

RESEARCH ARTICLE

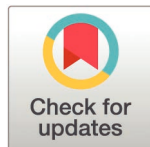
# PPDPF is not a key regulator of human pancreas development

Markus Breunig<sup>1\*</sup>, Meike Hohwieler<sup>1</sup>, Jasmin Haderspeck<sup>2</sup>, Felix von Zweyendorf<sup>3</sup>, Natalie Hauff<sup>1</sup>, Lino-Pascal Pasquini<sup>1</sup>, Christoph Wiegrefe<sup>4</sup>, Eleni Zimmer<sup>1</sup>, Medhanie A. Mulaw<sup>5</sup>, Cécile Julier<sup>6</sup>, Eric Simon<sup>7,8</sup>, Christian Johannes Gloeckner<sup>3,9</sup>, Stefan Liebau<sup>10,2†</sup>, Alexander Kleger<sup>1,10†\*\*</sup>

**1** Institute of Molecular Oncology and Stem Cell Biology (IMOS), Ulm University Hospital, Ulm, Germany, **2** Institute of Neuroanatomy & Developmental Biology (INDB), Eberhard Karls University Tübingen, Tübingen, Germany, **3** DZNE-German Center for Neurodegenerative Diseases, Tübingen, Germany, **4** Institute of Molecular and Cellular Anatomy, Ulm University, Ulm, Germany **5** Central Unit Single Cell Sequencing, Medical Faculty, Ulm University, Ulm, Germany, **6** Institut Cochin, Inserm U1016-CNRS UMR8104-Université Paris Descartes, Paris, France, **7** Cardio Metabolic Diseases Research, Boehringer Ingelheim Pharma GmbH & Co KG, Biberach, Germany, **8** Computational Biology & Genomics, Boehringer Ingelheim Pharma GmbH & Co KG, Biberach, Germany, **9** Institute for Ophthalmic Research, Eberhard Karls University Tübingen, Tübingen, Germany, **10** Division of Interdisciplinary Pancreatology, Department of Internal Medicine I, Ulm University Hospital, Ulm, Germany

† These authors jointly supervised the work.

\* [markus.breunig@uni-ulm.de](mailto:markus.breunig@uni-ulm.de) (MB); [alexander.kleger@uni-ulm.de](mailto:alexander.kleger@uni-ulm.de) (AK)



## OPEN ACCESS

**Citation:** Breunig M, Hohwieler M, Haderspeck J, von Zweyendorf F, Hauff N, Pasquini L-P, et al. (2025) PPDPF is not a key regulator of human pancreas development. PLoS Genet 21(4): e1011657. <https://doi.org/10.1371/journal.pgen.1011657>

**Editor:** Fengwei Yu, National University of Singapore, SINGAPORE

**Received:** July 8, 2024

**Accepted:** March 16, 2025

**Published:** April 7, 2025

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pgen.1011657>

**Copyright:** © 2025 Breunig et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## Abstract

Given their capability to differentiate into each cell type of the human body, human pluripotent stem cells (hPSCs) provide a unique platform for developmental studies. In the current study, we employed this cell system to understand the role of pancreatic progenitor differentiation and proliferation factor (PPDPF), a protein that has been little explored so far. While the zebrafish orthologue *exdpf* is essential for exocrine pancreas specification, its importance for mammalian and human development has not been studied yet. We implemented a four times CRISPR/Cas9 nicking approach to knockout *PPDPF* in human embryonic stem cells (hESCs) and differentiated *PPDPF*<sup>KO/KO</sup> and *PPDPF*<sup>WT/WT</sup> cells towards the pancreatic lineage. In contrast to data obtained from zebrafish, a very modest effect of the knockout was observed in the development of pancreatic progenitors *in vitro*, not affecting lineage specification upon orthotopic transplantation *in vivo*. The modest effect is in line with the finding that genetic variants near *PPDPF* are associated with random glucose levels in humans, but not with type 2 diabetes risk, supporting that dysregulation of this gene may only result in minor alterations of glycaemic balance in humans. In addition, PPDPF is less organ- and cell type specifically expressed in higher vertebrates and its so far reported functions appear highly context-dependent.

## Author summary

Human pluripotent stem cells (hPSCs) are capable of differentiating into nearly all cells of the human body and thereby serve in research as a powerful tool to study developmental processes in humans. In this study, we explored the role of *PPDPF*, a poorly

**Data availability statement:** The values used to build graphs are compiled in [S6 Table](#). Own RNA-seq data (FASTQ file) and pre-processed files (read count and R log transformed count matrix) are deposited at Gene Expression Omnibus (GSE282625). Publicly available data include ATAC-seq of CyT49 (GSE149148 [18]), ATAC-seq of HUES8 (GSE167606 [13]), histone ChIP-seq (H3K4me1, H3K27ac, and H3K4me3 (GSE149148) and transcription factor ChIP-seq. PP ChIP-seq of FOXA1, FOXA2, SOX9, GATA4, GATA6, PDX1 were retrieved from (GSE149148), NKX6-1 from (GSE167606), the differentiation time course of FOXA1, FOXA2 from (GSE54471 [17]) and murine E17.5 ChIP-seq from (GSE47459 [75]). Codes and data to run codes are also available under <https://github.com/MarkusBreunig1/PPDPF-is-not-a-key-regulator-of-human-pancreas-development>. Additional resources such as plasmids and generated cell lines are available from the corresponding authors on reasonable request, provided that national and international regulations (e.g. about the work with human embryonic stem cells) are met.

**Funding:** Main funding is provided to A.K. and S.L. by the Else Kröner Fresenius Foundation (EKFS) “2011\_A200”. A.K. was additionally supported by a Boehringer Ingelheim Ulm cooperation fund and by the DFG (K.L. 2544/7-1) and the Baden-Württemberg Stiftung GmbH („ISF-015 CrossingPanC“). S.L. received supportive funds by the DFG (LI 2044/5-1 and LI 2044/5-2), M.B. by the intramural Bausteinprogram of Ulm University (L.SBN.0207) and by the EKFS “2022\_EKEA.89” and C.J.G. by the Helmholtz-Gemeinschaft (iMed—the Helmholtz Initiative on Personalized Medicine). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

understood gene linked to pancreas development. While the zebrafish equivalent of this gene, *exdpf*, plays a critical role in forming the exocrine pancreas, its developmental function in mammals, including humans, had not been studied in detail. By using CRISPR/Cas9 technology, we “knocked out” the *PPDPF* gene in hPSCs and directed these cells to form pancreas-specific cells. Interestingly, unlike zebrafish, the “knock-out” of *PPDPF* was not essential for proper pancreatic cell type specification in human cells. Our findings underscore the importance of studying genes across different species to understand their evolutionary and functional nuances. The role of *PPDPF* appears to be more specific and critical in zebrafish, while in humans, it may rather act as a subtle regulator rather than a key driver. This less important role aligned with genetic data suggesting that variations in the *PPDPF* gene influence glucose levels but are not significantly associated with the risk to develop diabetes.

## Introduction

95% of the pancreas is composed of exocrine tissue with acinar glands producing and secreting digestive enzymes into a ductal network from which the pancreatic fluid is transported in the duodenum for processing of food. The remaining 5% of the pancreas is represented by endocrine islets of Langerhans, a specified tissue that is interspersed throughout the pancreas and is crucial for the regulation of blood glucose homeostasis. While physiological functions of the exocrine and endocrine pancreas are highly distinct, both compartments develop from a common ancestor, the multipotent pancreatic progenitor (PP) cells. Understanding embryonic and fetal pancreatic development, as well as the underlying molecular gatekeepers of exocrine and endocrine specification, is highly important for (i) studying causes of pancreatic agenesis, (ii) identifying genetic risk factors of diabetes, (iii) improving pancreatic *in vitro* differentiation strategies for regenerative medicine, and (iv) modelling endocrine and exocrine pancreatic diseases.

*Exdpf* has been identified in an RNA in situ hybridization screen as an essential regulator of exocrine pancreas development in zebrafish [1]. While the knockdown of *exdpf* nearly completely abolished exocrine cell development, its overexpression increased exocrine and reduced endocrine cell mass suggesting a critical and gatekeeping role during exocrine and endocrine lineage segregation [1]. Mechanistically, *exdpf* was found to act downstream of *ptf1a* and retinoic acid signalling [1]. Upon ablation, effectors of growth restriction such as *p21* were increased, and proliferation of exocrine cells was decreased [1]. This closely related function of *exdpf* for lineage specification and proliferation has drawn considerable attention to the human orthologue PPDPF for tumorigenesis. Indeed, several studies have revealed a pro-tumorigenic role of PPDPF in non-small cell lung cancer (NSCLC) [2–4], in ovarian cancer [5], in colorectal cancer [6] and in pancreatic ductal adenocarcinoma (PDAC) [7]. For hepatocellular carcinoma (HCC), findings were controversial. While a more recent study including explorative and mechanistical data found that decreased PPDPF expression correlated with poorer prognosis of HCC tumors [8], an initial HCC study postulated poorer survival in case of increased PPDPF expression [9]. In another study, PPDPF was utilized as part of a three-gene marker panel for biochemical recurrence of prostate cancer after treatment [10].

Studies in the zebrafish have greatly enhanced our understanding of endoderm development and organogenesis but species-specific differences have been also delineated lately [11]. The substantial degree of species conservation [1] together with the emerging literature on the role of PPDPF for tumor growth supported the hypothesis of a potentially important function of human PPDPF during pancreas development. Within this project, we now investigated

whether the human orthologue PPDPF indeed plays an important role for human pancreas development. Implementing a systematic and comprehensive analytical pipeline, we however demonstrate that PPDPF is not essential for proper specification of exocrine and endocrine tissue in a human pluripotent stem cell (PSC)-based differentiation model underpinning essential differences between the pancreas development of zebrafish and higher vertebrates.

## Results

### PPDPF is expressed in the pancreas and in other organs during mouse and human development

To investigate whether *Pdpf* is expressed in the pancreas during higher vertebrate development, we performed in situ hybridization (ISH) on mouse fetuses at E14.5. In addition to the pancreas, we detected similar high expression levels of *Pdpf* in other organs such as the stomach, hindgut, duodenum, and kidney (Fig 1A–C). A relatively broad expression pattern including additional organs such as dorsal root ganglion, umbilical artery, and urogenital sinus appears likely. ISH analysis on abdominal sections, however, did not allow mapping of *Pdpf* expression throughout all organs. We also enquired the human proteome map database [12], which revealed broad expression of PPDPF across multiple fetal and adult organs including the pancreas (Fig 1D). This clearly contrasts with *exdpf* zebrafish expression showing a more restricted organ-distribution to somites, liver, with the strongest signal in the pancreas [1]. Taken together, PPDPF is expressed in the developing and adult pancreas of mice and humans but with a broader expression pattern across multiple other organs than in zebrafish [1].

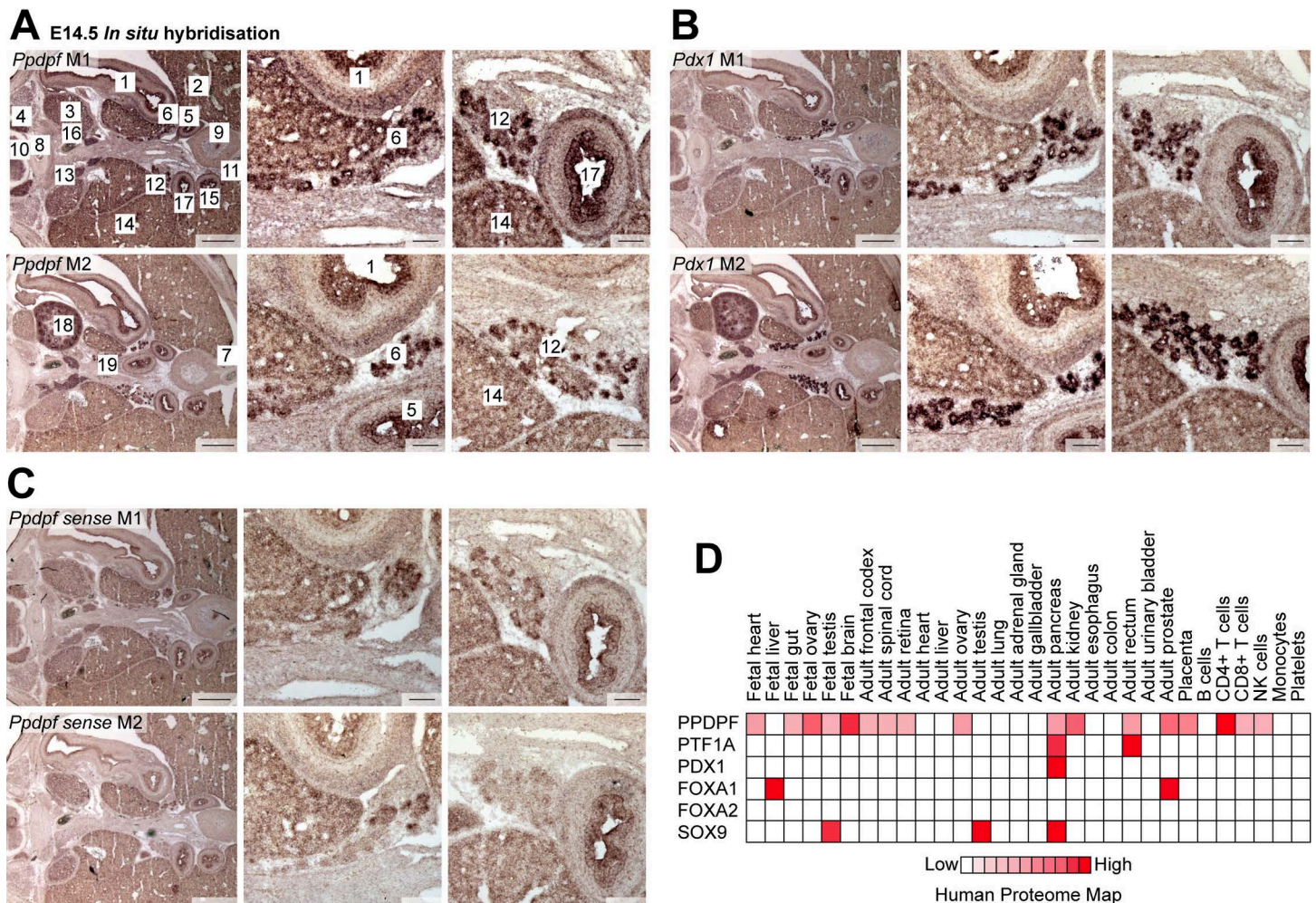
### PPDPF is expressed along pancreatic differentiation of hPSCs and is regulated by FOXA1/2 TFs

Our platform to differentiate hPSCs into pancreatic progenitor (PP) cells along definitive endoderm (DE), gut tube endoderm (GTE), and pancreatic endoderm (PE) allows the mimicry of and access to key stages of human pancreas development [13–16] (Fig 2A). While *PPDPF* mRNA expression gradually increased from hPSC to PP stage (Fig 2B), protein expression peaked at PE stage (Fig 2C). To understand how PPDPF expression is regulated, we examined publicly available ATAC-seq and ChIP-seq data of us [13] and others [17,18]. We found a gene regulatory promoter region (H3K4me3<sup>+</sup>) at the *PPDPF* locus with open chromatin as early as DE (Fig 2D) or PE in our data (S1A Fig). While the promoter was poised (H3K27ac<sup>-</sup>) at DE, it was active (H3K27ac<sup>+</sup>) at later pancreatic stages (Figs S1B and 2E). The promoter was bound by FOXA1/2 and SOX9 but not by transcription factors (TFs) being more specific to the pancreas such as PDX1 and NKX6-1 (Fig 2F). The endoderm-specifying TFs FOXA1/2 can regulate gene expression by either rendering the gene locus accessible at early stages for subsequent binding of tissue-specific TFs (pioneering role) or by being recruited themselves from tissue-specific TFs at later stages [19]. Here, FOXA1 and FOXA2 already bound to the *PPDPF* promoter at DE and GTE stages, respectively (Fig 2G), in line with a pioneering role and its broader gene expression profile across several tissues (Fig 1D). ChIP-seq data at a later time point of murine development confirmed *Foxa2* binding, while acinar-specific TFs, *Ptf1a* and *Rbpjl*, (additionally) bound a more downstream region (S1C Fig) that might act as putative enhancer.

### Genetic variants near PPDPF are associated with glycaemic control

Since several genes associated with diabetes are involved in PP development, we explored publicly available genetic data for *PPDPF* association with metabolic traits. We explored the *PPDPF* region in HugeAMP [20] and found a credible set of 4 single nucleotide

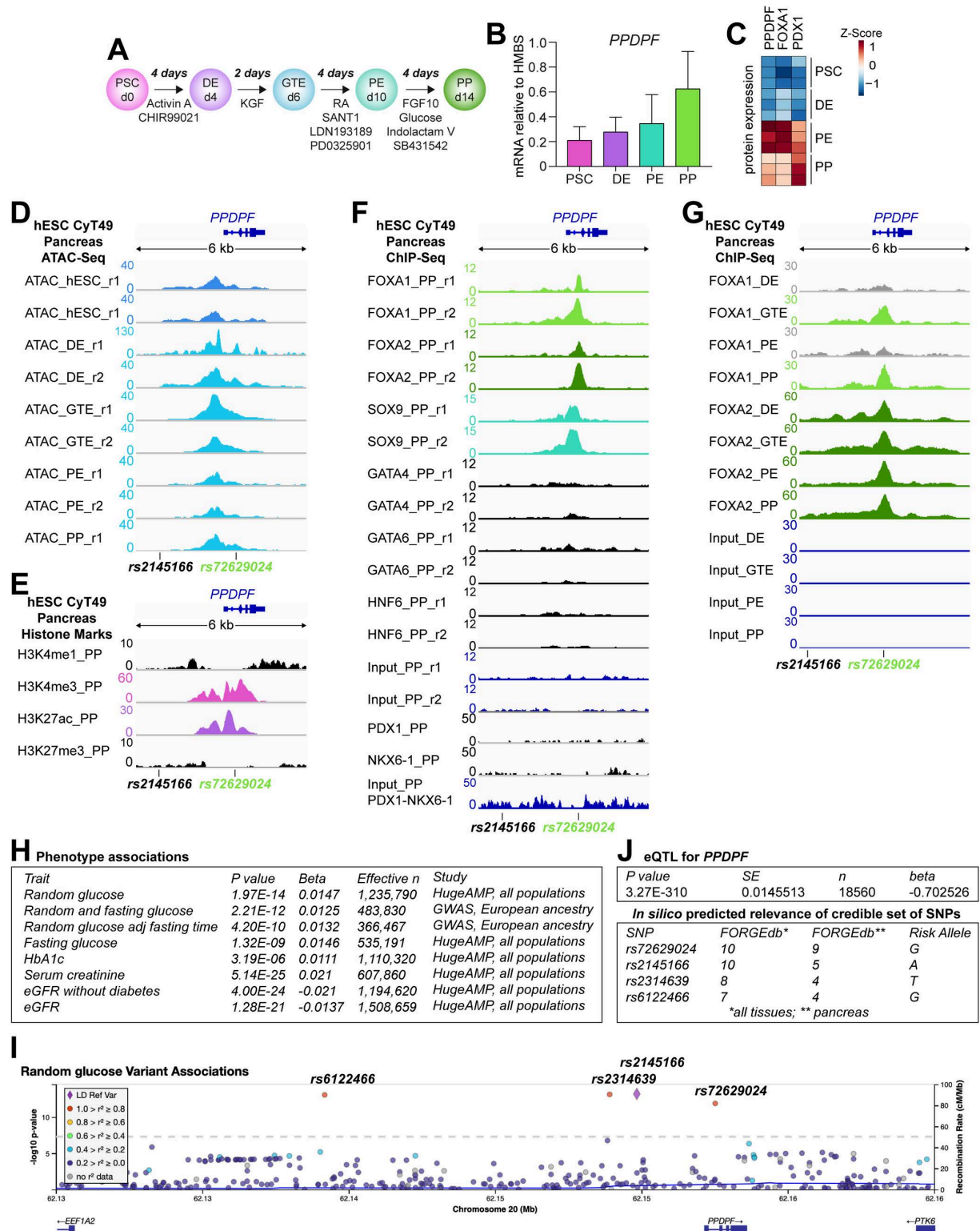




**Fig 1. PPDPF is expressed in the pancreas and in other organs during mouse and human development.** (A–C) *In situ* hybridization on E14.5 mouse fetuses in transverse plane against (A) *Pdpdf*, (B) a positive *Pdx1*, and (C) a negative *Pdpdf* sense control probe. M1: mouse 1; M2: mouse 2. [1] stomach; 2) left lobe of liver; 3) left adrenal primordium; 4) dorsal root ganglion; 5) hindgut; 6) primordium of the left lobe of the pancreas; 7) left umbilical artery; 8) notochord; 9) urogenital sinus; 10) spinal cord; 11) right umbilical cord; 12) primordium of right lobe of pancreas; 13) right adrenal primordium; 14) right lobe of liver; 15) descending part of duodenum; 16) aorta; 17) ascending part of duodenum; 18) left kidney; 19) portal vein). (D) PPDPF expression according to the human proteome map database [12] plotted against PTF1A, PDX1, FOXA1, FOXA2, and SOX9 expression profiles.

<https://doi.org/10.1371/journal.pgen.1011657.g001>

polymorphisms (SNPs) being associated with increased random glucose (Fig 2H and 2I and S1 Table). The same set was recently reported in a genome-wide association study (GWAS) of random glucose [21]. In addition, traits indicative of renal dysfunction such as increased serum creatine and decreased estimated glomerular filtration rates (eGFR), and other glycaemic traits such as increased fasting glucose and HbA1C levels were significantly associated with the identified SNPs (Fig 2H and S1 Table). Of note, the presence of the affect alleles is associated with decreased PPDPF expression, as measured by cis-expression quantitative trait locus (cis-eQTLs [22]) analysis in blood (Fig 2J and S1 Table). One of these 4 SNPs, rs72629024, is located in the first intron of PPDPF in a binding region of FOXA1/2 and SOX9 (Figs 2D–G, S1A and S1B). Of note, the homologous intron 1 region in zebrafish has been recognized as an important gene regulatory region for ptf1a binding [1] and FORGEDb analysis predicted that this specific SNP might be more important than the other three SNPs in the



**Fig 2. Regulation of PPDF expression in human in vitro pancreatic differentiation and correlation to human diseases.** (A) Schematic overview of pancreatic *in vitro* differentiation from human pluripotent stem cells (hPSCs) via definitive endoderm (DE) over gut tube endoderm (GTE) and pancreatic endoderm (PE) into pancreatic progenitors (PP). (B) PPDF mRNA expression relative to HMBS during *in vitro* differentiation of the human embryonic stem cell (hESC) line HUES8 (n = 4). (C) Protein expression during differentiation based on full proteome measurements [23]. (D) Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq)-based chromatin opening at PPDF locus from



publicly available *in vitro* differentiation data of the hESC line CyT49 [18]. (E) Histon marks indicating an active promoter region at PP stage in the same CyT49 differentiation. (F) Chromatin immunoprecipitation (ChIP)-seq binding peaks of key transcription factors (TFs) at PP stage of CyT49 differentiation [13,18]. (G) Time course of ChIP-seq FOXA1/2 binding peaks in CyT49 differentiation [17]. ATAC- and ChIP-seq peaks have been visualized via Integrative Genomics Viewer (IGV) [54]. (H) Significant associations of the lead SNP rs2145166 with phenotypic traits. Association data was extracted from the common metabolic diseases knowledge portal of HugeAMP [20]. (I) cis-eQTLs [22] correlation analysis between lead SNP and PPDPF expression (top). eQTLs: expression quantitative trait locus. *In silico* prediction of the relevance of individual SNPs of the credible set (bottom) using FORGEdb [55]. Localization of the two closer SNPs is displayed in E-H. Note that no chromatin or ChIP-seq peak was found close to the two more distant, not displayed, SNPs. (J) Exemplary plot of random glucose trait for lead SNP rs2145166 with linkage disequilibrium (LD) analysis emphasizing that individual effects of the 4 SNPs in the credible set cannot be distinguished (color code) and that these SNPs are the only SNPs close to the PPDPF locus that are associated with glycaemic traits.

<https://doi.org/10.1371/journal.pgen.1011657.g002>

pancreas (Fig 2J), while linkage disequilibrium analysis could not finally delineate the individual contribution of the 4 SNPs to the actual phenotypic traits (Fig 2I and S1 Table).

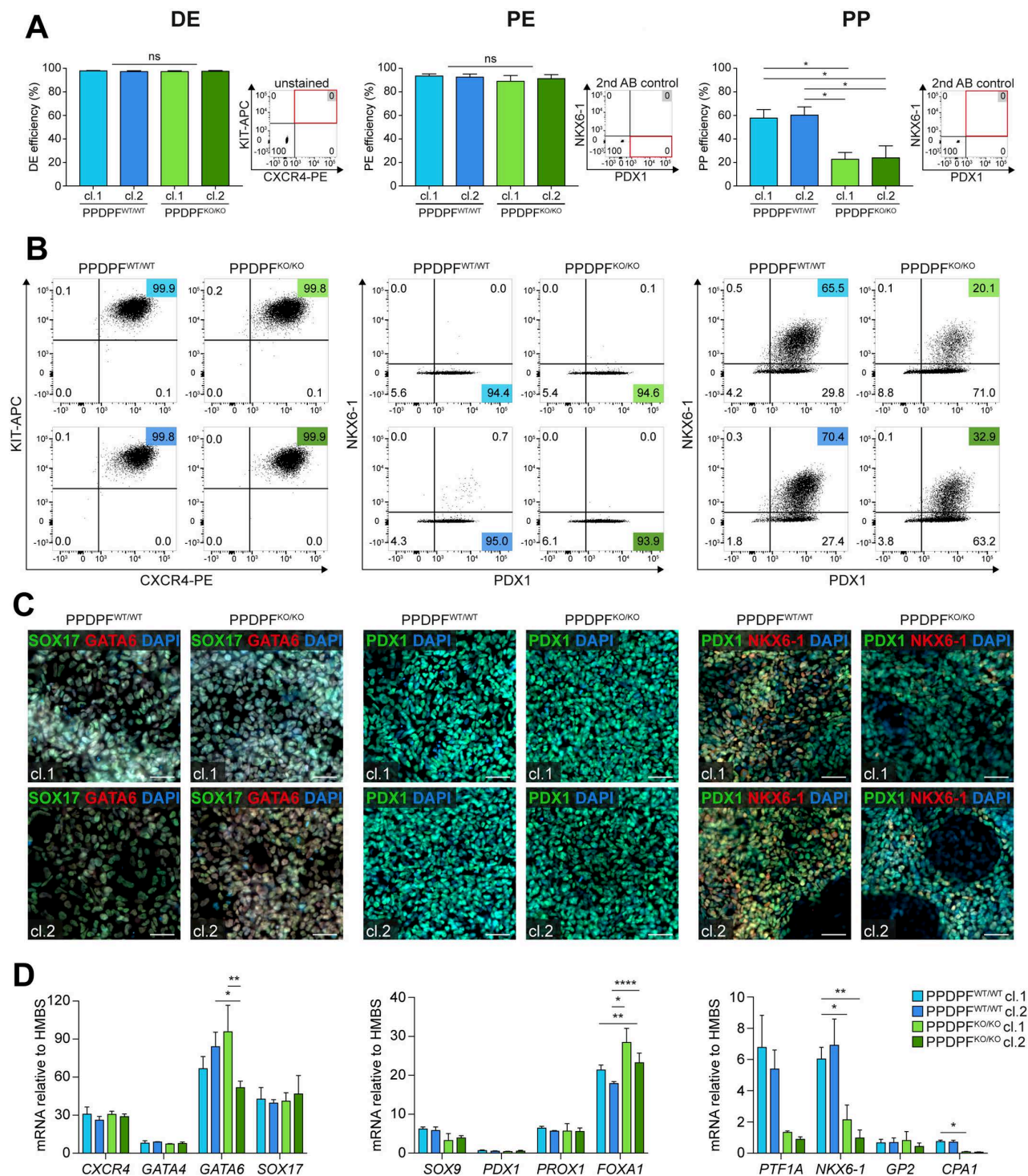
### PPDPF-deficient hESCs differentiate into PPs with reduced efficiency

To study the impact of PPDPF during pancreatic differentiation, we employed a four times CRISPR/Cas9 nicking strategy in the human embryonic stem cell (hESC) line HUES8 (S2A Fig). Based on PCR and sequencing (S2B and S2C Fig), we selected two clones harbouring a homozygous deletion of 63 bps of exon3 in the coding sequence. qPCR and Western Blotting revealed loss of transcript and protein expression of the PPDPF<sup>p.R19\_I40delInsL/p.R19\_I40delInsL</sup> clones (S2D-F Fig). Expression of pluripotency marker remained unaltered upon gene editing (S2G-H Fig). The two clones are from now on referred to as PPDPF<sup>KO/KO</sup> cl. 1 and cl. 2. While PPDPF-deficient hESCs differentiated into KIT<sup>+</sup>/CXCR4<sup>+</sup> DE and PDX1<sup>+</sup> PE with similar efficiencies as their wildtype (WT) controls, the capacity to generate PDX1<sup>+</sup>/NKX6-1<sup>+</sup> PPs was significantly diminished upon ablation of PPDPF (Fig 3A and 3B). Flow cytometry results could be confirmed by immunofluorescence staining (Fig 3C) and a panel of additional qPCR marker including GATA4 and GATA6 at DE stage, SOX9, PROX1, and FOXA1 at PE stage and PTF1A, GP2, and CPA1 at PP stage (Fig 3D).

### Identification of interaction partners not including pancreatic TFs

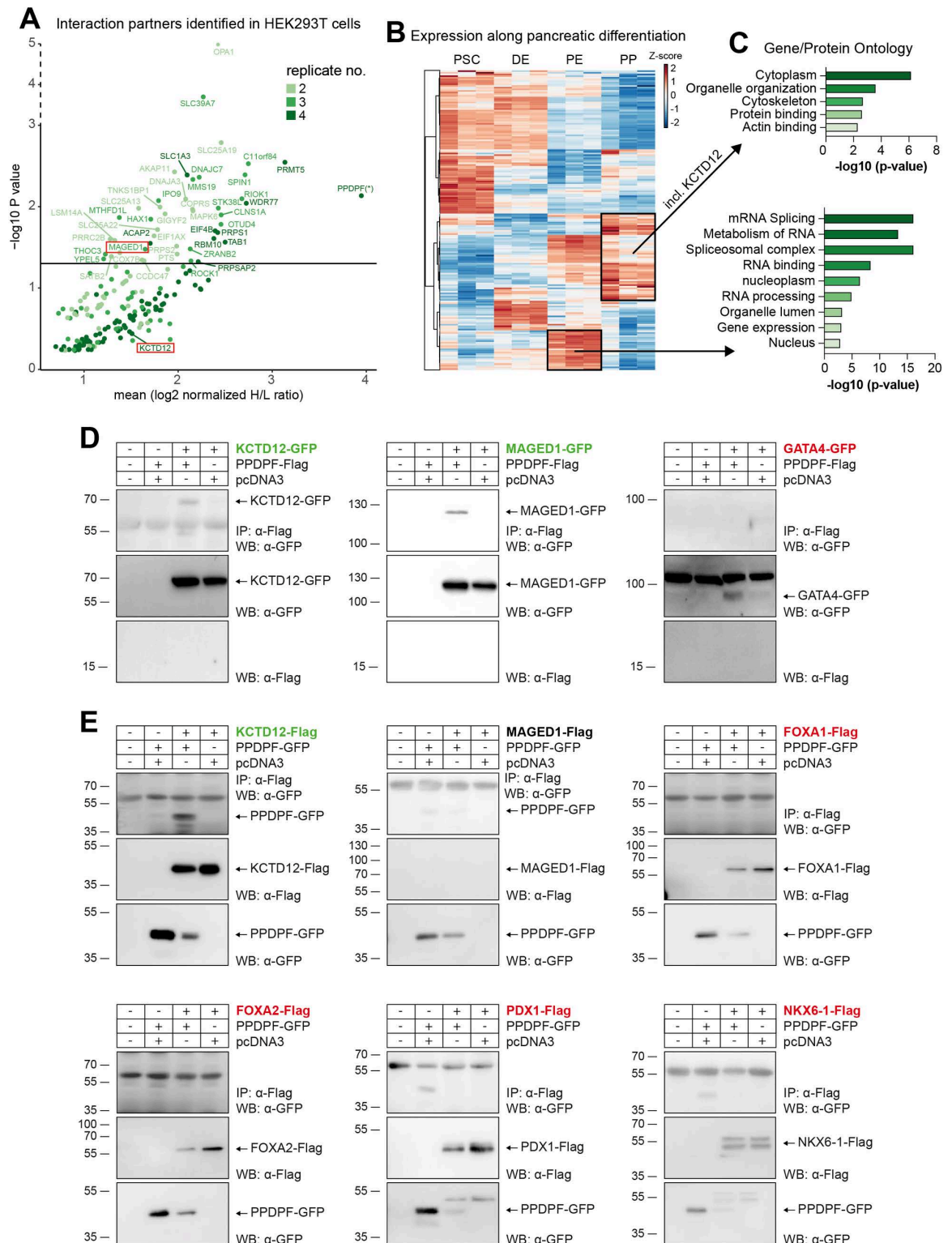
To further dissect this observation, we screened for potential interaction partners of PPDPF using a PPDPF-Flag overexpression SILAC-screen (stable isotope labeling by amino acids in cell culture) in HEK293T cells followed by mass spectrometry. Indeed, we detected 191 PPDPF-bound proteins from which 45 reached significance across four experiments (Fig 4A and S2 Table). Next, we clustered the list of putative interaction partners along a proteome time course of pancreatic *in vitro* differentiation [23] to obtain a context-relevant protein set followed by overrepresentation analysis of PE- and PP-specific proteins. While PE-specific interaction partners correlated with RNA binding (18 from 22 in query), nucleus [21], and mRNA splicing terms [14], PP-specific interaction partners were linked to protein binding (48 from 49 in query), cytoplasm (46 from 49), and cytoskeleton (17 from 49) terms (Fig 4B and 4C). We proved the validity of the identified interaction partners by exemplary validation of one significant hit (MAGED1) and one non-significant hit (KCTD12) by co-immunoprecipitation experiments and Western Blotting (Fig 4D and 4E).

Due to the prominent role of the zebrafish orthologue *exdpf* for pancreatic development, we had hypothesized a central function of PPDPF as signaling protein in the TF circuitry of the pancreas. Since no pancreas-specific TF could be identified in our HEK293T cell screen, we additionally analyzed the interaction of PPDPF to the pioneering TFs GATA4, FOXA1, and FOXA2 and the key pancreas determining TFs PDX1 and NKX6-1 but did not detect binding to any of the investigated TFs (Fig 4D and 4E).



**Fig 3. PPDPF-deficient hPSCs differentiate into PPs with reduced efficiency.** (A) Differentiation efficiencies of clonal PPDPF<sup>WT/WT</sup> and PPDPF<sup>KO/KO</sup> hPSC lines into DE, PE, or PP cells, measured by flow cytometry (DE = 6, PE = 4, PP: n = 5 (independent differentiations), in duplicates (two wells per differentiation); Mean  $\pm$  SEM; ordinary one-way ANOVA, not assuming sphericity, followed by Tukey's multiple comparison test). (B) Corresponding flow cytometry plots. (C) Immunocytochemistry/immunofluorescence staining confirming flow cytometry results at DE (left), PE (middle), and PP stage (right panel) (Scale bar: 50  $\mu$ m). (D) Stage-specific (DE, PE, PP) mRNA expression relative to HMBS (n = 3; Mean  $\pm$  SEM; ordinary two-way ANOVA, not assuming sphericity, followed by Tukey's multiple comparison test).

<https://doi.org/10.1371/journal.pgen.1011657.g003>



**Fig 4. Identification of interaction partners not including pancreatic TFs.** (A) Potential PPDPF interaction partners identified by immunoprecipitation experiments. PPDPF overexpression in HEK293T cells was followed by stable isotope labeling by amino acids in cell culture (SILAC) and mass spectrometry. H/L ratio: heavy/light isotope ratio. P-value of 0.05 is indicated by a horizontal line;  $n = 4$ , with color code illustrating in how many screens individual proteins were detected. (B) Expression of potential interaction partners is plotted along pancreatic differentiation using our previous proteome data [23]. Ward.D2 clustering identified PE and PE/PP-specific



proteins (C) that were applied to gProfiler overrepresentation analysis. (D) Anti-Flag co-immunoprecipitation blots after overexpression of PPDPF-Flag bait and either KCTD12-, MAGED1, or GATA4-GFP in HEK293T cells. (E) As in (D) but PPDPF-GFP was overexpressed together with either KCTD12-, MAGED1-, FOXA1-, FOXA2-, PDX1-, or NKX6-1-Flag.

<https://doi.org/10.1371/journal.pgen.1011657.g004>

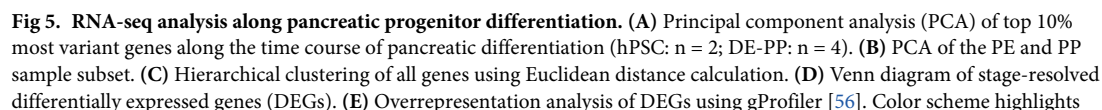
## RNA-seq analysis suggests incomplete PP differentiation and tissue bifurcation upon PPDPF ablation

To globally map developmental consequences of PPDPF ablation, RNA-seq experiments were performed from 2 PPDPF<sup>WT/WT</sup> and 2 PPDPF<sup>KO/KO</sup> clones at hPSC, DE, PE, and PP stage from four independent differentiations. First, we utilized global transcriptome data to check for potential off-target effects causing strong transcript expression changes after gene editing but found none in CRISPOR [24] predicted risk candidates (S3A-E Fig). RNA-seq samples clustered according to the developmental time point in principal component analysis (PCA) of the 10% most variant genes (Fig 5A), with a subtle separation of PPDPF<sup>WT/WT</sup> and PPDPF<sup>KO/KO</sup> samples at PE and PP stage (Fig 5B). In hierarchical clustering, 4 of 8 KO samples clustered closer to PE samples than to the other PP samples suggesting an incomplete differentiation into PPs upon PPDPF ablation (Fig 5C).

While very few genes were deregulated at hPSC and DE stage, around 300 differentially expressed genes (DEGs) were found at PP stage (Fig 5D and S3 Table). GO-term analysis of DEGs upregulated in PPDPF<sup>KO/KO</sup> samples pointed to a bifurcation towards non-pancreatic tissues with gene terms for gut enterocytes and lymphoid cells being overrepresented (Fig 5E and S4 Table). In addition, terms for pancreatic precursor cells and ECM-receptor interaction were downregulated (Fig 5E and S4 Table). Venn diagrams, depicting the top 10 DEGs, illustrated that several of these were altered at more than one developmental stage (Fig 5F). Among them the lncRNA H19 that was found to have diverse roles in proliferation and tumorigenesis by sponging microRNAs [25] was for example downregulated, while ZIM2 and PEG3 that can also regulate proliferation and tumorigenesis but also fetal growth [26] were upregulated in PPDPF<sup>KO/KO</sup> cells. We did however not detect specific pathways to be globally affected by loss of PPDPF. Although some pathways such as E2F targets and G2M checkpoints were upregulated in PPDPF<sup>KO/KO</sup> samples at PE stage (Fig 5G), (i) alterations did not persevere to PP stage (Fig 5G), and (ii) were likely rather caused by fluctuations of single samples instead of a consistent genotype-specific pattern (Fig 5H).

## Transcriptomic changes between genotypes are relatively small

To approach a potential developmental defect more specifically, we compiled tailored gene sets for hPSC-based pancreatic *in vitro* differentiations [23,27], human fetal tissue [28–30], and Ptf1a targets [31] (S5 Table). Gene set enrichment analysis substantiated our previous findings that PPDPF<sup>KO/KO</sup> PE/PP cells were depleted for a pancreatic progenitor identity while they were enriched for an endodermal identity (Fig 6A). Conclusively, genes that were differentially upregulated upon Ptf1a<sup>KO/KO</sup> in mice, were depleted in PPDPF<sup>KO/KO</sup> cells (S4A Fig). In addition, human fetal pancreatic signatures including the more mature tip and trunk states were depleted (S4B and S4C Fig), while several non-pancreatic tissue signatures such as intestine and placenta were enriched (S4D Fig). Related to latter, several pregnancy-specific glycoproteins (PSGs) were significantly upregulated in PPDPF<sup>KO/KO</sup> PE/PP samples, but their actual transcript levels remained extremely low (S4D Fig and S3 Table). Heatmap illustration of enriched and depleted gene sets, however, revealed that also PPDPF<sup>KO/KO</sup> cells have successfully acquired stage-specific signatures including a pancreatic progenitor expression profile and that differences between stages are much higher than between genotypes (Fig 6C and 6D).



differentiation stage (based on legend in C). Gene sets in top panel were enriched in PPDPF<sup>KO/KO</sup> cells, bottom panel were depleted. (F) Volcano plots of stage-specific PPDPF<sup>KO/KO</sup> versus PPDPF<sup>WT/WT</sup> gene expression comparison. DEGs that are up in PPDPF<sup>KO/KO</sup> samples are highlighted in green, genes that are down in blue. Top ten DEGs with highest FC change are labelled. (G) Gene set enrichment analysis (GSEA) against hallmark pathways. False discovery rates (FDRs)  $\leq 0.25$  are highlighted in red (left panel), while comparisons that are enriched in PPDPF<sup>KO/KO</sup> cells are depicted in green and comparisons that are depleted in blue (right panel). Additional information: For “G2M checkpoint”, 0 out of 197 genes were significantly upregulated (DEGs) in PPDPF<sup>KO/KO</sup> PPs over PPDPF<sup>WT/WT</sup> PPs, while 1 was downregulated. For “E2F targets” 0 out of 200 genes were up or down. (H) Heatmap illustrating row/gene-scaled gene expression at PE stage of two selected hallmark pathways. PSC: pluripotent stem cell; DE: definitive endoderm; PE: pancreatic endoderm; PP: pancreatic progenitor.

<https://doi.org/10.1371/journal.pgen.1011657.g005>

Indeed, only 19 of 421 genes within the “Definitive endoderm [23]” gene set were significantly upregulated (DEGs) in PPDPF<sup>KO/KO</sup> PPs over PPDPF<sup>WT/WT</sup> PPs, while 5 were downregulated (P-value < 0.01 and FC > 1.1). At the same time 5 of 451 genes within the “Pancreatic progenitor [23]” gene set were upregulated (DEGs) and only 15 downregulated in PPDPF<sup>KO/KO</sup> PPs (S3 and S5 Tables). Time-resolved gene expression profiles of key endoderm (Fig 6E) and pancreatic progenitor genes (Fig 6F) further highlights that transcriptomic changes were overall rather small.

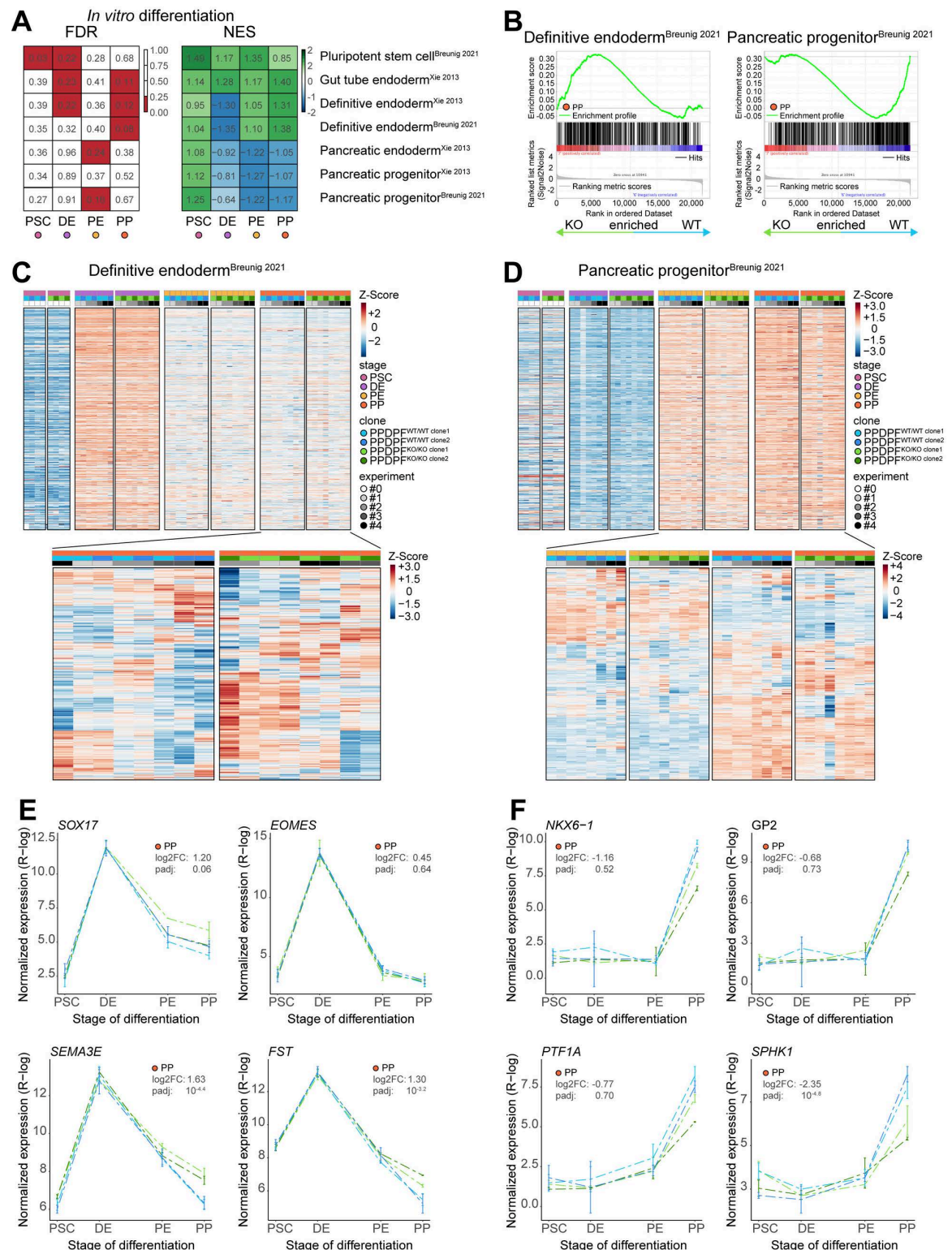
### No alteration in proliferation and growth characteristics of PPDPF<sup>KO/KO</sup> PPs

Since *exdpf* plays a role in the proliferation of exocrine cells in zebrafish [1], we specifically checked the effect of PPDPF ablation on proliferation and growth characteristics of pancreatic precursors. The total number of cells during pancreatic *in vitro* differentiation was however not reduced in PPDPF<sup>KO/KO</sup> cells (Fig 7A). Also, the rate of proliferation and the cell cycle distribution was not significantly reduced (Fig 7B–D). A slight trend of a reduction in S-phase was visible in PPDPF<sup>KO/KO</sup> PPs but did not manifest in changes in absolute cell numbers (Fig 7A). In addition, the number of cleaved CASP3<sup>+</sup> or Annexin V<sup>+</sup> cells was not altered (Fig 7E and 7F), and also global RNA-seq analysis did not suggest a change in growth- and cell death-related programs (Figs 7H, 5G and 5H). In summary, our data does not support a major impact of PPDPF on proliferation and growth of human pancreatic progenitors.

### PPDPF is broadly expressed across fetal pancreatic cell types beyond PP stage

Until now, we observed a reduction in PPs in *in vitro* differentiation experiments of PPDPF<sup>KO/KO</sup> hESCs and RNA-seq experiments revealed the preservation of endodermal gene expression and a lineage bifurcation into non-pancreatic tissue, however, both at relatively low expression level. To investigate how PPDPF is expressed beyond PP stage, we reanalysed a human fetal pancreas scRNA-seq data set covering key stages from PP formation, trunk and tip development until cell type specification (Fig 8A and 8B) [32] and found that PPDPF is broadly expressed at all investigated time points 4 to 11 weeks post conception (W4–11) (Fig 8C and 8D). While PPDPF expression increased from PP/MP stage to tip, trunk and endocrine precursors (EP), it remained expressed in all fetal pancreatic cell types (Fig 8C and 8D). Of note, albeit expressed, PPDPF expression was lowest in fetal acinar cells contrasting the exocrine-specific expression and function of its *exdpf* orthologue during zebrafish development (Figs 8C, 8D and S5A–H). In addition, we investigated PPDPF expression in an adult pancreatic scRNA-seq data set [33], where PPDPF appears also broadly expressed in different adult pancreatic cell types (S5I Fig).

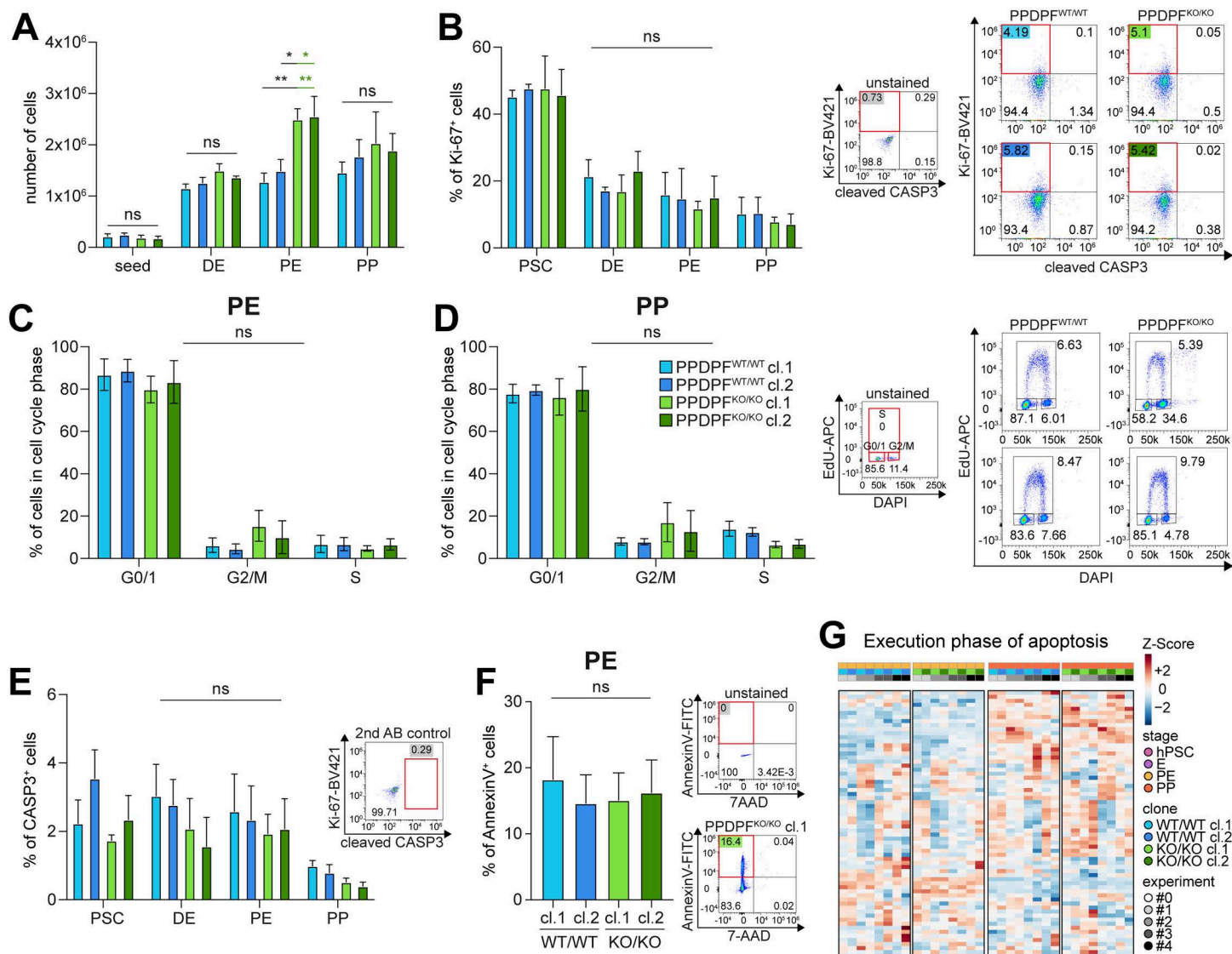




**Fig 6. Incomplete pancreatic progenitor differentiation upon PPDPF ablation.** (A) Gene set enrichment analysis (GSEA) against gene sets compiled from pancreatic differentiation experiments [23,27]. False discovery rates (FDRs)  $\leq 0.25$  are highlighted in red (left panel), while comparisons that are enriched in PPDPF<sup>KO/KO</sup> cells are depicted in green and comparisons that are depleted in blue (right panel). (B) GSEA plots against a definitive endodermal and pancreatic progenitor gene set are shown for the comparison of PPDPF<sup>WT/WT</sup> and PPDPF<sup>KO/KO</sup> PPs. (C) Heatmap illustrating row/gene-scaled gene expression at all stages and PP stage only (Zoom, and

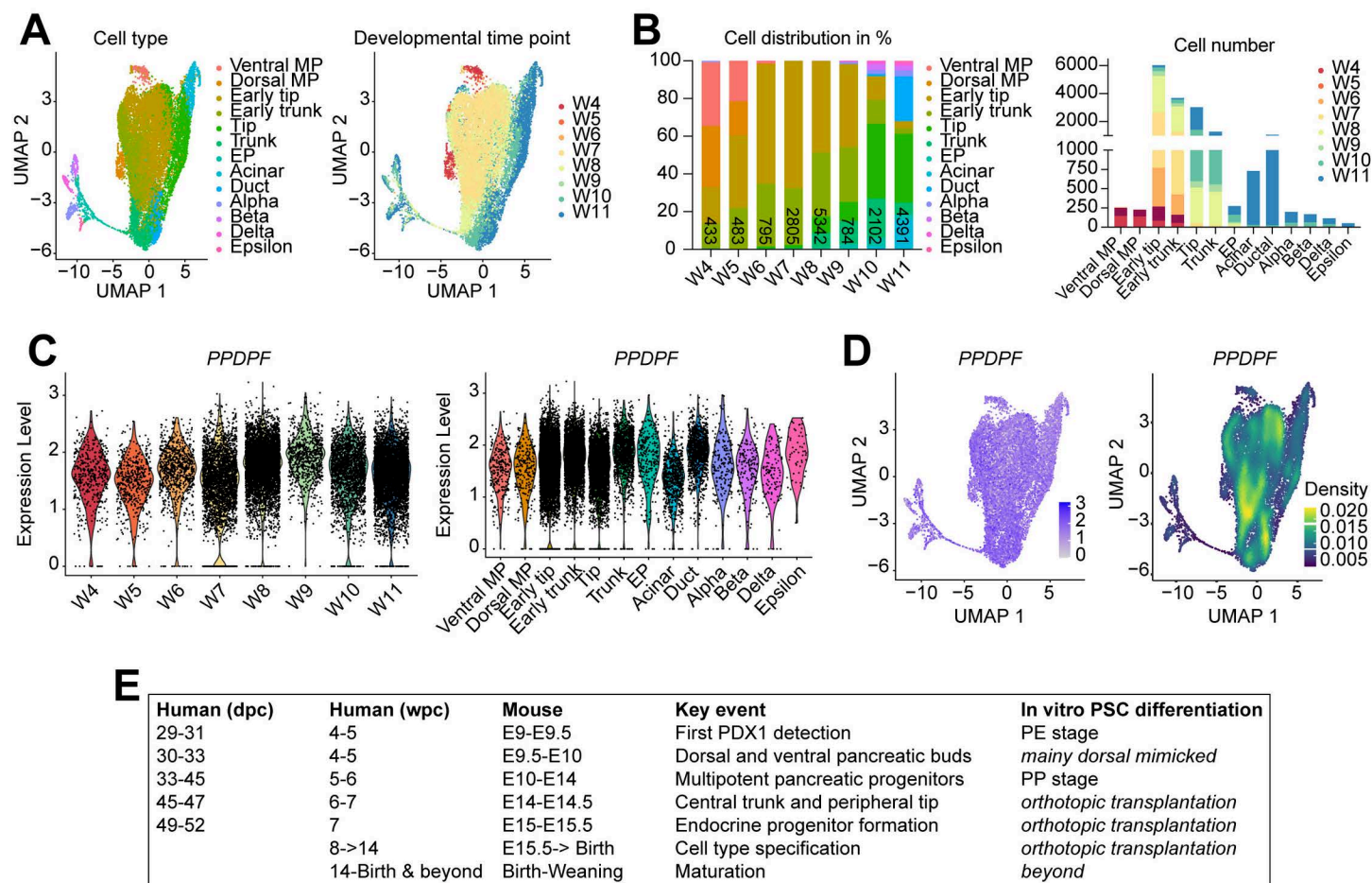
bottom panel) of all detected genes within the “Definitive Endoderm” gene set. (D) As in C but zoom (bottom panel) shows PE and PP stage, and “Pancreatic Progenitor” gene set is depicted. (E,F) Line plots (Mean  $\pm$  SEM) against key DE (SOX17, EOMES) and PP genes (NKX6-1, GP2, PTF1A) together with exemplary genes top ranked in GSEA (DE: SEMA3E, FTS; PP: SPHK1).

<https://doi.org/10.1371/journal.pgen.1011657.g006>



**Fig 7. No alteration in proliferation and growth characteristics of PPDPF<sup>KO/KO</sup> PPs.** (A) Cells per well of a 24-well plate counted with help of a Neubauer chamber. (B) Flow cytometry of proliferation marker Ki-67 and gating strategy. (C-D) EdU-based cell cycle analysis at PE (C) and PP (D) stage and gating strategy. (E) Flow cytometry of apoptosis marker cleaved Caspase 3 (CASP3). For A-E:  $n = 3$ , in duplicates; Mean  $\pm$  SEM; ordinary two-way ANOVA, not assuming sphericity, followed by Tukey's multiple comparison test. (F) Flow cytometry of Annexin V and gating strategy.  $n = 4$ , in duplicates; Mean  $\pm$  SEM; ordinary one-way ANOVA, not assuming sphericity, followed by Tukey's multiple comparison test. (G) Heatmap of RNA-seq data illustrating row/gene-scaled gene expression at all stages of all detected genes within the hallmark gene set “Execution phase of apoptosis”. Additional information: 0 out of 156 genes were significantly upregulated (DEGs) in PPDPF<sup>KO/KO</sup> PE cells over PPDPF<sup>WT/WT</sup> PE cells, while 1 was downregulated. At PP stage, 1 was up and 0 were down.

<https://doi.org/10.1371/journal.pgen.1011657.g007>



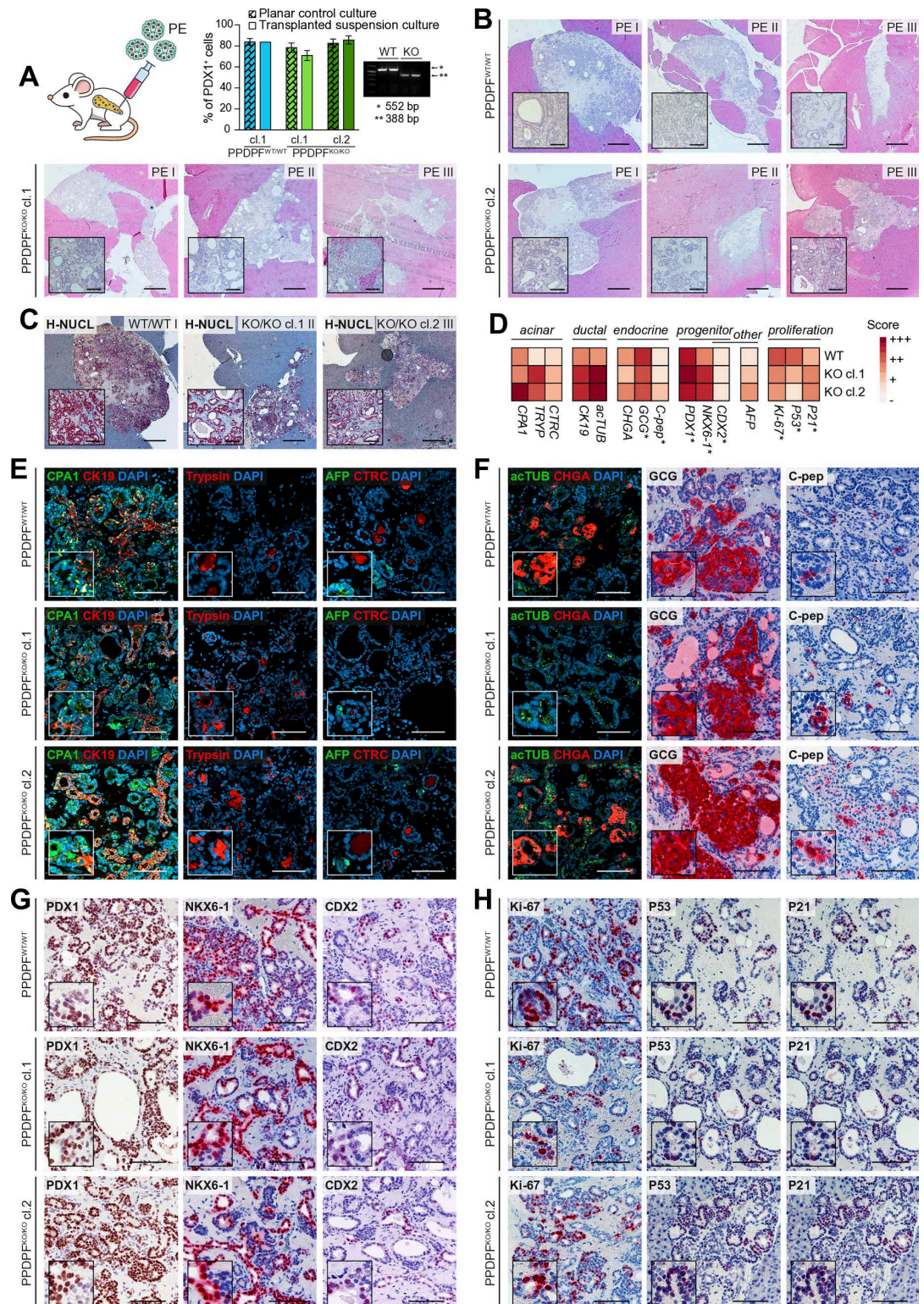
**Fig 8. PPDPF is broadly expressed across fetal pancreatic cell types beyond PP stage.** (A) Reanalysis of a publicly available human fetal pancreas scRNA-seq data set from Ma, Zhang [32]. UMAP representation with cell type annotation (left) along developmental time points (right). (B) Number of analyzed cells and time-resolved cell type distribution. (C) Violin plots of *PPDPF* expression along developmental time points (left) and within cell types (right). (D) *PPDPF* expression within UMAP space. (E) Classification of fetal stages of human and mouse pancreas development and estimation of appropriate stage of hPSC-based pancreatic differentiation. Information about early human stages (29-52 dpc) was compiled from Jennings, Berry [41], later stages from Nair and Hebrok [57]. Dpc: days post conception; wpc: weeks post conception; EP: endocrine progenitor; MP: multipotent progenitor; UMAP: Uniform Manifold Approximation and Projection.

<https://doi.org/10.1371/journal.pgen.1011657.g008>

## PPDPF<sup>KO/KO</sup> cells successfully develop into pancreatic tissue upon orthotopic transplantation

To experimentally investigate the role of *PPDPF* during pancreatic cell type specification *in vivo* (Fig 8E), we used our previously licensed orthotopic transplantation approach [14,23,34,35] using PE cells derived from 1 *PPDPF*<sup>WT/WT</sup> clone and 2 *PPDPF*<sup>KO/KO</sup> clones into 3 immunodeficient NOD/SCIDgamma (NSG) mice for each condition. Differentiation efficiencies of transplanted PE cells were very similar between genotypes, and we genetically confirmed genotypes of clones at the time point of transplantation as additional quality control (Fig 9A). All 9 engraftments efficiently formed fetal pancreas-like tissue (Fig 9B) of human origin (Fig 9C). While consecutive sections and engraftments were distinct in size and representation of tubular structures, we did not observe genotype-specific alterations. To determine proper lineage segregation, we quantified the expression of key acinar, ductal, endocrine, non-pancreatic and proliferation-related marker in engraftments (+++, ++, +, -)





**Fig 9. PPDPF<sup>KO/KO</sup> PE cells successfully develop into pancreatic tissue upon orthotopic transplantation.** (A) PE cells, that were cultured 2 days in suspension (d7-d9) for improved engrafting, were orthotopically transplanted in the pancreas of NSG mice and propagated *in vivo* for 8 weeks. Quality control of transplanted cells highlighting similar differentiation efficiencies between genotypes and to normal planar PP control (n = 1, in duplicates; Mean ± SD). DNA has been isolated directly prior to transplantation for an additional genotyping PCR control. (B) Hematoxylin and eosin (H&E) staining of

identified grafts with  $n = 3$  for PPDPF<sup>WT/WT</sup> and PPDPF<sup>KO/KO</sup> cl.1 and cl.2 PEs. (C) Human nucleoli (H-NUCL) staining illustrates human engraftment. (D) Quantification of depicted marker expression averaged over individual grafts of each genotype (– completely or nearly completely absent; + weak expression; ++ moderate expression; +++ strong expression). (E) Immunofluorescence (IF) staining of acinar marker CPA1, Trypsin, and CTSC, ductal marker CK19, and liver marker AFP. (F) IF and immunohistochemistry (IHC) of ductal cilia marker acetylated tubulin (acTUB), endocrine marker CHGA, GCG, and C-peptide (C-pep). (G) IHC of progenitor marker PDX1, NKX6-1, and CDX2. (H) IHC of proliferation marker Ki-67 and effectors of growth restriction P53 and P21. For all: Images are depicted that show strong marker expression to highlight trilineage competence. Scale bar: B–C: 500  $\mu$ m, inlet: 100  $\mu$ m. E–H: 100  $\mu$ m, inlets 2x enlarged.

<https://doi.org/10.1371/journal.pgen.1011657.g009>

(Fig 9D). We neither observed a reduction in acinar digestive enzymes (CPA1, TRYP, CTSC) nor in tubular ductal marker (CK19, acTUB) nor in endocrine (CHGA), alpha (GCG), and beta (C-pep) marker in PPDPF<sup>KO/KO</sup> grafts (Fig 9D–F). In addition, progenitor marker (PDX1, NKX6-1, CDX2) or non-pancreatic marker (AFP) were also not increased in PPDPF<sup>KO/KO</sup> grafts (Fig 9G). Finally, we also did not detect a genotype-consistent upregulation of cell-cycle checkpoint effectors (P53, P21) and a reduction of proliferation (Ki-67) (Fig 9H). In summary, PPDPF<sup>KO/KO</sup> PE cells differentiated efficiently into fetal pancreas-like tissue with similar size of engraftments and similar representation of acinar, ductal, and endocrine-specified cells as PPDPF<sup>WT/WT</sup> cells. In addition, we performed transplantation experiments of GP2 sorted/unsorted PPs according to our recently fine-tuned PP protocol [34] (S6A Fig). Importantly, PP transplantations confirmed the competence of PPDPF<sup>KO/KO</sup> cells to successfully differentiate into pancreatic-like tissue in a highly similar manner as PPDPF<sup>WT/WT</sup> cells (S6B–H Fig). Of note, the small observed changes in the quantification were not consistent between PE and PP transplantations (Figs 9D and S6D), underlining the preserved capacity of PPDPF<sup>KO/KO</sup> hPSCs to develop into the most important human pancreatic cell types *in vivo*.

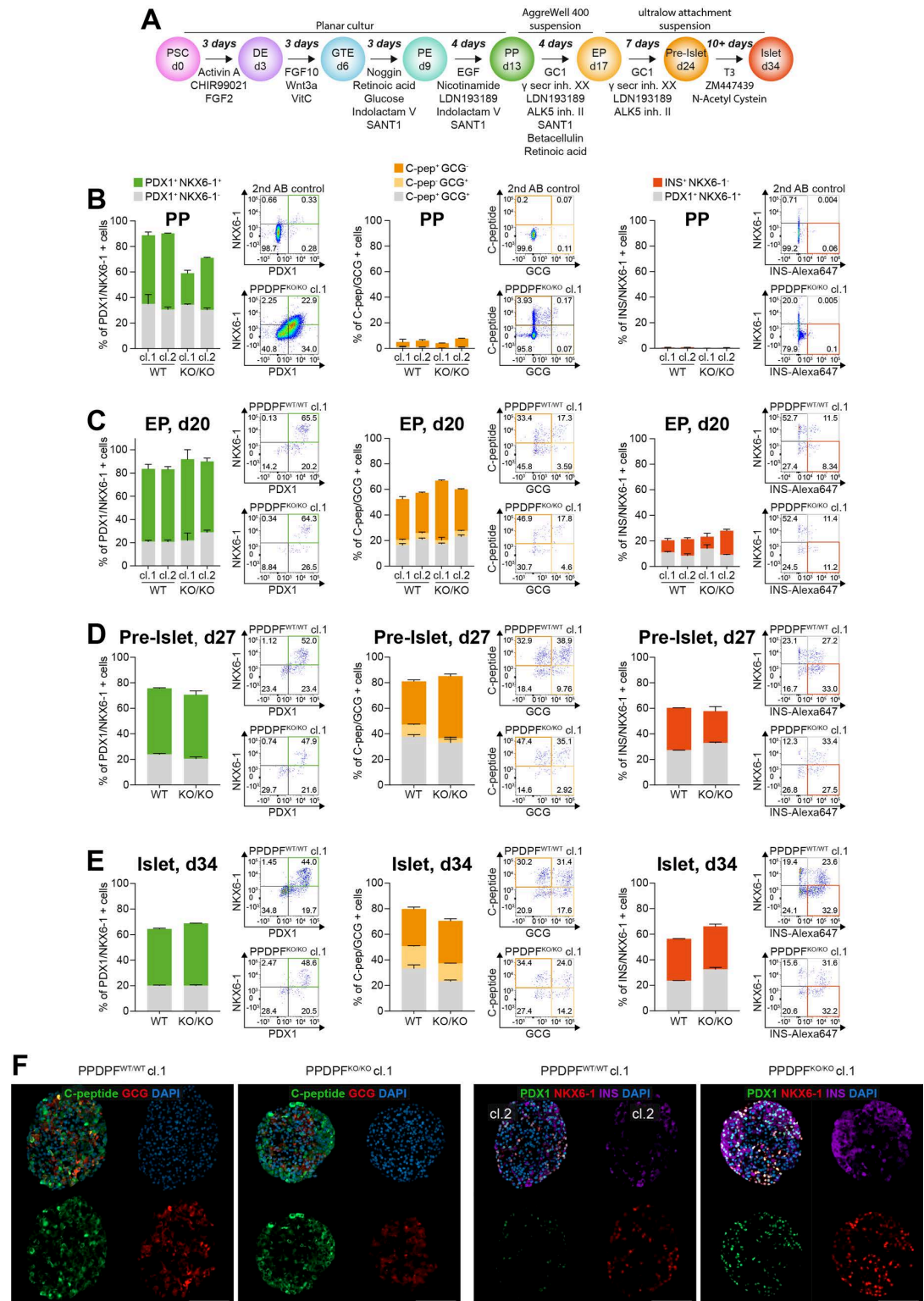
### PPDPF<sup>KO/KO</sup> cells efficiently differentiate into endocrine islet-like organoids *in vitro*

In a nutshell, observed differences between PPDPF proficient and deficient cells to form PP cells did not perpetuate to more mature pancreatic stages *in vivo*. It remained, however, unclear when PPDPF<sup>KO/KO</sup> PPs might have caught up to PPDPF<sup>WT/WT</sup> during PP specification. To dissect temporal aspects more closely, we differentiated PPs *in vitro* into endocrine islet-like organoids [36,37] in a proof-of-concept experiment over 34 days (Fig 10A). hPSCs were first differentiated into PP cells, where the decrease in NKX6-1<sup>+</sup> PPs upon PPDPF ablation was again reflected (Fig 10B, left), and no substantial premature activation of endocrine marker was found (Fig 10B, right). In contrast, upon further differentiation into endocrine progenitors (Fig 10C), pre-islets (Fig 10D) and more mature islet-like organoids (Fig 10E), the number of derived alpha-like cells (GCG) and beta-like cells (C-peptide, INS) was highly similar. Comparable endocrine differentiation efficiencies between PPDPF<sup>WT/WT</sup> and PPDPF<sup>KO/KO</sup> cells were further validated by immunofluorescence imaging on paraffin-embedded and sectioned islet-like organoids (Fig 10F). We therefore concluded that PPDPF<sup>KO/KO</sup> PPs harbour the capacity to efficiently specify pancreatic endocrine cells *in vitro* with no differences on marker expression being observed already at endocrine progenitor stage on day 20.

## Discussion

“Nomen est omen”, does not appear to keep its promise in case of PPDPF. The gene name *pancreatic progenitor differentiation and proliferation factor* suggests a key role of PPDPF for PP formation and expansion for tissue formation. In our comprehensive hPSC-based *in vitro* and *in vivo* experiments, however, we demonstrated that PPDPF is dispensable for





**Fig 10. PPDPF<sup>KO/KO</sup> cells efficiently differentiate into endocrine islet-like organoids *in vitro*.** (A) Schematic overview of human pluripotent stem (hPSC) cell-based differentiation into definitive endoderm (DE), gut tube endoderm (GTE), pancreatic endoderm (PE), and pancreatic progenitor (PP) cells in planar cultures followed by 3D endocrine progenitor (EP) formation and maturation into pre-islets and islet-like organoids at d34. (B-E) Differentiation efficiencies of clonal PPDPF<sup>WT/WT</sup> and PPDPF<sup>KO/KO</sup> hPSC lines, measured by flow cytometry ( $n = 1$  (independent differentiations), in duplicates (two wells per



differentiation); Mean  $\pm$  SD; at PP stage (B), EP stage (C), Pre-islet (D) and islet stage (E). (F) Representative immunofluorescence imaging of key endocrine marker (PDX1, NKX6-1, C-peptide, GCG, INS) on formalin-fixed paraffin-embedded (FFPE) organoid sections at d34 of differentiation. Scale bar: 50  $\mu$ m.

<https://doi.org/10.1371/journal.pgen.1011657.g010>

proper lineage specification and segregation into acinar, ductal, and endocrine cells. While initial *in vitro* differentiation experiments suggested a reduced capacity of PPDPF<sup>KO/KO</sup> hESCs to develop into PPs, global RNA-seq experiments revealed that stage-specific changes were rather small and might represent a slight delay in PP development together with a slight upregulation of non-pancreatic genes. PE- and PP-specific transplantation experiments and *in vitro* islet differentiations finally disclosed that PPDPF<sup>KO/KO</sup> hESCs indeed possess the competence to develop into all major cell types of the human fetal pancreas. The second aspect of its gene name suggests an important role of PPDPF for proliferation of PPs. However, proliferation marker and growth characteristics appeared not decreased upon loss of PPDPF during all investigated stages of pancreatic development including *in vitro* differentiation into DE, PE, and PP stage and *in vivo* specification into acinar, ductal, and endocrine cells.

Our data contrasts with the original finding in zebrafish, where the orthologue *exdpf* was found to be a novel player, that is, within the pancreas, specifically expressed in acinar cells, functions as a gatekeeper of exocrine and endocrine segregation, and that is essential for the formation of exocrine pancreatic tissue mass [1]. Utilizing publicly available scRNA-seq and proteome data sets as well as ISH of mouse fetuses, we were able to demonstrate a relatively broad expression profile of murine *Pdpf* and human PPDPF throughout development, but also in adult tissue. The lack of cell type and tissue specificity is in line with our observations, where we did not detect an exocrine to endocrine lineage switch upon ablation of PPDPF. Our results might be partly explained by an increasing evolutionary complexity from zebrafish to higher vertebrates and humans. While part of its developmental function might be residually visible in initial *in vitro* differentiation experiments demonstrating an effect on PP generation, the loss of PPDPF might be either compensated by other factors or might be simply not as decisive for the development of the more complex human pancreas.

Specifically, we found that PPDPF expression is regulated by FOXA1/2 and SOX9 TFs at DE to PP stage. FOXA1/2 play a central role for lineage specification and development of the neuronal tube, endodermal organs such as liver, lungs, kidney, prostate, and pancreas [38], the kidney and urethral tube [39]. The lack of binding of PP-specific TFs PDX1 and NKX6-1 fits to the relatively broad expression profile of PPDPF in non-pancreatic tissues. Interestingly, we observed binding of Ptf1a and Rbpjl to an additional distal gene region, likely possessing an enhancer function, at later murine E17.5 stage. While respective ChIP-seq data are not available for human cells or tissues, Ptf1a and Rbpjl are considered highly specific for acinar cells [40]. It remains therefore unclear, how PPDPF expression is regulated at later stages of human development and tissue maintenance. Its broad scRNA-seq expression profile in the pancreas however appears similar to FOXA2 expression (S5 Fig) rendering a sole dependence on PTF1A and acinar-specific function unlikely.

In humans, the pancreas develops from a separate ventral and dorsal anlage and multi-potent PTF1A positive PPs give rise to endocrine and exocrine cells [41,42]. Human *in vitro* PP differentiation protocols rather mimic the dorsal pancreas route [29]. Also in zebrafish, ventral and dorsal anlagen can be observed [43] but in contrast to mammals, amphibians and birds, only the ventral but not the dorsal PPs contribute to exocrine pancreas development [44] and *ptf1a* is only expressed in the ventral pancreas in zebrafish [45]. Since the function of zebrafish *exdpf* was placed downstream of *ptf1a* and since we also noted Ptf1a binding to the *Pdpf* locus at E17.5, the possibility that our hPSC-based differentiation approach might not

capture a potential phenotype due to the lack of the PTF1A/ventral pancreas route during PP differentiation, lack of full acinar maturity, or lack of morphogenesis-related processes cannot be completely excluded. However, we consider this rather unlikely, as (i) ventral and dorsal PPs, of which the latter also expresses PTF1A despite of lower expression levels ([S5A Fig](#)), contribute to exocrine and endocrine development in humans, (ii) analysis of human fetal development revealed similar PPDPF expression levels in dorsal and ventral PPs not following PTF1A expression ([Figs 8C and 5A](#)), and (iii) orthotopic transplantations of PPDPF<sup>KO</sup>/hPSC-derived pancreatic cells recapitulated acinar cell specification, secretion of digestive enzymes such as trypsin, and patterning into gland-like structures. In addition, we demonstrate that PPDPF expression can be regulated by FOXA1/2 and SOX9, providing an alternative or additional regulation mechanism to PTF1A TF activity.

A recent and large GWAS study specifically focusing on random glucose signals identified 143 SNPs including the PPDPF-associated SNPs highlighted in this study [[21](#)]. While the authors exemplarily demonstrated the relevance of such traits for mechanistical insights into glucose regulation, occurrence of potential diabetes complications and stratification of diabetes treatment [[21](#)], the underlying mechanism how the SNPs that correlate with decreased expression of PPDPF deteriorate glucose homeostasis remained unclear. Within this study, we can now show that one of the credible SNPs (rs72629024) is bound by endodermal FOXA1/2 TFs in a gene regulatory region associated with active promoter marks. The observed reduction in the generation of hPSC-derived PPs could explain changes in glycaemic control, especially in co-occurrence with additional unknown (non-genetic) risk factors. However, effects on glycaemic control can be caused by alterations in multiple organs such as kidney, liver, gut, adipocytes, cardiovascular or neuronal system or even eye and lungs. Hence, also a non-pancreatic function of PPDPF during development or tissue homeostasis could be the underlying cause of the observed glycaemic changes. PPDPF-associated SNPs also correlated with impaired renal activity, and renal and endocrine pancreatic function can affect each other in both directions. A recent GWAS study focusing on kidney function also reported on the correlation of the same PPDPF-associated SNP and impaired renal function [[46](#)]. In another study, liver-specific KO of *Pdpf* led to fatty liver formation in a mouse model at 32 weeks of age on chow diets, which was enhanced by high fat diet (HFD) [[47](#)]. Under both diets, higher blood glucose levels were found in the liver-specific *Pdpf* KO mice and *Pdpf* overexpression reduced blood glucose levels at least under HFD [[47](#)]. The changes in lipid metabolism offer an additional not directly pancreas-related explanation for the correlation between glycaemic alterations and PPDPF-associated SNPs. Importantly, the PPDPF-associated SNPs did not correlate with actual diabetes risk but only with glycaemic alterations thereby supporting the less important role for pancreas development in humans than in zebrafish.

While we initially had hypothesized a role as signaling protein in the developmental TF circuitry, instead, we have identified and exemplarily validated other novel interaction partners of PPDPF by SILAC mass spectrometry after PPDPF overexpression in HEK293T cells. Several of the hits including MAGED1 and KCTD12 have been described in both cancer context and general cellular processes. The list of putative interaction partners generated in HEK293T cells possesses, in our opinion, a valuable resource for future mechanistical studies in the rising field of PPDPF-related cancer studies. Some examples of identified interaction partners, where PPDPF interaction might be potentially involved in context-important gene and protein regulatory functions are USP15, a deubiquitinating enzyme involved in many protein regulatory functions [[48](#)], RBM10, a nuclear RNA-binding protein (RBP) that regulates the alternative splicing of primary transcripts [[49,50](#)], or the PRMT5-WDR77 complex together with its adapters (CLNS1A, COPRS, RIOK1) involved in methylation of the spliceosome, histones and ribosomal complexes [[51](#)] and a promising target of various anticancer therapies

[52,53]. Of note, current literature highly suggests context-specific mechanisms of PPDPF with a putatively converging theme of (indirect) protein modulatory functions [3,4,6–8,47]. We therefore hypothesize that PPDPF acts as a glue-like protein that alters the interaction of bound target proteins to other proteins thereby modulating function and regulation. In agreement, PPDPF is small and appeared very sticky in our SILAC screen experiments.

Furthermore, several studies have disclosed an oncogenic [2–7,9], pro-proliferative [1–3,5] and anti-apoptotic [2–4,8] role of PPDPF, while one study postulated a tumor suppressive and anti-proliferative role [8]. In contrast, we did not detect a pro-proliferative or anti-apoptotic function of PPDPF during our pancreatic differentiation experiments suggesting no cell-conserved and direct function of PPDPF in tissue growth. Instead, the indirect protein regulatory function of PPDPF might only become relevant upon additional inflammatory signals or disturbed cell homeostasis caused by e.g. oncogenic drivers. In conclusion, our study reveals that PPDPF is dispensable for human pancreas development, while current literature suggests an important protein regulatory role of PPDPF during tumor formation and progression in several organs.

## Materials and methods

### Ethics statement

All conducted experiments were carried out in compliance with German law and animal experiments were approved by the governmental office (“Regierungspräsidium”) of the state of Baden-Württemberg in Tuebingen, Germany: Xenotransplantation of hESC-derived pancreatic cells into the pancreas of NOD scid gamma (NSG) mice (NOD.Cg-Prkdcscid Il2rgtm-1Wjl/SzJ strain (Charles River); RRID:BCBC\_4142) was approved under TVA1406 and mouse organ isolation and processing of E14.5 C57BL/6J embryos under TV-Nr. o.161-3.

### Animal experiments

**Orthotopic transplantation.** NSG mice were used for xenotransplantation of hESC-derived pancreatic cells into the murine pancreas. Mice were sacrificed 8 weeks after transplantation (56 days for PE transplantation, 52 days for PP transplantation). While the experimental procedure of the transplantation has been detailed previously [23], cell preparation will be described in the pancreatic differentiation section.

**In situ hybridization (ISH).** For riboprobe generation, PCR was performed using GoTaq Flexi Polymerase (Promega) with the following primer pairs: Pdpf<sub>fwd</sub> 5'-gtcttccaaagccgaccc-3' and Pdpf<sub>rev</sub> 5'-gggtggctgctgtgagttta-3', Pdx1<sub>fwd</sub> 5'-gtcttccaaagccgaccc-3' and Pdx1<sub>rev</sub> 5'-caacagcctaagacaggct-3'. Pdpf and Pdx1 were amplified from cDNA derived from (E11) total mouse RNA (Takara) and adult mouse pancreas of C57BL/6J mouse strain, respectively. Purified PCR products were cloned into pCR 2.1 TOPO plasmid (Thermo) according to manufacturer's recommendations with insert:vector ratios of 2:1 and 5:1 with 50 ng vector overnight (o/n) at 16°C. Insert sequences were verified by sanger sequencing (Eurofins Genomics). Plasmid DNA was linearized using HindIII (NEB). To generate DIG-labeled riboprobes, *in vitro* transcription using T7 RNA polymerases (Roche) was performed. Pdx1 antisense [58] and Pdpf sense probes served as positive and negative controls, respectively. For *in situ* hybridization, mouse tissue was fixed in 4% PFA, cryoprotected in 20% sucrose, and embedded in OCT compound. 18 µm thin frozen sections through the pancreatic anlage of E14.5 C57BL/6J embryos were collected on SuperFrost Plus slides (Thermo). *In situ* hybridization was performed according to [59] with minor modifications. Briefly, after proteinase K treatment and acetylation, sections were permeabilized using 1% TritonX100 and incubated o/n at 63°C with riboprobes diluted



in hybridization buffer [50% formamide, 5x SSC, 5x Denhardt's solution, 250 µg/mL yeast RNA, 500 µg/mL herring sperm DNA]. Sections were washed three times in 0.2x SSC at 65°C, blocked with 10% goat serum, and incubated with alkaline phosphatase conjugated anti-DIG antibody (Roche) at 4° o/n. Alkaline phosphatase activity was detected using nitro blue tetrazolium (NBT) and 5-bromo-4-chloro-3-indolyl-phosphate (BCIP) (Roche).

## Embryonic stem cells

**Maintenance culture.** In this study, the human embryonic stem cell line HUES8 (Harvard University; RRID:CVCL\_B207) was used. Culture and differentiation towards the pancreatic lineage was performed with permission from the Robert Koch Institute according to the “79. Genehmigung nach dem Stammzellgesetz, AZ 3.04.02/0084”. HUES8 cell authentication was confirmed with a DNA profile using nonaplex PCR of Short Tandem Repeats (STRs) done by the Leibniz-Institute.

Culture conditions for hESCs were 5% CO<sub>2</sub>, 5% O<sub>2</sub>, and 37°C. Cells were cultured on hESC Matrigel (Corning) coated plates (according to manufacturer's recommendations) in FTDA medium [60] or mTesR1 medium (STEMCELL Technologies) with daily media change. Splitting in a 1:4 to 1:6 ratio was done twice a week with TrypLE Express (Thermo) to enable feeder-free homogenous monolayer cultures. For this, cells were washed once with PBS and incubated with TrypLE for 3–5 min at 37°C. By adding DMEM-F12+Glutamax (Gibco) the enzymatic reaction was stopped and after centrifugation at 200 x g for 5 min, cells were resuspended in FTDA or mTesR1 supplemented with 10 µM ROCK inhibitor (Y-27632, Abcam) and seeded on a new, coated well.

**Pancreatic progenitor differentiation for *in vitro* experiments.** Differentiation of hESCs into PPs was performed as described in [14]. Briefly, hESCs were seeded on 24-well plates, which were coated with growth factor reduced Matrigel (Corning, diluted 1:18 in DMEM-F12+Glutamax), 24 h prior to the start of differentiation. Cells were cultured in basal media BE1 at d0–d5: MCDB131 (Thermo) with 2 mM L-Glutamine (Gibco), 1.174 g/l sodium bicarbonate (Sigma), 0.8 g/l cell culture tested glucose (Sigma) and 0.1% fatty acid free BSA (Sigma/Proliant). At d6–14 BE3 was used as basal medium: MCDB131 with 2 mM L-Glutamine, 1.754 g/l sodium bicarbonate, 0.44 g/l glucose, 0.5% ITS-X (Gibco), 44 mg/l L-Ascorbic acid (Sigma) and 2% fatty acid free BSA. At a confluency of about 70–95%, differentiation was initialized. For that, cells were washed with PBS (Gibco), and BE1 medium containing 3 µM CHIR99021 (Axon Medchem) and 100 ng/ml ActivinA (Peprotech/R&D) (d0 medium) was added. After initializing differentiation, cells were cultured at 5% CO<sub>2</sub>, 21% O<sub>2</sub>, and at 37°C. After 24 h, medium was replaced by BE1 medium supplemented with 100 ng/ml ActivinA (d1 – d3 medium). During the next two days d1 – d3 medium was added to allow formation of definitive endoderm (DE), and on day 4 and day 5 BE1 medium containing 50 ng/ml KGF (Peprotech) was added inducing foregut specification. From day 6 until day 9, cells were cultured in BE3 medium supplemented with 0.25 µM SANT1 (Sigma), 2 µM retinoic acid (Sigma), 0.2 µM LDN193189 (Sigma), and 0.5 µM PD0325901 (Calbiochem). From day 10 on cells received 50 ng/ml FGF10 (Peprotech/R&D), 0.33 µM indolactamV (Stemcell Technologies), 10 µM SB431542 (Axon Medchem), and 16 mM glucose (resulting in a final glucose concentration of 24 mM) in BE3 medium for four days.

**Pancreatic progenitor differentiation for *in vivo* experiments.** For orthotopic transplantation, our most recent protocol was implemented as detailed in [34]. This refined protocol allows higher PP efficiencies and better representation of the multipotency marker GP2. For PE transplantations, planar cultures were harvested at d7 2 days prior to PE and transferred to Costar ultra-low attachment plates (Corning) at 37°C 5% CO<sub>2</sub> without orbital shaking. Suspension culture spheres were harvested at the day of transplantation and cells

corresponding to  $1 \times 10^6$  PE cells were transplanted. For PP transplants, cells were GP2 MACS sorted at d12 as optimized in [34] (based on [61,62]) and cultured in ultra-low attachment plates at  $37^\circ\text{C}$  5%  $\text{CO}_2$  without orbital shaking for 1 day. While  $0.56 \times 10^6$  sorted PP cells were transplanted in most cases,  $10 \times 10^6$  unsorted planar PP cells were transplanted when sorted PP cells were not sufficient for 3 transplantation experiments. In summary, following cells have been transplanted:

- (1)  $1 \times 10^6$  suspension PE: 3x PPDPF<sup>WT/WT</sup>, 3x PPDPF<sup>KO/KO</sup> cl.1, 3x PPDPF<sup>KO/KO</sup> KO cl.2 (all engrafted well).
- (2A)  $0.56 \times 10^6$  GP2-sorted suspension PP: 2x PPDPF<sup>WT/WT</sup> (1 not engrafted), 2x PPDPF<sup>KO/KO</sup> cl.1, 3x PPDPF<sup>KO/KO</sup> KO cl.2
- (2B)  $10 \times 10^6$  unsorted planar PP: 1x PPDPF<sup>WT/WT</sup>, 2x PPDPF<sup>KO/KO</sup> cl.1 (1 teratom, excluded from analysis).

Based on our experience, we suggest transplanting  $1 \times 10^6$  unsorted PE or PP cells that were cultured as spheroids at least 1 day prior to transplantation for future experiments.

***In vitro* differentiation into pancreatic endocrine organoids.** We performed stem cell-derived islet differentiation as recently detailed in [36].

## Genome editing by CRISPR/Cas9 in hESCs

**Construct generation.** A large deletion in the *PPDPF* gene locus was generated by CRISPR/Cas9 gene editing using a four times nicking strategy, thereby inducing two DSBs. Two nicks at each intended cleavage site were required to lead to one DSB, thereby minimizing off-target cleavage. crRNAs were designed in intronic regions flanking exon 3 of the *PPDPF* gene locus: guide 1a: GCTGGTTCCGCAGCCTGCAC; guide 1b: GCCTGAAGGCCGGTGGGCTG; guide 2a: GGGGCACTCGGTACTGCTGC; and guide 2b: GAAGCCATTCCCCACCCCC. CRISPR plasmids were based on pX335-U6-Chimeric\_BB-CBh-hSpCas9n(D10A)-2A-GFP vector (a gift from Boris Greber at MPI Münster [63], itself based on Addgene #42335, a gift from Feng Zhang [64]). Phosphorylation for subsequent ligation of complementary oligonucleotides was performed using the T4 Polynucleotide Kinase (NEB). Oligonucleotides were annealed at  $37^\circ\text{C}$  for 30 min and  $95^\circ\text{C}$  for 5 min with subsequent cooling to  $25^\circ\text{C}$  at  $0.1^\circ\text{C}/\text{sec}$ . Restriction digestion of the vector was performed using FastDigest BbsI (Thermo, FD1014) for 30 min at  $37^\circ\text{C}$ . After digestion, the vector was purified using the JetQuick Genomic DNA Clean up Spin Kit (Genomed) according to the manufacturer's protocol. Ligation of the annealed oligonucleotides with the BbsI-digested vector was performed with a T7 DNA Ligase (NEB) for 1h at room temperature (RT). To prevent undirected recombination, ligation products were treated with a Plasmid Safe ATP-dependent DNase (Biozym, 161010). Electrocompetent DH5A E. coli cells were transformed with final constructs for plasmid amplification. The final DNA of amplified plasmids was checked by PCR and Sanger sequencing.

**Cell line generation.** Before transfection, HUES8 cells were dissociated into single cells using Accutase and 200,000 cells were seeded per well of a 6-well plate. Cells were cultured o/n in FTDA media supplemented with  $10 \mu\text{M}$  ROCK inhibitor.  $2 \mu\text{g}$  DNA (consisting of  $0.5 \mu\text{g}$  of each construct) was transfected via FuGene HD Transfection Reagent (Promega) in OptiMEM medium (Thermo) to one well of 6-well plate for a 4h incubation. After 1 day, transfected hESCs were selected with  $1 \mu\text{g}/\text{mL}$  puromycin for 1 day. Multiple single cell-derived colonies were picked mechanically as soon as small colonies had formed for screening of gene edited clones.

**Screening of edited clones: DNA isolation and PCR reaction.** Clonal colonies were mechanically dissociated and one half of the cells was further cultivated and expanded while the other half was used for genotyping. DNA was isolated using the QIAamp DNA Micro Kit (Qiagen), according to manufacturer's instructions and amplified using a high-fidelity PrimeSTAR GXL DNA Polymerase (Takara). PCR screening was based on a primer pair spanning the target region (fwd: GAGCCCCGCCTCGGTAAATAAC; rev: AGGGTGGACTTCCCGAAAAAGAA). PCR products were sent for Sanger sequencing (GATC Biotech AG Tuebingen) for validation of clonal genotypes (sequencing primer: CTCGGTAAATAACCCAGC). For off-target sequencing, following primer pairs were used: For MGMT, fwd: TCAGGAGCACCCATTTGGTC, and rev: GGCCCCCATTTAAGGGTTCA and for TTC34, fwd: AAACGCACCCACAGACCG, and rev: TGAGGGTGGGGTTTCTGTTC.

## Sample processing and molecular biology assays

### Co-Immunoprecipitation experiments.

#### Construct generation

While expression plasmids of GATA4, FOXA1, FOXA2, PDX1, NKX6-1 with C-terminal Flag- and GFP-tag have been previously generated in a pcDNA3 backbone [13], PPDPF, KCTD12, and MAGED1 expression plasmids with the same tags were cloned in this study. PPDPF constructs were commercially synthesized (Eurofins) into a pEX-A backbone and recloned into our library of pcDNA3, pcDNA3-Flag, and pcDNA3-GFP via EcoRI and NotI restriction enzymes. For MAGED1 and KCTD12, the coding sequence (CDS) was amplified from human cDNA and first TOPO-cloned into pCR XL TOPO and pCR 2.1 TOPO, respectively, using TA cloning kits (Thermo). During PCR amplification, enzyme recognition sites were introduced with following primers and GoTaq Flexi Polymerase (Promega): MAGED1 (NotI fwd: ctgaacGCG GCCGCaaGCTCAGAAAATGGACTGTGGTG; XbaI rev: ctgcacTCTAGATgcTCACTCAAC CCAGAAGAAACCAA); KCTD12 (EcoRI fwd: ctgaacGAATTCGCTCTGGCGGACAGCAC; XbaI rev: ctgcacTCTAGATCACTCCCTGCAGAAGACGTAC). For KCTD12 PCR reaction, 3% DMSO was added. Topo cloning was performed according to manufacturer's recommendation with insert:vector ratios of 2:1 and 5:1 with 50 ng vector and o/n TA reaction at 16°C. For transformation and bacteria amplification, chemical competent DH5a were used for PPDPF and MAGED1, while dam incompetent chemical competent bacteria (JM110) were employed for KCTD12 due to a Dam methylation blocked XbaI site. After topo cloning and validation by sanger sequencing (Eurofins Genomics), CDSs of interest were cloned into target pcDNA3-Flag and pcDNA-GFP vectors via restriction enzyme digestion. Sticky ends were produced for MAGED1 plasmids with NotI-XbaI, for PPDPF and KCTD12 with EcoRI-XbaI and cloned via T4 DNA ligation (NEB) using insert:vector ratios of 3:1 with 100 ng vector input.

#### Lipofection and protein lysis of HEK293T cells

2.1 x 10<sup>6</sup> HEK293T cells were seeded in DMEM with 10% FCS and 1x Penicillin/Streptomycin on a 10 cm dish 1 day prior to transfection. For transfection, 10 µg plasmid DNA was transfected with 18 µL Lipofectamine 2000 (Thermo). After 24 h, cells were washed with PBS and harvested with 0.25% Trypsin/EDTA. After one additional PBS washing step, cells were resuspended in 600 µL chaps lysis buffer (0.01 M Chaps + 50 mM Tris-HCL (pH 7.8) + 150 mM NaCl) supplemented with 5 mM NaF + 0.5 mM PMSF + 1x protease inhibitor cocktail (Sigma) and lysed on ice for 1 h.

#### Immunoprecipitation experiments

After washing agarose-conjugated α-Flag M2 beads (Sigma) 6 times with 0.5% BSA in PBS, beads were centrifuged at 420 x g for 1 min at 4°C and protein extracts were added to 40 µL



beads and put on a shaker o/n at 4°C. After incubation, samples were washed 6 times with chaps buffer and resuspended after centrifugation at 20,800 x g for 1 min at 4°C in 1x laemmli buffer. Proteins were detected via Western Blotting as detailed below.

**Western Blotting.** For protein extraction of hESCs and differentiated pancreatic cells, cell lysates were generated by incubating cell pellets in RIPA buffer supplemented with 1 mM PMSF, phosphatase inhibitor (1x) and protease inhibitor cocktail (1x) for 30 min on ice and vortexing every 10 min. After 8 min centrifugation at 10,000 x g, supernatant containing the protein fraction was collected. Protein concentration was determined using a Bradford reagent (Bio Rad) and equalized amounts of protein lysates were separated on a 10–12% polyacrylamide gel in SDS-buffer followed by blotting to a methanol-activated Immobilon-P PVDF membrane (Millipore) by using semidry transfer buffer (32 mM glycine (Appllichem), 44 mM Tris, and 20% methanol (Sigma)) and the Transblot semidry transfer system (Bio-Rad). Effective protein transfer was confirmed by Ponceau staining (AppliChem) before membrane was blocked with 5% BSA and 0.1% Tween20 (Sigma) in TBS (TBS-T) for at least 1 h at RT. Membranes were incubated with primary antibodies diluted in blocking solution o/n (o/n) at 4°C. After washing three times with TBS-T, incubation with secondary antibody anti-mouse-horseradish peroxidase (HRP) or anti-rabbit-HRP (ECL anti-rabbit or mouse IgG, GE Healthcare) was performed for 1 h at RT. HRP was detected with the SuperSignal West Dura kit (Thermo) on a chemiluminescence imaging fusion SL system (VILBER). The following primary antibodies were used: anti-GFP (Roche, 1181 446 0001, 1:500), anti-Flag M2 (Sigma, F3165, 1:500), anti-PDX1 (R&D, AF2419, 1:500), anti-PPDPF (Atlas Antibodies, HPA040929, 1:150), and anti-Vinculin (Sigma, V9264, 1:1000).

**Flow cytometry of surface marker.** At PSC stage, pluripotency marker SSEA4 and TRA1-60 were checked after gene editing. At DE stage, differentiation efficiency was determined by KIT (CD117) and CXCR4 (CD184) surface marker staining. For each measurement, cells from one well of a 24-well plate were harvested with TrypLE, enzymatic reaction was stopped with FC buffer containing 2% FCS in PBS and washed once with FC buffer (200 x g, 5 min). Cells were blocked for at least 20 min on ice with blocking buffer consisting of 10% FCS in PBS and washed again with FC buffer. After resuspension of the cell pellets in 50 µl FC buffer, DE cells were incubated with PE-conjugated CXCR4 antibody (Thermo, MHCXCR404, 1:50) and APC-conjugated KIT antibody (Thermo, CD11705, 1:100) for 45 min on ice. hESCs were incubated with PE-conjugated SSEA4 (BD Bioscience, 560128, 1:10) and FITC-conjugated TRA1-60 (BD Bioscience, 560380, 1:10) for 60 min on ice. Samples were washed and resuspended in FC buffer and filtered using a 50 µm polyamide mesh (Hartenstein). DAPI was added after the last washing step at a concentration of 150 ng/µl to distinguish viable (DAPI<sup>-</sup>) and dead (DAPI<sup>+</sup>) cells during analysis.

**Flow Cytometry of Annexin V.** For Annexin V staining, cells were harvested as described for surface marker staining and staining was performed according to manufacturer's instructions of the FITC Annexin V Apoptosis Detection Kit with 7-AAD (Biolegend, 640922). Instead of using the recommended volumes of reagents, applied volumes were halved without changing reaction mix concentrations.

**Flow cytometry of intracellular marker.** Differentiation efficiency was analyzed at PE and PP stage by FC-based analysis of intracellular PDX1 and NKX6-1 expression. Endocrine organoids were singularized with TrypLE for around 9–12 min with pipetting up and down in-between and control of proper singularization under the microscope. Planar cells were harvested and singularized as for surface marker staining. Singularized cells were washed with PBS (200 x g, 5 min) and fixed on ice for 25 min in 4% PFA in PBS with 100 mM sucrose (Sigma). Samples were washed twice with PBS and blocked with 5% donkey serum (Jackson ImmunoResearch) in 0.1% Triton-X/PBS for 30 min on ice. After centrifugation (3000 x

g, 5 min) and two rounds of washing with 2% donkey serum, 0.1% Triton-X in PBS (wash solution), cells were resuspended in blocking solution with primary antibodies. Incubation was performed o/n at 4°C or 90 min on ice with combination of following antibodies: anti-PDX1 (R&D, AF2419, 1:500), anti-NKX6-1 (DSHB, F55A12, 1:150), anti-Ki67 (Thermo, MA5-14520, 1:1000 or DAKO, M7240, 1:200 or BV421-coupled (BD Horizon, 562899, 1:20)), anti cl-CASP3 (Cell Signaling, 9661, 1:800), GCG (Sigma, G2654, 1:500), C-pep (Cell Signaling, 4593, 1:100), and INS-Alexa647 (Cell Signaling, 3014, 1:80). Next, the samples were washed twice and incubated with Alexa Fluor secondary antibodies (Thermo, 1:500) for 90 min on ice. After one additional washing step, cells were finally resuspended in washing solution and filtered to obtain single cells before measurement. FC measurements were performed on an LSR II (BD) or an AttuneNxT (Thermo) flow cytometer.

**Flow cytometry of EdU staining.** EdU staining was performed using the Click-iT EdU Alexa Fluor 647 Assay Kit (Thermo, C10635). EdU at a concentration of 10  $\mu$ M was added to the cell cultures prior to harvesting. Chosen incubation time was optimized for each cell type: At hPSC stage, EdU was incubated for 45 min, at DE stage for 2 h and at PE and PP stage for 4 h. Cells were harvested, fixated as described in the intracellular flow cytometry section and staining was performed according to manufacturer's recommendations with minor changes. Instead of using the recommended volumes of reagents, applied volumes were halved without changing reaction mix concentrations. DNA was finally incubated with 1  $\mu$ g/ $\mu$ L DAPI for around 30min prior to measurements on an AttuneNxT (Thermo) flow cytometer.

**Flow cytometry analysis.** Analysis of flow cytometry data including gating was performed using FLOWJo version 10.8.1. Briefly, Cells were gated on the area of forward and sideward scatters. Single cells were subsequently defined by the area and width of the forward scatter followed by an additional stringency step using the area and width of the sideward scatter. In case of live cell measurements, DAPI<sup>-</sup> or 7AAD<sup>-</sup> cells were gated as living cells for subsequent marker analysis (area of specific channels).

**Immunocytochemistry staining.** Prior to immunocytochemistry (ICC/IF) in-well staining, stem cells were cultivated and differentiated on Ibidi-precoated glass-bottom IBIDI-24-well plates (IBIDI) with additional Matrigel coating as outlined in maintenance and differentiation culture. Cells were washed with PBS, fixed in 4% PFA+100 mM sucrose solution at RT for 20 min, and washed with PBS three times. Quenching was performed with 50 mM NH<sub>4</sub>Cl (Sigma) for 10 min and wells were washed three times with PBS before permeabilization with 0.1% Triton-X/PBS was performed for 30 min at 4°C. Cells were blocked with 5% normal goat or donkey serum (Jackson ImmunoResearch) in 0.1% Triton-X/PBS for 45 min and incubated with the primary antibody solution at 4°C o/n. The following antibodies were used: anti-OCT4 (Santa Cruz, sc-5279, 1:100), anti-SOX2 (R&D, MAB2018, 1:300), anti-SOX17 (R&D, AF1924, 1:500), anti-PDX1 (R&D, AF2419, 1:300), and anti-NKX6-1 (DSHB, F55A12 (concentrate), 1:150). On the next day, cells were washed three times with PBS and incubated with secondary Alexa Fluor antibodies (Thermo) 1:500 diluted in blocking solution at RT for 1 h in the dark. After washing with PBS, 500 ng/ml DAPI in PBS was added to the cells for 10 min. Wells were either directly imaged or stored in PBS at 4°C prior to imaging. Cells on IBIDI-plates were imaged on an AxioObserver 7 microscope (Zeiss) equipped with a AxioCam 702 camera (Zeiss) using ZEN blue software (Zeiss).

## Histology and immunohistochemical staining

### Histological processing

Pancreata of transplanted mice and endocrine organoids were fixated with 4% paraformaldehyde solution (PFA) at 4°C o/n. Dehydration, paraffin embedding, and sectioning (4  $\mu$ m)

was performed according to standard histology procedures. Hematoxylin and eosin (H&E) staining was applied to identify human engraftments.

#### Immunofluorescence (IF)

Rehydration in an ethanol series was followed by heat-mediated antigen retrieval using either tris buffer (pH 9) or citrate buffer (pH 6, both Vector Laboratories) for 20–25 min in a steamer. For IF, tissue permeabilization was performed with 0.5% Triton X-100/PBS (PBS-T) for 30 min at RT. After washing twice, primary antibodies diluted in Antibody Diluent (Zytomed) were added to the slides, which were then incubated o/n at 4°C in a wet chamber. After washing three times with PBS-T for 5 min, pancreatic tissue slides were quenched with TrueView (Vector Laboratories) for 5 min at RT. No quenching was performed for endocrine organoid sections. Next, slides were stained with Alexa Fluor secondary antibodies (Thermo) and 1 µg/ml DAPI diluted in Antibody Diluent for 90 min at RT in the dark. Slides were washed three times with PBS-T and finally with dH<sub>2</sub>O before sections were mounted with Vectashield Vibrance Antifade mounting medium (TrueView kit, different mounting medium might affect quenching). Following ABs were employed: AcTUB (Abcam, ab179484, rabbit, 1:1000, steamer citrate), CHGA (DAKO, M0869, mouse, 1:200, steamer citrate), KRT19 (DAKO, M0888, mouse 1:100, steamer tris), CPA1 (BioRad, AHP2054, rabbit, 1:1000, steamer tris), TRYP (Santa Cruz, sc-137077, mouse, 1:50, steamer tris), AFP (DAKO, A0008, rabbit, 1:500, steamer citrate), and CTRC (Millipore, MAB1476, mouse, 1:500, steamer citrate). For endocrine organoids, following ABs were used: PDX1 (R&D, AF2419, 1:500), NKX6-1 (DSHB, F55A12 (concentrate), 1:150), GCG (Sigma, G2654, 1:1000), C-pep (Cell Signaling, 4593, 1:200), and INS (Abcam, ab181547, 1:5000). IF images were acquired on an Axio Observer 7 microscope (Zeiss) equipped with a AxioCam 702 camera (Zeiss) using ZEN blue software (Zeiss).

#### Immunohistochemistry (IHC)

For immunohistochemistry (IHC) staining, antigen retrieval was directly followed by incubation with primary antibodies diluted in Antibody Diluent (Zytomed) for 30 min at RT in a wet chamber. Slides were washed and antibody detection was performed applying alkaline phosphatase-based Dako real Detection Kit (Agilent). In brief, slides were incubated first with a biotinylated secondary antibody, washed and next incubated with Streptavidin both for 30 min at RT. After another washing step, slides were incubated with Red Detection Dye for around 10 min, counterstained with 20% Hematoxylin solution (Merck) for 30 sec, washed with tap water for 5 min and finally mounted with Aquatex (Merck). For IHC of PDX1 raised in goat, a biotinylated secondary antibody against goat IgG (Vector Laboratories) was used in combination with Vectastain Elite ABC-HRP kit and Novared peroxidase substrate kit (both Vector Laboratories). Following Abs were employed: NKX6.1 (DSHB Hybridoma, F55A12, mouse, 1:150, steamer tris), CDX2 (Cell Marque, 235R-14, rabbit, 1:500, steamer tris), PDX1 (PDX1 (R&D, AF2419, goat, 250, steamer tris), C-pep (Cell Signaling, 4593, rabbit, 1:200, steamer citrate), GCG (Sigma, G2654, mouse, 1:1000, steamer citrate), P53 (Santa Cruz, sc-47698, mouse, 1:100, steamer citrate), P21 (Abcam, ab109520, rabbit, 1:300, steamer citrate), Ki-67 (DAKO, M7240, mouse, 1:500, pressure cooker), Human Nucleoli antibody (Abcam, ab190710, mouse, 1:200, pressure cooker). IHC images were acquired on a DM5500B microscope equipped with a DMC5400 camera (Leica) using the Application Suite software (Leica).

**ICC, IF, and IHC image analysis.** For ICC/IF and IF, brightness and contrast was adjusted in ZEN blue (Zeiss), for IHC in Photoshop (Adobe) according to good scientific practice. Images were then exported as jpeg files and cropped in Adobe Illustrator (Adobe) for final image compilation.

**Analysis of the PPDPF interactome.** The SILAC-based analysis of the PPDPF interactome was performed as previously described (PMID: 250s59612). Briefly,



HEK293T cells were metabolically labeled with lysine and arginine containing light (K0/R0) (Sigma; control cells) or heavy (K8/R10) carbon, nitrogen or hydrogen isotopes (Silantes) for ten passages to allow complete labeling of proteins. Labeled cells were transfected with N-terminally Strep/FLAG-tagged PPDPF or a vector control by lipophilic transfection via a home-made PEI reagent. To rule out labeling artifacts, labels were also switched between the conditions. The bait protein and associated proteins were purified via StrepTactin resin (IBA) and eluted with Desthiobiotin elution buffer (IBA). The eluates of the heavy and light condition were mixed 1:1 and subjected to mass spectrometric analysis. For this purpose, samples were precipitated by Methanol/Chloroform and cysteines were alkylated by DTT/Iodoacetamide prior to a tryptic proteolysis following standard conditions. After the proteolysis step, samples were desalted by StageTips (Thermo-Fisher) and analysed on a Q-Exactive Orbitrap mass spectrometer (Thermo-Fisher) coupled to a nano-flow high-performance liquid chromatography (HPLC) system (Dionex Ultimate 3000 RSLC, Thermo-Fisher). For separation, 180 min standard gradients were used. To obtain MS/MS spectra, the ten most intense precursor peptides were subjected to fragmentation (TOP10 method). Resulting RAW files were analysed by MaxQuant (ver. 1.6.2.10) setting K8/R10 as heavy and K0/R0 as light condition. The search was performed against the human subset of the Swissprot/UniprotKB database (release: 2016\_09, 23442 entries) using carbamylated cysteine as fixed and methionine oxidation and N-terminal acetylation as variable modifications. The statistical analysis was performed with Perseus (ver. 1.6.7.0).

**RNA isolation, reverse transcription and qPCR.** Cells were harvested with TrypLE as outlined above. RNA was extracted from cell pellets with the GeneJET RNA Purification Kit (Thermo) according to manufacturer's instructions. Reverse transcription of 200–1000 ng of total RNA was performed with the iScript cDNA Synthesis Kit (Bio-Rad) and cDNA was utilized for quantitative real-time PCR (qPCR) with SensiMix SYBR No-ROX Kit (Bioline) on the QuantStudio 3 Real-Time PCR System (Thermo). Following commercial QuantiTect primers (Qiagen) of target genes were employed: Hs\_SOX17\_1\_SG (QT00204099), Hs\_CXCR4\_2\_SG (QT02311841), Hs\_GATA6\_1\_SG (QT00233331), Hs\_GATA4\_1\_SG (QT00031997), Hs\_FOXA1\_1\_SG (QT00212828), Hs\_SOX9\_1\_SG (QT00001498), Hs\_PDX1\_1\_SG (QT00201859), Hs\_PROX1\_1\_SG (QT01006670), Hs\_NKX6-1\_1\_SG (QT00092379), Hs\_PTF1A\_2\_SG (QT01033396), Hs\_GP2\_1\_SG (QT00010535), Hs\_CPA1\_1\_SG (QT00001736), Hs\_PPDPF\_1\_SG (QT00203469). Target genes were normalized against the housekeeping gene Hs\_HMBS\_1\_SG (QT00014462) using  $2^{-\Delta\Delta C_t}$ .

## RNA-seq experiments

**RNA isolation, library preparation and sequencing.** RNA from hPSC, DE, PE and PP stage was isolated as described above. For DE, PE, and PP, RNA was isolated from four independent differentiations (differentiation started on different days). For hPSC control samples, 2 RNA samples were isolated. Quality and concentration were measured on an RNA Pico chip on an Agilent Bioanalyzer. Library preparation and sequencing was performed as detailed in [65]. Briefly, the Sequencing library preparation has been performed with the TrueSeq RNA Sample Prep Kit v2-Set B (RS-122–2002, Illumina Inc, San Diego, CA) producing 275 bp fragments including adapters in average size. Pooled libraries have then been clustered on the cBot Instrument from Illumina using the TruSeq SR Cluster Kit v3—cBot—HS(GD-401–3001, Illumina Inc, San Diego, CA). Sequencing was performed as 50 bp single reads and 7 bases index reads on an Illumina HiSeq2000 instrument using the TruSeq SBS Kit HSv3 (50-cycle) (FC-401–3002, Illumina Inc, San Diego, CA).

**Data processing.** RNA-seq reads were aligned to the human genome using STAR Aligner v.2.5.2a [66] with the Ensembl 86 reference genome. Sequenced read quality was checked with FastQC v0.11.2 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and alignment quality metrics were calculated using the RNASeQC v1.18 [67]. Following read alignment, duplication rates of the RNA-seq samples were computed with bamUtil v1.0.11 to mark duplicate reads and the dupRadar v1.4 Bioconductor R package for assessment [68]. Unique read counts were retrieved from the “featureCounts” software package [69]. As quality control measure, Reads Per Kilobase of transcript per Million mapped reads (RPKM) were quantified using Cufflinks software version 2.2.1 [70].

For quality control, total read counts, the rate of ribosomal RNA, the rate of intergenic and intronic reads and distribution of logarithmic RPKM values were assessed for each sample. While all samples except of one were of very good quality with 20–40 million 85 bp single end reads and very low ribosomal, intergenic, and intronic rates, one sample (PPDPF<sup>WT/WT</sup> cl. 1 at DE, replicate 3) was excluded from downstream analysis. Following samples were finally analyzed:

hESC/hPSC (day 0): 4 genotypes x 2 replicates = 8 samples

DE (day 4): 4 genotypes x 4 replicates - 1 sample exclusion = 15 samples

PE (day 10): 4 genotypes x 4 replicates = 16 samples

PP (day 14): 4 genotypes x 4 replicates = 16 samples

#### **RNA-seq data analysis.** Generation of DEG lists and transformed read counts

Raw read counts were imported to R v4.2.1 and differential expression was analyzed with “DESeq2” [71]. A gene was determined to be differentially expressed if the absolute log<sub>2</sub> FC > 1 and the adjusted P-value < 0.01. Read counts were Rlog transformed, before subjected to downstream statistical analysis.

#### PCA

Prior PCA analysis, Rlog transformed read counts were filtered for the top 10% most variant genes. For the PE and PP subset PCA, Rlog transformed read counts were first filtered for PE and PP samples before filtering for the top 10% most variant genes. Subsequently, PCA analysis was performed using the “prcomp” function and standard settings in R. PCA was visualized using “ggbiplot” function in conform package.

#### Hierarchical clustering

Sample distances have been calculated from Rlog transformed read counts using the “sample-Dists” function with default parameters such as Euclidean distance calculation.

#### Overrepresentation analysis

Overrepresentation analysis was conducted on pairwise DEG lists using the web browser tool of gProfiler [56].

#### Venn diagrams and Volcano plots

Venn diagrams were generated using “ggvenn” function and package, while “ggplot” was used for plotting Volcano plots.

#### Line plots and heatmaps

Line plots and heatmaps illustrating target gene expression were performed on Rlog transformed read counts. The Mean ± SEM was plotted with “ggplot” along the time course of differentiation for line plots. Heatmaps were generated using the packages (“pheatmap”) for plotting and the package (“biomaRt”) for retrieval of gene sets from common databases. The function (“pheatmap”) was used with parameter scale=“row” and clustering\_method=“ward.D2” with further parameters set as default.

### Gene set enrichment analysis (GSEA)

GSEA was conducted with the desktop version of GSEA version 4.0.3 (Broad Institute) [72]. Run GSEA function (not preranked function) was conducted on the Rlog transformed read count matrix and an additionally prepared data information file. Except of Collapse parameters (“No\_collaps”), default settings including 1000 permutations and a weighted statistical analysis were used. For comparison with reference gene sets, gene lists have been either directly compiled from literature or raw data has been reanalyzed. A complete list of applied gene sets can be found in [S5 Table](#).

### Generation of reference gene list

**Differentiation:** For stage-specific gene sets of hPSC-based pancreatic differentiations, two different RNA-seq data sets were utilized. The gene sets for Definitive endoderm\_Xie2013, Gute tube\_Xie2013, Pancreatic endoderm\_Xie2013, Pancreatic progenitor\_Xie2013 were directly retrieved from [27]. For Pluripotent stem cell\_Breunig2021 upregulated DEGs of the pairwise RNA-seq comparison between hPSC and DE from [23] were extracted, for Definitive endoderm\_Breunig2021 the downregulated DEGs of the same comparison and for Pancreatic progenitor\_Breunig2021 upregulated DEGs of the comparison PPs versus DE were listed. **Mouse Ptf1a Targets\_Thompson2012:** Gene sets from an RNA microarray of Ptf1a<sup>KO/KO</sup> versus Ptf1a<sup>KO/WT</sup> murine pancreatic progenitors could be directly retrieved from [31] (Ptf1a<sup>KO/KO</sup> up\_Thompson2012 and Ptf1a<sup>KO/KO</sup> down\_Thompson2012).

**Human Fetal Tissue\_Jennings2017:** A human fetal gene set was generated from raw FASTQ files (friendly provided by Neil Hanley) [29]. Briefly, FASTQ files have been mapped to the human genome, normalized using batch correction, and significant gene lists have been generated by Limma pairwise comparisons between all investigated organs. All genes found significant in at least on pairwise comparison were subsequently sorted according to their peaked expression to assign organ-specific gene lists (Liver, Extrahepatic biliary duct, Pancreas).

**Human Fetal Pancreas\_Gonçalves2021:** Gene lists were directly retrieved from cluster analysis in supplementary tables of [73] (Unknown1\_Gonçalves2021, Erythroblasts\_Gonçalves2021, Proliferative\_Gonçalves2021, Mesenchyme\_Gonçalves2021, Tip\_Gonçalves2021, Trunk\_Gonçalves2021, Endocrine\_Gonçalves2021, Neural\_Gonçalves2021).

**Human Fetal Tissue\_Cao2020:** Organ-specific pseudobulk expression profiles were retrieved from [S2 Table](#) of [30]. For final gene lists, genes with P-value < 0.01 and log2 FC > 121 were selected (Adrenal\_Cao2020, Placenta\_Cao2020, Cerebellum\_Cao2020, Lungs\_Cao2020, Liver\_Cao2020, Stomach\_Cao2020, Intestine\_Cao2020, Kidney\_Cao2020, Brain\_Cao2020, Muscle\_Cao2020, Eye\_Cao2020, Spleen\_Cao2020, Heart\_Cao2020, Pancreas\_Cao2020, Thymus\_Cao2020).

**Reanalysis of human fetal and adult pancreas scRNA-seq data.** The processed gene expression matrix and annotation data (including developmental time point, cell type annotation and UMAP coordinates) of the human fetal pancreas dataset [32] were downloaded from OMIX database under accession code [OMIX001616](#) and reanalyzed using Seurat v5.1.0 [74]. Counts were normalized using Seurat’s “NormalizeData” function. The expression of selected genes across developmental time points and cell types and within UMAP space was visualized using Seurat’s “VlnPlot” and “FeaturePlot” functions, respectively, and kernel density estimates of gene expression were plotted using “Nebulosa’s plot\_density” function.

The adult pancreas scRNA-seq data, downloaded from [GSE84133](#) has been previously reanalyzed as detailed here [23].

**ChIP-seq and ATAC-seq analysis.** ChIP-seq and ATAC bigwig files have been downloaded from publicly available data as described in the data availability section. Data has



been plotted at the PPDPF locus with the desktop version of Integrative Genomics Viewer (IGV) version 2.6.3 [54].

**PheWAS studies.** Association results were extracted from HugeAMP (<https://hugeamp.org>) [20], cis-eQTLs from (<https://eqtlgen.org/cis-eqtl.html>) [22], FORGEdb scores from <https://foragedb.cancer.gov> [55], and SNP frequencies in ethnical cohorts (only listed in S1 Table) from <https://www.ncbi.nlm.nih.gov/snp>.

**Figure preparation.** Adobe Illustrator was used for compilation and arrangement of multipaneled figures and for the generation of schematic views. The in Figs 9A and S6A depicted scheme of an orthotopic transplantation was produced with the help of the generic artificial intelligence function of Adobe Illustrator.

**Statistical analysis.** If not stated elsewhere, statistical analysis was performed using the GraphPad Prism 8 software and detailed information regarding the different applied tests are indicated in figure legends. In general, data summarize three independent experiments (independently started hPSC differentiations) with each analysis performed in duplicates (two wells per condition), unless otherwise stated. Statistical analyses was performed on the respective means of technical replicates. Statistical significance was defined as follows: \* P-value < 0.05, \*\* P-value < 0.01, \*\*\* P-value < 0.001, \*\*\*\* P-value < 0.0001. For differentiation experiments, independently started wells were considered independent experiments, while wells of the same differentiation were treated as technical replicates.

## Supporting Information

**S1 Fig. Regulation of PPDPF expression in human *in vitro* pancreatic differentiation.** (A) ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing)-based chromatin opening at PPDPF locus from our previous *in vitro* differentiation of the hESC line HUES8 [13]. (B) Histone marks along the time course of differentiation in CyT49 differentiation complementing main Fig 2F. (C) Chromatin immunoprecipitation (ChIP)-seq binding peaks of murine Foxa2, Ptf1a, and Rbpjl from publicly available ChIP-seq data at E17.5 [75]. ATAC- and ChIP-seq peaks have been visualized via Integrative Genomics Viewer (IGV) [54]. (TIF)

**S2 Fig. CRISPR/Cas9 gene editing of PPDPF in hPSCs.** (A) Scheme of the applied four times CRISPR/Cas9 nicking strategy to delete exon 3 of PPDPF. (B) PCR-based clone screening after gene editing in the hESC line HUES8 highlighting the two employed PPDPF<sup>KO/KO</sup> clones. Note that the PCR band size, which was generated with primers spanning the target region, decreases after successful gene editing (from around 550 bp to around 400 bp). (C) Sanger sequencing of the two PPDPF<sup>KO/KO</sup> clones. Both clones harbored identical homozygous mutations. (D) Predicted alterations on protein level caused by the induced genetic deletion. (E) PPDPF mRNA expression relative to HMBS in PPDPF<sup>KO/KO</sup> and PPDPF<sup>WT/WT</sup> PP (n = 3; Mean ± SEM; ordinary one-way ANOVA followed by Dunett's multiple comparison test; cl.: clone). (F) Western Blot at PE stage confirming KO of PPDPF on protein level (200 µg, n = 2). (G) Flow cytometry analysis of pluripotency marker SSEA4 and TRA1-60 in PPDPF<sup>KO/KO</sup> and PPDPF<sup>WT/WT</sup> hESCs (n = 1, in duplicates; Mean ± SD). (H) ICC/IF staining of pluripotency marker OCT4 and SOX2 in PPDPF<sup>KO/KO</sup> and PPDPF<sup>WT/WT</sup> hPSCs (scale bar: 100 µm, n = 1). (TIF)

**S3 Fig. Off-target analysis in gene edited clones.** (A) Overlap between CRISPR [24] predicted potential off-targets of different guides. Two nicks on complementary strands and in close proximity are required for induction of a double strand break (DSB). (B) Overlap between predicted potential off-targets and RNA-seq DEGs. Only 2 genes were found, of

which expression was significantly reduced in at least 3 stages. (C) Line plots depicting Mean  $\pm$  SEM of the two genes, *MGMT* and *TTC34*, with (D) subsequent sanger sequencing on the potential off-target sites. Sequences were unaltered in all clones. One representative PPDPF<sup>WT</sup> and PPDPF<sup>KO/KO</sup> hESC clone sequence is shown. (E) Exemplary line plots of genes (Mean  $\pm$  SEM), in which potential off-targets were shared between guide1a and 1b, respectively 2a and 2b. Also, other pairwise comparisons (e.g. guide 1a with 2a) were investigated with no indications for a loss- or gain-of-transcript mutation. (TIF)

#### S4 Fig. Depletion of pancreatic identity in PP differentiation upon PPDPF ablation.

Gene set enrichment analysis (GSEA) for gene sets compiled from (A) an RNA microarray of Ptf1a<sup>KO/KO</sup> versus Ptf1a<sup>KO/WT</sup> murine pancreatic progenitors [31], (B) bulk RNA-seq from human fetal pancreas [29], (C) scRNA-seq from human fetal pancreas [73] and (D) scRNA-seq from human fetal tissues [30]. For all: False discovery rates (FDRs)  $\leq 0.25$  are highlighted in red (left panel), while comparisons that were enriched in PPDPF<sup>KO/KO</sup> cells are depicted in green and comparisons that were depleted in blue (right panel). GSEA plots against depicted reference gene sets are shown for the comparison of PPDPF<sup>WT/WT</sup> and PPDPF<sup>KO/KO</sup> PPs (only for proliferative progenitor of Gonçalves PE plotted). Line plots (Mean  $\pm$  SEM) are displayed for genes, top ranked in GSEA. Additional information: For “PTF1A<sup>KO/KO</sup> up<sup>Thompson 2012</sup>”, 4 out of 64 genes were significantly upregulated (DEGs) in PPDPF<sup>KO/KO</sup> PPs over PPDPF<sup>WT/WT</sup> PPs, while 0 were downregulated. For “PTF1A<sup>KO/KO</sup> down<sup>Thompson 2012</sup>”, 0 out of 66 genes were up, 1 was down. For “liver<sup>Jennings 2017</sup>”, 4 out of 79 genes were up, 2 were down. For “Extrahepatic biliary duct<sup>Jennings 2017</sup>”, 1 out of 32 genes was up, 5 down. For “Pancreas<sup>Jennings 2017</sup>”, 0 out of 32 genes were up, 3 down. For “Proliferative/Progenitor<sup>Gonçalves 2021</sup>”, 2 out of 195 genes were up, 1 was down (at PE stage). For “Tip<sup>Gonçalves 2021</sup>”, 0 out of 111 genes were up, 1 was down. For “Trunk<sup>Gonçalves 2021</sup>”, 0 out of 55 genes were up, 2 were down. For “Placenta Pancreas<sup>Cao 2020</sup>”, 20 out of 307 genes were up, 4 were down. For “Intestine Pancreas<sup>Cao 2020</sup>”, 7 out of 239 genes were up, 0 down. For “Pancreas<sup>Cao 2020</sup>”, 0 out of 107 gene were up, 1 was down. (TIF)

**S5 Fig. scRNA-seq analysis of human fetal and adult pancreas.** UMAP representations and violin plots for expression of (A) progenitor marker *PTF1A*, (B) *PDX1*, (C) *FOXA1*, (D) *FOXA2*, (E) acinar marker *CEL*, (F) ductal marker *CFTR*, (G) endocrine marker *CHGA*, and (H) trunk (and tip) marker *TTYH1* in human fetal pancreas [32] complementing main Fig. 8A–D. (I) Expression of *PPDPF* and selected cell type marker in human adult pancreas [33] revealing relatively broad expression of *PPDPF*. EP: endocrine progenitor; MP: multipotent progenitor; UMAP: Uniform Manifold Approximation and Projection. (TIF)

**S6 Fig: PPDPF<sup>KO/KO</sup> PPs successfully develop into pancreatic tissue upon orthotopic transplantation.** (A) Schematic overview of orthotopic transplantation experiments and flow cytometry-based quality control of GP2 sorting ( $n = 1$ ). (B) Hematoxylin and eosin (H&E) staining of identified grafts with  $n = 2$  for PPDPF<sup>WT/WT</sup> PPs and  $n = 3$  for PPDPF<sup>KO/KO</sup> cl.1 and cl.2 PPs. Unsorted planar PPs were used when cell numbers after GP2 sort were not sufficient. (C) Human nucleoli (H-NUCL) staining illustrates human engraftment. (D) Quantification of depicted marker expression averaged over individual grafts of each genotype (– completely or nearly completely absent; + weak expression; ++ moderate expression; +++ strong expression). (E) Immunofluorescence (IF) staining of acinar marker CPA1, Trypsin, and CTRC, ductal marker CK19, and liver marker AFP. (F) IF and immunohistochemistry (IHC) of ductal cilia marker acetylated tubulin (acTUB), endocrine marker CHGA, GCG, and C-peptide (C-pep).

(G) IHC of progenitor marker PDX1, NKX6-1, and CDX2. (H) IHC of proliferation marker Ki-67 and effectors of growth restriction P53 and P21. For all: Images are depicted that show strong marker expression to highlight trilineage competence. Scale bar: B-C: 500  $\mu$ m, inlet: 100  $\mu$ m. E-H: 100 $\mu$ m, inlets 2x enlarged.  
(TIF)

**S1 Table. GWAS traits, frequency of SNPs, and information about credible set.**  
(XLSX)

**S2 Table. List of PPDPF-bound proteins identified in HEK293T cells via SILAC and gProfiler of PE/PP intersected proteins.**  
(XLSX)

**S3 Table. Stage-specific lists of differentially regulated genes (DEGs).**  
(XLSX)

**S4 Table. Stage-specific overrepresentation analysis using gProfiler.**  
(XLSX)

**S5 Table. Developmental gene sets compiled from literature.**  
(XLSX)

**S6 Table. Values behind depicted graphs.**  
(XLSX)

## Acknowledgements

The authors thank Katrin Jochmann, Katrin Köhn, Ulrike Mayr-Beyrle, and Ralf Köhntop for excellent technical assistance, and Justin Antony, Michael Kormann (University of Tuebingen), Franz Oswald, Sabine Schirmer (Ulm, University), Ninel Azoitei, Johann Gout, Elodie Roger, Michael Karl Melzer, and Sandra Heller (IMOS) for helpful project discussions. They acknowledge Rachel E. Jennings and Neil A. Hanley (University of Manchester) for providing FASTQ raw data on human fetal RNA-seq and thank Thomas Engleitner, Gaurav Jain, and Roland Rad (Technical University Munich) for their support in reanalysis of this data set and the additional human adult scRNA-seq pancreas data set. They thank Ivan Costa, Nicole Zhang, and Kevin Menden for providing input on RNA-seq data evaluation and O.D. Madsen and the Developmental Studies Hybridoma Bank (DSHB) for providing the NKX6-1 antibody.

## Author contributions

**Conceptualization:** Markus Breunig, Stefan Liebau, Alexander Kleger.

**Data curation:** Markus Breunig, Meike Hohwieler, Eleni Zimmer, Medhanie A. Mulaw, Eric Simon.

**Formal analysis:** Markus Breunig, Meike Hohwieler, Felix von Zweydorf, Natalie Hauff, Cécile Julier, Christian Johannes Gloeckner.

**Funding acquisition:** Markus Breunig, Christian Johannes Gloeckner, Stefan Liebau, Alexander Kleger.

**Investigation:** Markus Breunig, Meike Hohwieler, Jasmin Haderspeck, Felix von Zweydorf, Natalie Hauff, Lino-Pascal Pasquini, Christoph Wiegrefe.

**Project administration:** Markus Breunig, Stefan Liebau, Alexander Kleger.



**Resources:** Christoph Wiegrefe, Medhanie A. Mulaw, Cécile Julier, Eric Simon, Christian Johannes Gloeckner, Stefan Liebau, Alexander Kleger.

**Software:** Markus Breunig, Eleni Zimmer, Medhanie A. Mulaw, Eric Simon.

**Supervision:** Markus Breunig, Meike Hohwieler, Christian Johannes Gloeckner, Stefan Liebau, Alexander Kleger.

**Validation:** Markus Breunig.

**Visualization:** Markus Breunig.

**Writing – original draft:** Markus Breunig.

**Writing – review & editing:** Meike Hohwieler, Jasmin Haderspeck, Lino-Pascal Pasquini, Christoph Wiegrefe, Eleni Zimmer, Cécile Julier, Stefan Liebau, Alexander Kleger.

## References

1. Jiang Z, Song J, Qi F, Xiao A, An X, Liu N, et al. Exdpf is a key regulator of exocrine pancreas development controlled by retinoic acid and ptf1a in zebrafish. *PLoS Biol.* 2008;6(11):e293. <https://doi.org/10.1371/journal.pbio.0060293> PMID: 19067490
2. Liu G, Shi H, Deng L, Zheng H, Kong W, Wen X, et al. Circular RNA circ-FOXN1 facilitates cell progression as ceRNA to target PPDPF and MACC1 by sponging miR-1304-5p in non-small cell lung cancer. *Biochem Biophys Res Commun.* 2019;513(1):207–12. <https://doi.org/10.1016/j.bbrc.2019.03.213> PMID: 30954221
3. Yun M, Yingzi L, Jie G, Guanxin L, Zimei Z, Zhen C, et al. PPDPF promotes the progression and acts as an antiapoptotic protein in non-small cell lung cancer. *Int J Biol Sci.* 2022;18(1):214–28. <https://doi.org/10.7150/ijbs.65654> PMID: 34975328
4. Zheng Q-W, Ni Q-Z, Zhu B, Liang X, Ma N, Wang Y-K, et al. PPDPF promotes lung adenocarcinoma progression via inhibiting apoptosis and NK cell-mediated cytotoxicity through STAT3. *Oncogene.* 2022;41(36):4244–56. <https://doi.org/10.1038/s41388-022-02418-3> PMID: 35906391
5. Xiao Y, Lai Y, Yu Y, Jiang P, Li Y, Wang C, et al. The exocrine differentiation and proliferation factor (EXDPF) gene promotes ovarian cancer tumorigenesis by up-regulating DNA replication pathway. *Front Oncol.* 2021;11:669603. <https://doi.org/10.3389/fonc.2021.669603> PMID: 34041032
6. Zi Y, Liu L, Gao J, Xu X, Guan Y, Rong Z, et al. Phosphorylation of PPDPF via IL6-JAK2 activates the Wnt/ $\beta$ -catenin pathway in colorectal cancer. *EMBO Rep.* 2023;24(9):e55060. <https://doi.org/10.15252/embr.202255060> PMID: 37477088
7. Ni Q-Z, Zhu B, Ji Y, Zheng Q-W, Liang X, Ma N, et al. PPDPF Promotes the development of mutant KRAS-driven pancreatic ductal adenocarcinoma by regulating the GEF activity of SOS1. *Adv Sci (Weinh).* 2023;10(2):e2202448. <https://doi.org/10.1002/adv.202202448> PMID: 36453576
8. Wang Y-K, Ma N, Xu S, Huang J-Y, Ni Q-Z, Cao H-J, et al. PPDPF suppresses the development of hepatocellular carcinoma through TRIM21-mediated ubiquitination of RIPK1. *Cell Rep.* 2023;42(4):112340. <https://doi.org/10.1016/j.celrep.2023.112340> PMID: 37027301
9. Mao Z, Li X, Ma X, Wang X, Zhang J, Fan X. Pancreatic progenitor cell differentiation and proliferation factor predicts poor prognosis in hepatocellular carcinoma. *Medicine (Baltimore).* 2019;98(9):e14552. <https://doi.org/10.1097/MD.00000000000014552> PMID: 30817571
10. Vittrant B, Leclercq M, Martin-Magniette ML, Collins C, Bergeron A, Fradet Y, et al. Identification of a transcriptomic prognostic signature by machine learning using a combination of small cohorts of prostate cancer. *Front Genet.* 2020;11:550894. <https://doi.org/10.3389/fgene.2020.550894> PMID: 33324443
11. Zorn AM, Wells JM. Vertebrate endoderm development and organ formation. *Annu Rev Cell Dev Biol.* 2009;25:221–51. <https://doi.org/10.1146/annurev.cellbio.042308.113344> PMID: 19575677
12. Kim M-S, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaekady R, et al. A draft map of the human proteome. *Nature.* 2014;509(7502):575–81. <https://doi.org/10.1038/nature13302> PMID: 24870542
13. Heller S, Li Z, Lin Q, Geusz R, Breunig M, Hohwieler M. Transcriptional changes and the role of ONECUT1 in hPSC pancreatic differentiation. *Communications Biology.* 2021;4(1):1298. <https://doi.org/10.1038/s42003-021-02729-5>
14. Hohwieler M, Illing A, Hermann PC, Mayer T, Stockmann M, Perkhof L, et al. Human pluripotent stem cell-derived acinar/ductal organoids generate human pancreas upon orthotopic transplantation

- and allow disease modelling. *Gut*. 2017;66(3):473–86. <https://doi.org/10.1136/gutjnl-2016-312423> PMID: [27633923](#)
15. Krüger J, Breunig M, Pasquini LP, Morawe M, Groß A, Arnold F, et al. Functional genomic screening in human pluripotent stem cells reveals new roadblocks in early pancreatic endoderm formation. *Cells*. 2022;11(3):582. <https://doi.org/10.3390/cells11030582> PMID: [35159392](#)
  16. Philippi A, Heller S, Costa I, Senée V, Breunig M, Li Z, et al. Mutations and variants of ONECUT1 in diabetes. *Nature Medicine*. 2021;27(11):1928–40. <https://doi.org/10.1038/s41591-021-01502-7> PMID: [34663987](#)
  17. Wang A, Yue F, Li Y, Xie R, Harper T, Patel NA, et al. Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates. *Cell Stem Cell*. 2015;16(4):386–99. <https://doi.org/10.1016/j.stem.2015.02.013> PMID: [25842977](#)
  18. Geusz RJ, Wang A, Chiou J, Lancman JJ, Wetton N, Kefalopoulou S, et al. Pancreatic progenitor epigenome maps prioritize type 2 diabetes risk genes with roles in development. *Elife*. 2021;10:e59067. <https://doi.org/10.7554/eLife.59067> PMID: [33544077](#)
  19. Geusz RJ, Wang A, Lam DK, Vinckier NK, Alysandratos K-D, Roberts DA, et al. Sequence logic at enhancers governs a dual mechanism of endodermal organ fate induction by FOXA pioneer factors. *Nat Commun*. 2021;12(1):6636. <https://doi.org/10.1038/s41467-021-26950-0> PMID: [34789735](#)
  20. Common Metabolic Diseases Knowledge Portal (CMDKP) [Internet]. 2024 [cited 2024-05-08]. Available from: <https://md.hugeamp.org> (RRID: SCR\_020937)
  21. Lagou V, Jiang L, Ulrich A, Zudina L, González KSG, Balkhiyarova Z, et al. GWAS of random glucose in 476,326 individuals provide insights into diabetes pathophysiology, complications and treatment stratification. *Nat Genet*. 2023;55(9):1448–61. <https://doi.org/10.1038/s41588-023-01462-3> PMID: [37679419](#)
  22. Vösa U, Claringbould A, Westra H-J, Bonder MJ, Deelen P, Zeng B, et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet*. 2021;53(9):1300–10. <https://doi.org/10.1038/s41588-021-00913-z> PMID: [34475573](#)
  23. Breunig M, Merkle J, Wagner M, Melzer M, Barth T, Engleitner T. Modeling plasticity and dysplasia of pancreatic ductal organoids derived from human pluripotent stem cells. *Cell Stem Cell*. 2021;28(6):1105–24. <https://doi.org/10.1016/j.stem.2021.03.005> PMID: [33915078](#)
  24. Haeussler M, Schöning K, Eckert H, Eschstruth A, Mianné J, Renaud J-B, et al. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol*. 2016;17(1):148. <https://doi.org/10.1186/s13059-016-1012-2> PMID: [27380939](#)
  25. Ye Y, Shen A, Liu A. Long non-coding RNA H19 and cancer: A competing endogenous RNA. *Bull Cancer*. 2019;106(12):1152–9. <https://doi.org/10.1016/j.bulcan.2019.08.011> PMID: [31753509](#)
  26. He H, Kim J. Regulation and function of the peg3 imprinted domain. *Genomics Inform*. 2014;12(3):105–13. <https://doi.org/10.5808/GI.2014.12.3.105> PMID: [25317109](#)
  27. Xie R, Everett LJ, Lim H-W, Patel NA, Schug J, Kroon E, et al. Dynamic chromatin remodeling mediated by polycomb proteins orchestrates pancreatic differentiation of human embryonic stem cells. *Cell Stem Cell*. 2013;12(2):224–37. <https://doi.org/10.1016/j.stem.2012.11.023> PMID: [23318056](#)
  28. Gerrard DT, Berry AA, Jennings RE, Piper Hanley K, Bobola N, Hanley NA. An integrative transcriptomic atlas of organogenesis in human embryos. *Elife*. 2016;5:e15657. <https://doi.org/10.7554/eLife.15657> PMID: [27557446](#)
  29. Jennings RE, Berry AA, Gerrard DT, Wearne SJ, Strutt J, Withey S, et al. Laser capture and deep sequencing reveals the transcriptomic programmes regulating the onset of pancreas and liver differentiation in human embryos. *Stem Cell Reports*. 2017;9(5):1387–94. <https://doi.org/10.1016/j.stemcr.2017.09.018> PMID: [29056335](#)
  30. Cao J, O'Day DR, Pliner HA, Kingsley PD, Deng M, Daza RM, et al. A human cell atlas of fetal gene expression. *Science*. 2020;370(6518):eaba7721. <https://doi.org/10.1126/science.aba7721> PMID: [33184181](#)
  31. Thompson N, Gésina E, Scheinert P, Bucher P, Grapin-Botton A. RNA profiling and chromatin immunoprecipitation-sequencing reveal that PTF1a stabilizes pancreas progenitor identity via the control of MNX1/HLXB9 and a network of other transcription factors. *Mol Cell Biol*. 2012;32(6):1189–99. <https://doi.org/10.1128/MCB.06318-11> PMID: [22232429](#)
  32. Ma Z, Zhang X, Zhong W, Yi H, Chen X, Zhao Y, et al. Deciphering early human pancreas development at the single-cell level. *Nat Commun*. 2023;14(1):5354. <https://doi.org/10.1038/s41467-023-40893-8> PMID: [37660175](#)

33. Baron M, Veres A, Wolock SL, Faust AL, Gaujoux R, Vetere A, et al. A Single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure. *Cell Syst*. 2016;3(4):346–360.e4. <https://doi.org/10.1016/j.cels.2016.08.011> PMID: 27667365
34. Merz S, Breunig M, Melzer MK, Heller S, Wiedenmann S, Seufferlein T, et al. Single-cell profiling of GP2-enriched pancreatic progenitors to simultaneously create acinar, ductal, and endocrine organoids. *Theranostics*. 2023;13(6):1949–73. <https://doi.org/10.7150/thno.78323> PMID: 37064874
35. Krüger J, Fischer A, Breunig M, Allgöwer C, Schulte L, Merkle J. DNA methylation-associated allelic inactivation regulates Keratin 19 gene expression during pancreatic development and carcinogenesis. *J Pathol*. 2023;261(2):139–55. <https://doi.org/10.1002/path.6156> PMID: 37555362
36. Merz S, Senée V, Philippi A, Oswald F, Shaigan M, Führer M. A ONECUT1 regulatory, non-coding region in pancreatic development and diabetes. *Cell Rep*. 2024;43(11):114853. <https://doi.org/10.1016/j.celrep.2024.114853> PMID: 39427318
37. Balboa D, Barsby T, Lithovius V, Saarimäki-Vire J, Omar-Hmeadi M, Dyachok O. Functional, metabolic and transcriptional maturation of human pancreatic islets derived from stem cells. *Nature Biotechnology*. 2022;40(7):1042–55. <https://doi.org/10.1101/2021.03.31.437748> PMID: 35444444
38. Friedman JR, Kaestner KH. The Foxa family of transcription factors in development and metabolism. *Cell Mol Life Sci*. 2006;63(19–20):2317–28. <https://doi.org/10.1007/s00018-006-6095-6> PMID: 16909212
39. Gredler ML, Patterson SE, Seifert AW, Cohn MJ. Foxa1 and Foxa2 orchestrate development of the urethral tube and division of the embryonic cloaca through an autoregulatory loop with Shh. *Developmental Biology*. 2020;465(1):23–30. <https://doi.org/10.1016/j.ydbio.2020.06.009> PMID: 32645357
40. Masui T, Swift GH, Deering T, Shen C, Coats WS, Long Q, et al. Replacement of Rbpj with Rbpjl in the PTF1 complex controls the final maturation of pancreatic acinar cells. *Gastroenterology*. 2010;139(1):270–80. <https://doi.org/10.1053/j.gastro.2010.04.003> PMID: 20398665
41. Jennings RE, Berry AA, Strutt JP, Gerrard DT, Hanley NA. Human pancreas development. *Development*. 2015;142(18):3126–37. <https://doi.org/10.1242/dev.120063> PMID: 26395141
42. Duque M, Amorim JP, Bessa J. Ptf1a function and transcriptional cis-regulation, a cornerstone in vertebrate pancreas development. *FEBS J*. 2022;289(17):5121–36. <https://doi.org/10.1111/febs.16075> PMID: 34125483
43. Field HA, Dong PDS, Beis D, Stainier DYR. Formation of the digestive system in zebrafish. II. Pancreas morphogenesis. *Dev Biol*. 2003;261(1):197–208. [https://doi.org/10.1016/s0012-1606\(03\)00308-7](https://doi.org/10.1016/s0012-1606(03)00308-7) PMID: 12941629
44. Matsuda H. Zebrafish as a model for studying functional pancreatic  $\beta$  cells development and regeneration. *Dev Growth Differ*. 2018;60(6):393–9. <https://doi.org/10.1111/dgd.12565> PMID: 30133710
45. Lin JW, Biankin AV, Horb ME, Ghosh B, Prasad NB, Yee NS, et al. Differential requirement for ptf1a in endocrine and exocrine lineages of developing zebrafish pancreas. *Dev Biol*. 2004;274(2):491–503. <https://doi.org/10.1016/j.ydbio.2004.07.001> PMID: 15570689
46. Sheng X, Guan Y, Ma Z, Wu J, Liu H, Qiu C, et al. Mapping the genetic architecture of human traits to cell types in the kidney identifies mechanisms of disease and potential treatments. *Nat Genet*. 2021;53(9):1322–33. <https://doi.org/10.1038/s41588-021-00909-9> PMID: 34385711
47. Ma N, Wang Y-K, Xu S, Ni Q-Z, Zheng Q-W, Zhu B, et al. PPDPF alleviates hepatic steatosis through inhibition of mTOR signaling. *Nat Commun*. 2021;12(1):3059. <https://doi.org/10.1038/s41467-021-23285-8> PMID: 34031390
48. Li YC, Cai SW, Shu YB, Chen MW, Shi Z. USP15 in cancer and other diseases: From diverse functions to therapeutic targets. *Biomedicines*. 2022;10(2).
49. Cao Y, Di X, Zhang Q, Li R, Wang K. RBM10 regulates tumor apoptosis, proliferation, and metastasis. *Front Oncol*. 2021;11:603932. <https://doi.org/10.3389/fonc.2021.603932> PMID: 33718153
50. Inoue A. RBM10: Structure, functions, and associated diseases. *Gene*. 2021;783:145463. <https://doi.org/10.1016/j.gene.2021.145463> PMID: 33515724
51. Mulvaney KM, Blomquist C, Acharya N, Li R, Ranaghan MJ, O'Keefe M, et al. Molecular basis for substrate recruitment to the PRMT5 methylosome. *Mol Cell*. 2021;81(17):3481–3495.e7. <https://doi.org/10.1016/j.molcel.2021.07.019> PMID: 34358446
52. Feustel K, Falchook GS. Protein arginine methyltransferase 5 (PRMT5) inhibitors in oncology clinical trials: a review. *J Immunother Precis Oncol*. 2022;5(3):58–67. <https://doi.org/10.36401/JIPO-22-1> PMID: 36034581



53. Zheng J, Li B, Wu Y, Wu X, Wang Y. Targeting arginine methyltransferase PRMT5 for cancer therapy: updated progress and novel strategies. *J Med Chem*. 2023;66(13):8407–27. <https://doi.org/10.1021/acs.jmedchem.3c00250> PMID: 37366223
54. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 2012;14(2):178–92. <https://doi.org/10.1093/bib/bbs017> PMID: 22517427
55. Breeze CE, Haugen E, Gutierrez-Arcelus M, Yao X, Teschendorff A, Beck S, et al. FORGEdb: a tool for identifying candidate functional variants and uncovering target genes and mechanisms for complex diseases. *Genome Biol*. 2024;25(1):3. <https://doi.org/10.1186/s13059-023-03126-1> PMID: 38167104
56. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res*. 2019;47(W1):W191–8. <https://doi.org/10.1093/nar/gkz369> PMID: 31066453
57. 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*. 2015;32:171–80. <https://doi.org/10.1016/j.gde.2015.03.004> PMID: 25909383
58. Koya V, Lu S, Sun Y-P, Purich DL, Atkinson MA, Li S-W, et al. Reversal of streptozotocin-induced diabetes in mice by cellular transduction with recombinant pancreatic transcription factor pancreatic duodenal homeobox-1: a novel protein transduction domain-based therapy. *Diabetes*. 2008;57(3):757–69. <https://doi.org/10.2337/db07-1441> PMID: 18086901
59. Schaeren-Wiemers N, Gerfin-Moser A. A single protocol to detect transcripts of various types and expression levels in neural tissue and cultured cells: in situ hybridization using digoxigenin-labelled cRNA probes. *Histochemistry*. 1993;100(6):431–40. <https://doi.org/10.1007/BF00267823> PMID: 7512949
60. Frank S, Zhang M, Schöler HR, Greber B. Small molecule-assisted, line-independent maintenance of human pluripotent stem cells in defined conditions. *PLoS One*. 2012;7(7):e41958. <https://doi.org/10.1371/journal.pone.0041958> PMID: 22860038
61. Cogger KF, Sinha A, Sarangi F, McGaugh EC, Saunders D, Dorrell C, et al. Glycoprotein 2 is a specific cell surface marker of human pancreatic progenitors. *Nat Commun*. 2017;8(1):331. <https://doi.org/10.1038/s41467-017-00561-0> PMID: 28835709
62. Aghazadeh Y, Sarangi F, Poon F, Nkennor B, McGaugh EC, Nunes SS, et al. GP2-enriched pancreatic progenitors give rise to functional beta cells in vivo and eliminate the risk of teratoma formation. *Stem Cell Reports*. 2022;17(4):964–78. <https://doi.org/10.1016/j.stemcr.2022.03.004> PMID: 35364010
63. Marzenke M, Fell J, Piccini I, Röpke A, Seeböhm G, Greber B. Generation and cardiac subtype-specific differentiation of PITX2-deficient human iPS cell lines for exploring familial atrial fibrillation. *Stem Cell Res*. 2017;21:26–8. <https://doi.org/10.1016/j.scr.2017.03.015> PMID: 28677534
64. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science*. 2013;339(6121):819–23. <https://doi.org/10.1126/science.1231143> PMID: 23287718
65. Söllner JF, Leparc G, Hildebrandt T, Klein H, Thomas L, Stupka E, et al. An RNA-Seq atlas of gene expression in mouse and rat normal tissues. *Sci Data*. 2017;4:170185. <https://doi.org/10.1038/sdata.2017.185> PMID: 29231921
66. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21. <https://doi.org/10.1093/bioinformatics/bts635> PMID: 23104886
67. DeLuca DS, Levin JZ, Sivachenko A, Fennell T, Nazaire M-D, Williams C, et al. RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics*. 2012;28(11):1530–2. <https://doi.org/10.1093/bioinformatics/bts196> PMID: 22539670
68. Sayols S, Scherzinger D, Klein H. dupRadar: a Bioconductor package for the assessment of PCR artifacts in RNA-Seq data. *BMC Bioinformatics*. 2016;17(1):428. <https://doi.org/10.1186/s12859-016-1276-2> PMID: 27769170
69. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923–30. <https://doi.org/10.1093/bioinformatics/btt656> PMID: 24227677
70. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol*. 2013;31(1):46–53. <https://doi.org/10.1038/nbt.2450> PMID: 23222703

71. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550. <https://doi.org/10.1186/s13059-014-0550-8> PMID: [25516281](https://pubmed.ncbi.nlm.nih.gov/25516281/)
72. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102(43):15545–50. <https://doi.org/10.1073/pnas.0506580102> PMID: [16199517](https://pubmed.ncbi.nlm.nih.gov/16199517/)
73. Gonçalves CA, Larsen M, Jung S, Stratmann J, Nakamura A, Leuschner M, et al. A 3D system to model human pancreas development and its reference single-cell transcriptome atlas identify signaling pathways required for progenitor expansion. *Nat Commun.* 2021;12(1):3144. <https://doi.org/10.1038/s41467-021-23295-6> PMID: [34035279](https://pubmed.ncbi.nlm.nih.gov/34035279/)
74. Hao Y, Stuart T, Kowalski M, Choudhary S, Hoffman P, Hartman A. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nature Biotechnology.* 2024;42(2):293–304. <https://doi.org/10.1038/s41587-023-01767-y> PMID: [37231261](https://pubmed.ncbi.nlm.nih.gov/37231261/)
75. Meredith DM, Borromeo MD, Deering TG, Casey BH, Savage TK, Mayer PR, et al. Program specificity for Ptf1a in pancreas versus neural tube development correlates with distinct collaborating cofactors and chromatin accessibility. *Mol Cell Biol.* 2013;33(16):3166–79. <https://doi.org/10.1128/MCB.00364-13> PMID: [23754747](https://pubmed.ncbi.nlm.nih.gov/23754747/)