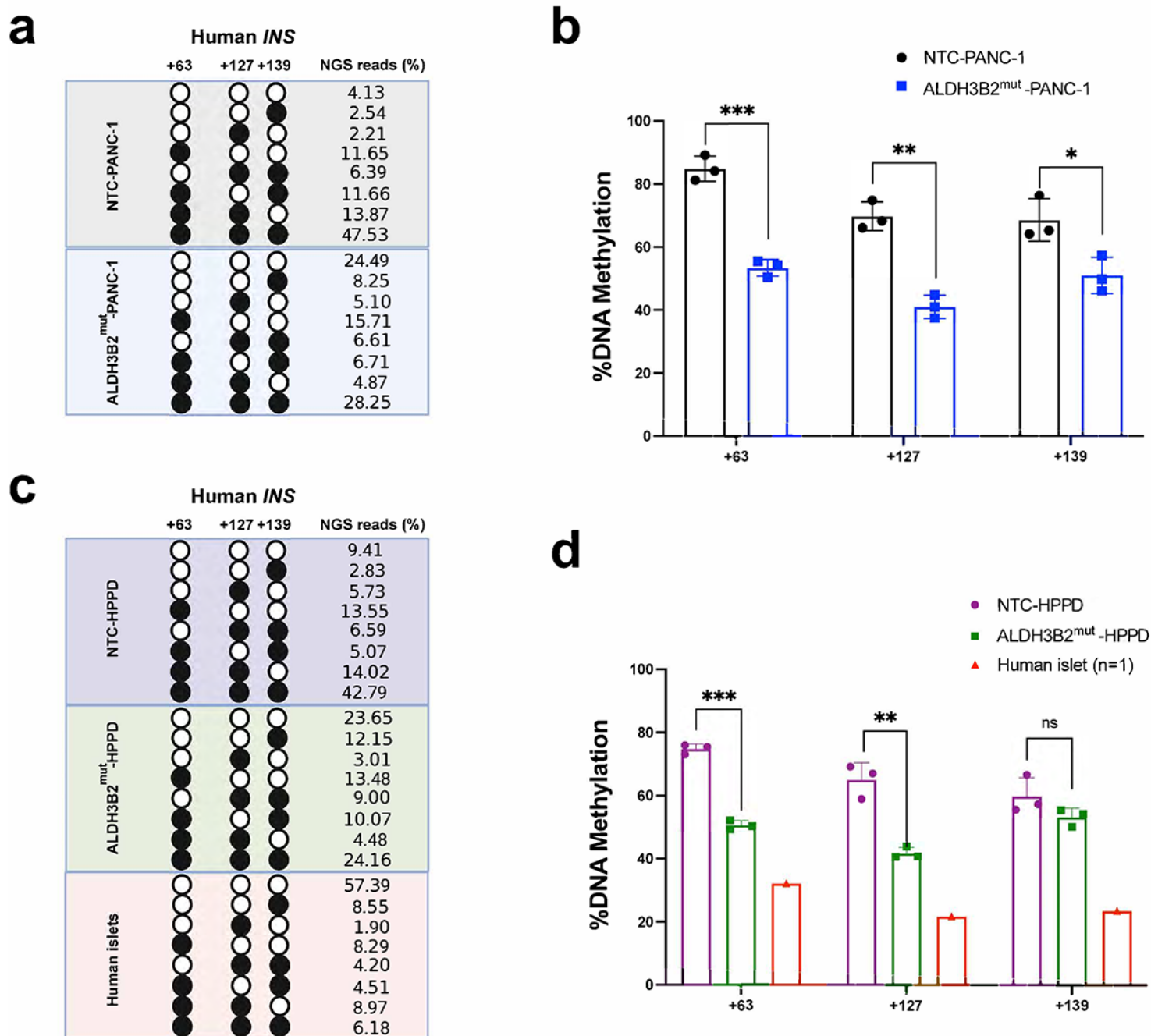


Fig. 6

Figure 6

ALDH3B2 loss-of-function in pancreatic duct cells leads to epigenetic changes.

a, Pattern and percentage of the DNA methylation at the position +63, +127 and +139 of the human insulin gene locus in NTC or ALDH3B2^{mut}-PANC-1 cells. **b**, quantification of the DNA methylation percentage in **a**. **c**, Pattern and percentage of the DNA methylation at the position +63, +127 and +139 of

the human insulin gene locus in HPPD-NTC, HPPD-ALDH3B2^{mut} and human primary pancreatic islet cells. **d**, quantification of the DNA methylation percentage in **c**.

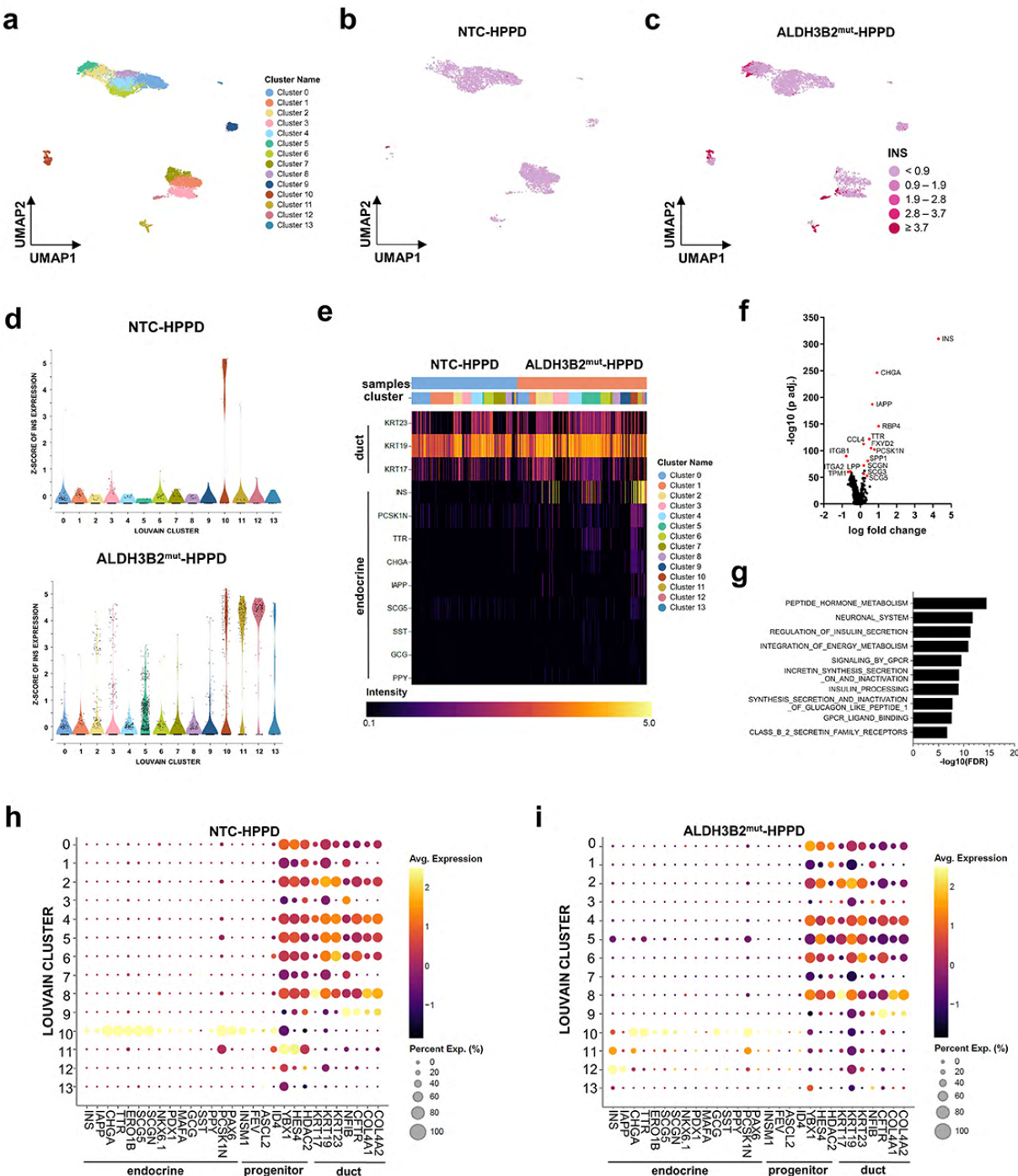


Fig. 7

Figure 7

Single cell RNA sequencing of ALDH3B2 mutant human primary pancreatic duct cells identified multiple clusters of insulin producing beta-like cells with overlapping duct cell gene expression profiles.

a-c: UMAP plots showing all identified cell cluster (**a**), insulin (INS) expression level of non-target control (NTC) (**b**) and ALDH3B2^{mut} (**c**) human primary pancreatic duct cells (HPPD). **d**, violin plot showing INS expression level of each individual cell grouped by cell cluster for NTC (left) and ALDH3B2^{mut} HPPD cells (right). **e**, Heatmap showing indicated endocrine cell and duct cell maker gene expression of control (blue) and ALDH3B2^{mut} (red) HPPD cells at single cell level grouped by cell cluster. **f**, volcano plot showing differential expressed genes in ALDH3B2^{mut} HPPD cells comparing cells with high INS expression (intensity >1) vs. low INS expression (intensity <1). **g**, Gene set enrichment analysis showing the upregulated Reactome gene sets of ALDH3B2^{mut} HPPD cells using the top 300 most significant upregulated genes of high INS-expressing (intensity >1) vs. low INS-expressing (intensity <1) cells. **h** and **i**, Dot plots showing average gene expression of indicated endocrine, progenitor, and duct cell marker genes in all identified cell cluster of NTC (**h**) or ALDH3B2^{mut} HPPD cells (**i**).

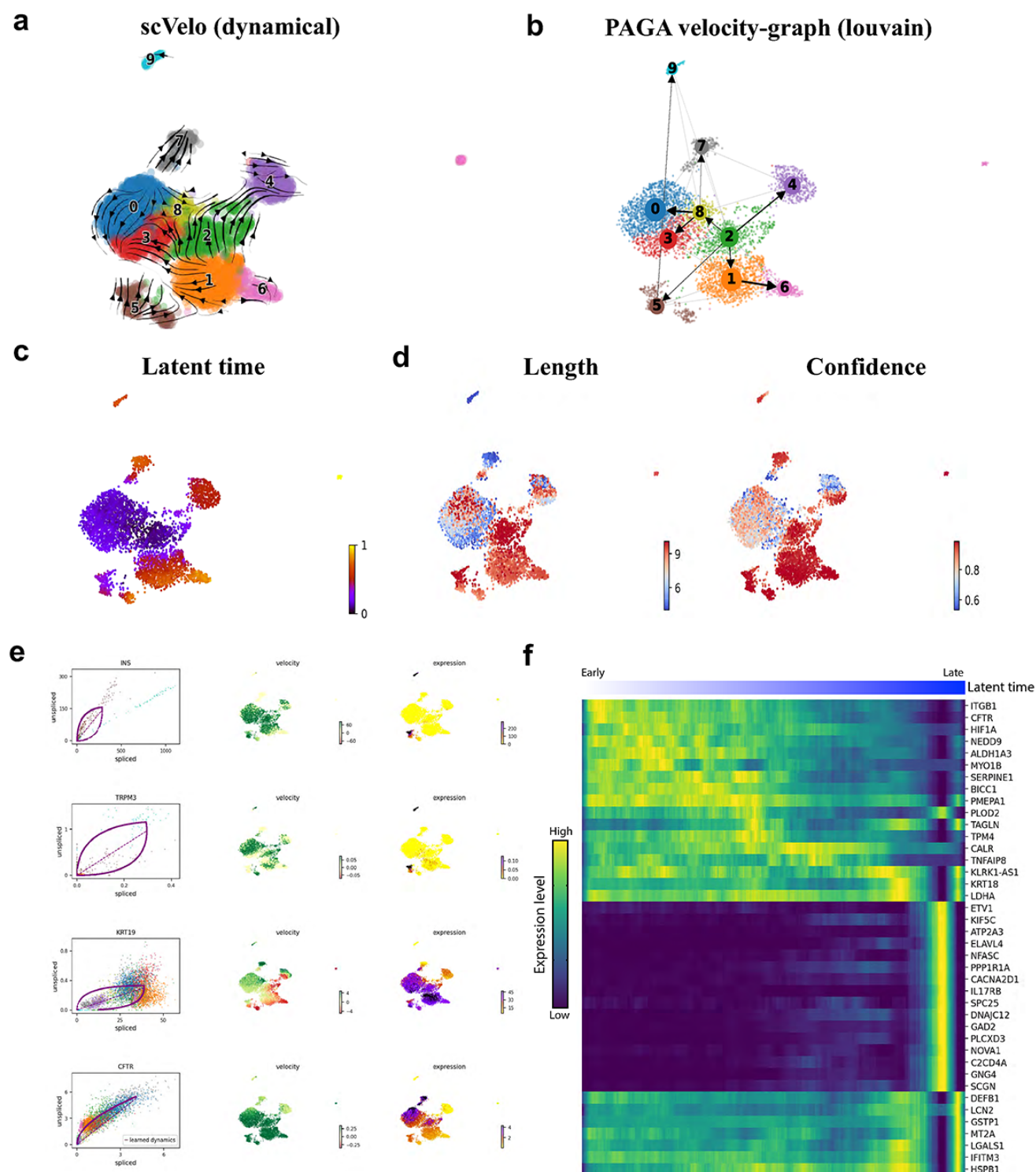


Fig. 8

Figure 8

Dynamic change in ALDH3B2 mutant human primary pancreatic duct cells transdifferentiation is revealed by RNA Velocity and PAGA Trajectory Analysis.

a. Velocity stream from scVelo analysis. The direction and magnitude of velocities are projected as arrows onto the Uniform Manifold Approximation and Projection (UMAP) plot of gene expression values

across cells. Arrows depict predicted fate trajectories. **b.** Velocity-based PAGA trajectory inference using scVelo dynamic model. The predicted trajectory reflects the developmental relations shown in **a.** **c.** Latent time plot (scale from black [0] to yellow [1] indicating initial and terminal states, respectively). **d.** The velocity-based length indicating rate of transition (left), and velocity confidence (right). **e.** Single gene velocity visualization with spliced/unspliced scatter plot (left), RNA velocity (mid), and gene expression (right) for INS, TRPM3, KRT19, and CFTR. **f.** Heatmap of velocity genes along INS-expressing cell development showing the means of spliced counts of the top-ranked dynamical genes in clusters 2, 5, and 9 for ALDH3B2^{mut}-HPPD. The columns were sorted by the latent time of cells in those clusters.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementalTable1.xlsx](#)
- [SupplementalTable2.pdf](#)