Transcriptome analysis of pancreatic cells across distant species highlights novel important regulator genes.

Estefania Tarifeño-Saldivia, Arnaud Lavergne, Alice Bernard, Keerthana Padamata, David Bergemann, Marianne L. Voz, Isabelle Manfroid, Bernard Peers*.

E-mail : Estefania Tarifeño-Saldivia : etarisal@gmail.com; Arnaud Lavergne : arnaud.lavergne@student.ulg.ac.be; Alice Bernard : alice.bernard@student.ulg.ac.be; Keerthana Padamata : keerthanaperike@googlemail.com; David Bergemann : david.bergemann@student.ulg.ac.be; Marianne L. Voz : mvoz@ulg.ac.be; Isabelle Manfroid : Isabelle.Manfroid@ulg.ac.be; Bernard Peers : bpeers@ulg.ac.be;

Laboratory of Zebrafish Development and Disease Models (ZDDM), GIGA, Avenue de l'Hôpital 1, B34, 4000 SART TILMAN, University of Liège, Belgium

* Corresponding author ; E-mail : bpeers@ulg.ac.be

Abstract

Background

Defining the transcriptome and the genetic pathways of pancreatic cells is of great interest for elucidating the molecular attributes of pancreas disorders such as diabetes and cancer. As the function of the different pancreatic cell types has been maintained during vertebrate evolution, the comparison of their transcriptomes across distant vertebrate species is a mean to pinpoint genes under strong evolutionary constrains due to their crucial function, which have therefore preserved their selective expression in these pancreatic cell types.

Results

In this study, RNA-sequencing was performed on pancreatic alpha-, beta- and delta endocrine cells as well as the acinar and ductal exocrine cells isolated from adult zebrafish transgenic lines. Comparison of these transcriptomes identified many novel markers, including transcription factors and signaling pathways components, specific for each cell type. By performing interspecies comparisons, we identified hundreds of genes with conserved enriched expression in endocrine and exocrine cells among human, mouse and zebrafish. This list includes many genes known as crucial for pancreatic cell formation or function, but also pinpoints many factors whose pancreatic function is still unknown. A large set of endocrine-enriched genes can already be detected at early developmental stages as revealed by the transcriptomic profiling of embryonic endocrine cells, indicating a potential role in cell differentiation. The actual involvement of conserved endocrine genes in pancreatic cell differentiation was demonstrated in zebrafish for *myt1b*, whose invalidation leads to a reduction of alpha-cells, and for *cdx4,* selectively expressed in endocrine delta-cells and crucial for their specification. Intriguingly, comparison of the endocrine alpha- and beta-cell subtypes from

human, mouse and zebrafish reveals a much lower conservation of the transcriptomic signatures for these two endocrine cell subtypes compared to the signatures of pan-endocrine and exocrine cells. These data suggest that the identity of the alpha- and beta-cells relies on a few key factors, corroborating numerous examples of inter-conversion between these two endocrine cell subtypes.

Conclusion.

This study highlights both evolutionary conserved and species-specific features that will help to unveil universal and fundamental regulatory pathways as well as pathways specific to human and laboratory animal models such as mouse and zebrafish.

Keywords:

RNA-seq, comparative transcriptomics, pancreas, endocrine cells, acinar cells, ductal cells.

Background

Pancreas is a vital organ playing crucial function in the metabolism of all vertebrates. Acinar cells, the most abundant cell type of the pancreas, produce the digestive enzymes which are conveyed to the gut by the pancreatic ducts. The pancreatic endocrine cells are grouped in the Langerhans islets and secrete diverse hormones controlling metabolism and glucose homeostasis. Five endocrine cell subtypes (alpha, beta, delta, PP and epsilon cells) have been described in pancreatic islets, each characterized by the expression of a particular hormone (glucagon, insulin, somatostatin, pancreatic polypeptide and ghrelin, respectively). Many transcriptomic studies have been focused on pancreatic endocrine cells, and notably on beta-cells, due to their implication in the development of diabetes. Microarrays and RNA sequencing (RNA-seq) were conducted on endocrine pancreatic cells isolated from human (healthy or diabetic persons) or rodents at adult and embryonic stages [1–9]. More recently, RNA-seq performed on endocrine cell types isolated by FACS or at single-cell level allowed to define the genes enriched in each endocrine cell type in human and mice [7, 10–21]. However, no comprehensive interspecies comparison has been performed so far to define the conserved signatures for each pancreatic cell type, except for two studies comparing human and murine beta-cells and reporting some notable differences between these two mammalian species [12, 21]. Comparison of transcriptomes between species is a straightforward approach to identify tissue-specific or cell type-specific genes playing crucial functions in the physiology of the studied tissue/cell. Indeed, if a gene is essential in a differentiated cell, strong constraints will maintain its expression throughout evolution and its tissue-specific expression will be detected in most species. In accordance to this view, several studies have reported conservation of organ-specific expression for a large set of genes among species [22–24]. Nevertheless, there are also striking divergences in gene expression pattern even between close vertebrate species such as human and mice, which may contribute to physiological adaptations [25, 26]. Comparative studies between evolutionary distant species are useful to identify the set of genes displaying highly conserved cell-type specific expression and likely playing a fundamental function in the studied cells. Such analyses have not been performed yet on pancreatic cells due to the lack of transcriptomic data from pancreatic cells isolated from lower vertebrates, such as zebrafish. To

tackle this lack of knowledge, we first determined the transcriptomic landscape of the three major endocrine cell types (alpha-, beta- and delta-cells) as well as of the acinar and ductal cells from zebrafish. Analysis of these zebrafish datasets allowed us to define the signature of each cell type. Then, comparison with published human and murine pancreas data led to define the conserved signatures. Furthermore, by determining the transcriptome of endocrine cells from early stage embryos, we identified genes expressed during endocrine cell differentiation and putatively involved in this process; among them, *myt1b* and *cdx4* are shown to be essential for endocrine cell differentiation in zebrafish. Thus, our list of pancreatic conserved genes represents a useful resource for studies related to pancreatic development and disease such as diabetes and pancreatic cancer.

Results

Transcriptomic profiles of the different pancreatic cell types isolated from adult zebrafish.

We purified the different pancreatic cell types from adult zebrafish using a series of transgenic reporter lines allowing the selection of these distinct cells by fluorescence-activated cell sorting (FACS). Acinar cells were obtained from the BAC transgenic lines *Tg(ptf1a:GFP)*[27]. The endocrine beta- and delta-cells were isolated respectively from the transgenic lines Tg*(ins:GFP)*^{ulg021Tg} (see MM)and *Tg(sst2:GFP)* [28]*;* the alpha-cells were obtained from Tg(*gcga:GFP*)/ Tg(*ins:NTR-mCherry*) line through selection of GFP+/mCherry- cells (as many beta cells were found to express Tg(gcga:GFP) transgene at a lower level, S1 Fig). RNA-seq was performed on 3 independent preparations for each cell type, except for acinar cells for which 4 replicates were prepared. About 60 million of paired-end reads were obtained from each Illumina library, 80% of which mapped to the zebrafish genome. We previously reported the transcriptome of pancreatic ductal cells by using the same procedure on the *Tg(nkx6.1:GFP)* ^{ulg004Tg}transgenic line [29], and these data were compared in the present study with endocrine and acinar cell transcriptomes.

Principal component (PC) analysis of all these pancreatic RNA-seq data sets showed a tight clustering of all replicates for each pancreatic cell type (Fig 1A), underscoring the high reproducibility of the data. As expected, the PC analysis also revealed a closer clustering of the 3 endocrine cell subtypes

compared to the ductal and acinar cell types; though, when the PC analysis is performed only with the endocrine datasets, clear distinct transcriptome profiles are observed for the alpha-, beta- and delta-cell subtypes (Fig 1B). Comparison of the expression levels of various known markers of each pancreatic cell type confirmed the high purity of each cell preparation. Indeed, *glucagon a* (*gcga*), *insulin* (*ins*) and *somatostatin 2* (*sst2*) were selectively detected at very high levels in alpha-, beta- and delta-cell libraries respectively, representing in average 24%, 10% and 32% of the total reads number in the corresponding libraries, while being detected at much lower levels in the other libraries (Table 1). The *trypsin* (*try*) and *chymotrypsin-like elastase member1*(*cela1*) genes were the strongest expressed genes in acinar cells representing each about 10% of all total reads of the acinar cell libraries. These two acinar markers were not detected at significant levels in endocrine data sets consistent with an accurate cell sorting. Among the genes expressed at highest levels in ducts, we find the *cldnb, sdc4* and the *epcam* genes coding for cell adhesion molecules, each representing less than 1% of total reads of ductal data sets. All these results indicate an accurate and reproducible sorting of the different pancreatic cells allowing the identification of genes selectively expressed in each pancreatic cell type. Expression values for all genes in all samples are shown in S1 and S2 Tables.

Identification of genes enriched in endocrine, acinar and ductal pancreatic cells.

The clear distinct transcriptomic profiles observed for endocrine, acinar and ductal cells prompt us to identify, in a first step, all genes presenting a differential expression in these three pancreatic tissues (with at least a 4-fold enrichment and adjusted P-value <0.05). By using these cut-off values, 1853, 1430 and 492 genes were found to be enriched in endocrine, ductal and acinar, respectively (Fig 2A and Fig 2B, gene lists are given in S3 Table). As expected, the zebrafish endocrine-specific genes include several orthologs of mammalian genes known as endocrine markers such as those involved in hormone regulation and secretion, like the proprotein convertases (*pcsk1* and *pcsk2*), carboxypeptidase E (*cpe*), secretogranins (*scg2a*, *scg3,* and *scg5*), ATP-dependent potassium channels (the *kcnj11* and *Sur1*/*abcc8* subunits), the voltage-dependent type calcium channels (*cacna1c* and *cacna1da*) among others. Moreover, most of the transcription factors previously shown to be crucial for the differentiation

of endocrine cells (*neurod*, *isl1*, *pax6b*, *insm1a, etc …)*[30] are detected in the endocrine signature validating our RNA-seq data. Interestingly, many other transcription factors, whose function in endocrine cells is still not known, are present in this list such as *egr4*, *creb3l1, lmo1, cdx1b or cdx4* (Fig 2B). Gene ontology (GO) enrichment analysis using DAVID revealed known biological pathways in endocrine cells such as "Potassium ion transport", "regulation of secretion" and "Regulation of exocytosis" as well as "Response to glucose" and "G-protein coupled receptor signaling" (Fig 2C). Indeed, many G-couple protein receptors (GPCRs) and several regulators of G-protein signaling, like *gpr12*, *gpr22, gpr27,* or *gpr63* as well as *rgs4, rgs5a, rgs8* or *rgs17* among others were found enriched in the endocrine cells. While some of these regulatory proteins have been previously reported to control pancreatic islet activity [31, 32]*,* others were not yet known to have a selective expression in pancreatic endocrine cells nor to play a role in islet physiology.

Similar observations were done for the acinar and ductal cell transcriptomic signature. Indeed, as expected, many genes coding for digestive enzymes such as *trypsin* (*try*), *elastase* (i.e. *ela* and *cela*), *trypsin-like* (*tryl*/*zgc:66382*), *amylase* (*amy2a*) as well as the known acinar transcription factors *ptf1a* and *rbpjl* were found enriched in acinar cells. The acinar gene list also includes novel markers such as genes encoding for the transcription factors *esr2a*, *klf15* or *nr0b2a* (Fig2B and S3 Table). As expected, GO enrichment analysis revealed, among the biological processes active in acinar cells, "Gland development" and "Digestive system development", "Cellular amino acid metabolic process", "Proteolysis" as well as "ATP synthesis coupled electron transport" (Figure 2D). As for the duct transcriptome, the analysis reveals novel markers such as *id2a* and *frzb* in addition to known markers like *sox9b*, *nkx6.1*, *onecut1* and *ctgfa,* as we recently described [29]. All together, these analyses confirm that many known markers and biological pathways display the same pancreatic enrichment in zebrafish as in mammals, and also highlight many novel cell-type specific genes not previously reported to display such selective expression in mammals. This led us to compare comprehensively the endocrine- and exocrine-enriched genes across zebrafish and mammalian species, thereby defining the conserved specific signatures among vertebrates.

Comparison of the pancreatic endocrine and exocrine transcriptomic signatures across zebrafish, mouse and human.

The evolutionary conserved expression of a gene in a specific cell type is a strong argument for its crucial function in that cell. This concept is widely supported by numerous studies showing conserved expression of many transcription factors in development and cell differentiation [23, 33], including pancreatic cells [34, 35]. To perform interspecies transcriptome comparison, we retrieved from public databases several RNA-seq data sets obtained on murine and human pancreatic islets, whole pancreas [11, 36–38] as well as a human pancreatic sample enriched in acinar tissue [9]. While mouse RNA-seq data are not presently available for acinar enriched preparations, the genes with either endocrine-enriched or exocrine-enriched expression (named hereafter as “endocrine or exocrine genes”) can nevertheless be identified by comparing RNA-seq from purified pancreatic islets versus RNA-seq from whole pancreas (composed of more than 90% of exocrine cells). When a global comparison of the zebrafish endocrine and acinar data sets with the murine and human pancreatic data is performed using a PC plot analysis, we can observe that the data cluster according to the tissue type (endocrine and acinar) in the first component (PC1 : representing 55% of variance), while the data tend to cluster according to the species in the second component (PC2 : representing 22% of variance)(Fig 3A). Thus, this global analysis suggests the existence of sets of genes displaying an evolutionary conserved endocrine and exocrine expression. Endocrine and exocrine genes were identified from human and murine data sets by selecting all genes presenting at least 4-fold higher expression in each pancreatic tissue (with adjusted P-value<0.05). Next, we compared the endocrine and exocrine gene lists between the three vertebrate species. The Venn diagram on Fig3B indicates that, while a large fraction of endocrine-enriched genes are species-specific, some sets of genes display a conserved endocrine-enriched expression in two species or in the three species. Indeed, among the 1853 zebrafish endocrine genes, 251 are also classified as “endocrine” in human and mice. This endocrine signature, conserved between zebrafish, mice and human (named “ZMH”), includes many factors known to be involved in hormone maturation, secretion and regulation. Many transcription factors crucial for endocrine cell differentiation [30, 39] are found in this ZMH conserved endocrine signature (see Table 2), thereby

validating this method to identify genes with important physiological function. The ZMH endocrine genes (given in S4 Table) also comprise components of signaling pathways reported to control pancreatic endocrine cells in mammals such as the glutamate receptor *gria2a* and *gria2b* [40], regulators of G-protein signaling (*rgs4* and *rgs16)* [32, 41], *urocortin3* [5, 42], ion channels (*abcc8, cacna1* and *scn1lab*) and calcium dependent proteins (*c2cd4* and *scgn*). This indicates that many regulatory processes controlling endocrine cell development and physiology have been maintained from fish to human. Interestingly, the ZMH endocrine genes include several regulatory genes whose function in pancreatic cells is still unknown or not well defined such as the kinases *map3k15*, and *mast1*, the signaling factor *gpr158* as well as the transcription factors *npas4a, lmo1, fev* and *etv1*. Similarly, the comparison of the exocrine-enriched genes across zebrafish, mice and human shows that a fraction of genes displays conserved exocrine enrichment (S2 Fig): among the 2361 zebrafish exocrine genes (combined acinar and ductal genes from Fig2A), 127 show exocrine enriched expression also in human and mice. This conserved exocrine signature (S5 Table) includes as expected many digestive enzymes as well as the known exocrine transcription factors as *ptf1a*, *rbpjl* and *nr5a2* [36, 43] but also regulatory genes with unknown pancreatic function, like *klf15*.

The interspecies comparison of both exocrine and endocrine signature also indicates that many genes display the same enrichment in only two species (e.g., “HM”: between human and mice; “ZM”: zebrafish and mouse). As expected, the number of conserved HM endocrine and exocrine genes is higher than the number of conserved ZH and ZM genes. Overall, this global analysis indicates that, while significant divergence is observed between species at the level of endocrine and exocrine signatures, hundreds of genes have nevertheless maintained a common expression pattern, and among them are found most of the known pancreatic regulatory genes.

Identification of genes expressed in embryonic endocrine cells.

While many transcription factors controlling pancreatic endocrine cell differentiation in embryos remain expressed and functional in mature cells from adults, there are nevertheless striking differences between the transcriptomes of fetal and adult cells [5]. In order to determine the fraction of the endocrine-enriched

genes in adult zebrafish that are already expressed in the first embryonic endocrine cells, we determined the transcriptomic profile of endocrine cells isolated from *Tg(pax6b:GFP)* embryos at 27 hpf. We detected 9919 genes expressed in embryonic pancreatic cells above the threshold level of 100 Normalized count. By comparing these data with adult endocrine data, we found that, among the 1853 genes enriched in adult endocrine cells, 911 genes (49 %) were already detected in embryonic cells (full gene list shown in S6 table). The expression of some of these endocrine-enriched genes was further characterized by *in situ* hybridization (ISH) on whole zebrafish embryos (Fig 4). We analyzed genes identified as endocrine-enriched in the three species (the "ZMH" genes : *pcsk1*, *pcsk2*, *fev*, *cpe*, *etv1*, *map3k15*, *and lmo1*), in two species (the "ZM" gene *cdx4* and the "ZH" gene *tbx2b*), or only in zebrafish (the "Z" genes *gpr22*, *dkk3b*, *pnoca*, *ppdpfb, scinlb* and *spon1b*). All these genes were detected in the embryonic endocrine pancreatic cells by ISH, validating the RNA-seq data. Thus, these data indicate that the early embryonic endocrine cells from the dorsal pancreatic bud express a large fraction of the genes constituting the adult endocrine signature.

Characterization of the transcriptomic signatures for the endocrine cell subtypes.

In order to define the molecular signatures of alpha-, beta- and delta-cell subtypes, the 1854 endocrine genes identified in adult zebrafish (from Fig 2A) were classified according to their differential expression in the distinct endocrine cell subtype (above the threshold of 4-fold enrichment, adjusted-P value <0.05). 73, 70 and 192 endocrine genes were found to be enriched in alpha-, beta- and delta-cells, respectively (Fig 5A). Heatmap plot in Figure 5B presents an overview of the expression pattern for all differentially expressed genes (gene list available in S7 Table). Principal component analysis using all 1854 endocrine-enriched genes displays a clear discrimination of the alpha, beta and delta cell transcriptomes (Fig 5C). GO enrichment analysis performed on beta-cell specific genes identifies "response to organic substrate", "ion transport" or " response to nutrient levels" as enriched pathways among others (see S3A Fig). The gene expressed at the highest level in the beta-cells after *insulin* is *ppdpfb*. This gene is the paralog gene of the *ppdpfa* gene (*pancreatic progenitor cell differentiation and proliferation factor a)* which was reported as specifically expressed in acinar cells and control their differentiation [44].

Interestingly, the paralog *ppdpfb* is specifically expressed in the endocrine pancreas and mainly in beta-cells. We verified the expression of *ppdpfb* in 24 hpf zebrafish embryos by fluorescent ISH and, according to the RNA-seq data, we confirmed its expression in many beta-cells and in a few alpha-cells but not in delta-cells (Fig 6A-C). The RNA-seq data also confirmed the selective expression of the zebrafish *nkx6.2* gene in beta-cells, as previously reported [45]. Expression of the zebrafish *pcsk2* gene is more enriched in beta-cells compared to the *pcsk1* gene (beta/alpha enrichment of 5.5- and 2.5-fold, respectively), while in mouse and human only PCSK1 was reported to be more enriched in beta- compared to alpha-cells [4, 12, 46]. The enrichment of *pcsk2* in beta-cells was confirmed by FISH on zebrafish embryos (Fig 6D-F). The zebrafish estrogen receptor 1, *esr1,* is also strongly enriched in beta-cells (10575 normalized counts in beta-cells versus 671 in alpha- and 71 in delta-cells). Finally, the beta-cells selective expression of *spondin 1b* (*spon1b*), an activator of the Wnt pathway, was also validated by FISH on zebrafish embryos (S4 Fig).

As expected, the genes specifically expressed in alpha-cells include those coding for the two zebrafish glucagon hormones (*gcga* and *gcgb*) as well as for the Arx transcription factor (*arxa*) but also novel markers including peptide hormones such as *prepronociceptin* (*pnoca*) and *neuropeptide B* (*npb*) or the calcium regulated factor scinderin like b (*scinlb*). The alpha-enriched genes also comprise the transcription factors *etv1* and *si:ch211-145o7.3* (a forkhead domain factor). The alpha-selective expression of *pnoca* and *scinlb* was confirmed by FISH, validating the RNA-seq data (S5 Fig).

As for the delta-cells, besides the somatostatin genes [47], the RNA-seq data revealed many novel markers, which include genes coding for transcription factors such as *dlx3b*, *cdx1*, *cdx4* and *tbx2b,* as well as genes coding for signaling factors like *bmp7b,* or the kinase Map3k15. Interestingly, as demonstrated by the GO enrichment analysis (S3 Fig), many G-protein coupled receptor (GPCR) are specifically enriched in delta cells such as npy8br (neuropeptide Y receptor), glra4a (glycine receptor), uts2r (urotensin receptor), ptger1b & 2a (prostaglandin receptor), adra1d (adrenoreceptor), grm3 (glutamate receptor), oprd1b (opioid receptor) and gpr123, pinpointing these endocrine cells as targets of diverse external metabolic signals. We confirmed the selective expression for several genes, including the *cdx4, cdx1, lamc2* and *map3k15* genes, detected already at 24 hpf in delta cells (S6 Fig).

All these results validate the RNA-seq data and show that many endocrine cell subtype markers acquire their selective expression in the first embryonic endocrine cells.

Comparison of Transcriptomic signatures of alpha- and beta-cells across species.

The zebrafish RNA-seq data led to the identification of many novel markers for each endocrine cell subtypes raising the question whether such markers display the same cell subtype-specific expression in other species. As the transcriptome of human and murine alpha- and beta-cells have been determined by a similar approach using highly purified FACS cell preparations [12, 48], this allowed us to perform an interspecies comparison. In order to identify all alpha- and beta-cell differentially expressed genes (with no restriction on endocrine enriched genes), we performed exactly the same procedure on the zebrafish, mouse and human RNA-seq data sets : a comparison of the entire alpha-cell and beta-cell transcriptomes within each species using the same software and selection criteria (ratio of alpha- versus beta-cells above 4-fold enrichment, with adjusted P-value <0.05). This led to the identification of 747, 1330 and 1102 alpha-enriched genes as well as 544, 381 and 465 beta-enriched genes in human, mouse and zebrafish, respectively. Surprisingly, the interspecies comparison of these sets of genes revealed that the large majority of alpha- and beta- enriched genes are species specific (Fig 7). Indeed, among the 465 zebrafish beta-enriched genes, only 3 have maintained beta-cell preferential expression in human and mice : *insulin*, the transcription factor *pdx1* and the glucagon receptor *gcgra* (Table 3 and S8 Table for all ZH, ZM and HM beta-enriched genes). Similarly, among the 1102 zebrafish alpha-cell enriched genes, only 20 have human and murine orthologs with alpha-cell enriched expression (Fig7, Table 3 and S9 Table). As expected, *glucagon* and *arx* are part of the conserved alpha-cell genes. The *adcy2* and *adcy7* genes coding for adenylate cyclase are also alpha-enriched in the three vertebrate species supporting the important role of cAMP in alpha-cells [49]. The *gc* gene (coding for vitamin-D binding protein)[14] and *fev* (coding for a Ets transcription factor) are among the conserved alpha-cell signature suggesting a specific function in this endocrine cell subtype. Taken together, all these analyses indicated that, while hundreds of genes display conserved expression in endocrine and exocrine cells, there is much less conservation for the genes differentially

expressed between the endocrine cell subtypes alpha and beta. These data confirm the striking differences between species as previously reported for the human and murine alpha- and beta-enriched genes [12, 21]. Importantly, many of the conserved alpha- and beta-specific genes identified in the present study were also recently highlighted as enriched in these endocrine cell types in several single cell RNA-seq studies [19–21](see S8 (beta-cells) and S9 (alpha-cells) Tables for comparison).

Role of the zebrafish *myt1b* and *cdx4* genes in endocrine cell differentiation.

To validate our cross-species approach to identify important pancreatic regulatory factors, we selected two transcriptional regulators which were expressed at high levels in endocrine cells at early developmental stage (27 hpf) and performed loss of function studies. Mutations of *myt1b* gene ("ZMH" conserved endocrine genes) and of the *cdx4* gene ("ZM" conserved endocrine gene) were found to affect endocrine cell differentiation. As *cdx4* is selectively expressed in delta-cells, we analyzed the expression of several novel delta-cell markers in the *cdx4*^{tv205} null zebrafish mutant [50]. A previous study has shown that the zebrafish *cdx4* mutant displays defects in the anterio-posterior patterning of the endoderm, with notably a posteriorly-shifted pancreas and an increase of pancreatic beta-cell number [51], however, the selective expression of *cdx4* in delta-cells was unknown as well as its function in this endocrine cell type. Analysis of the null *cdx4*^{tv205} mutant embryos revealed a complete loss of *somatostatin2* gene expression (Fig 8A). Similarly, expression of the new delta-cell markers *lamc2* and *slc7a7* was almost undetectable in the mutant embryos, in contrast to the increase of insulin expression (Fig8 B-C). Expression of *map3k15* gene in delta pancreatic cells was also specifically abrogated by *cdx4* mutation while expression of this gene was not affected anteriorly at the level of the pronephric glomeruli (white arrows Fig 8D). All these data demonstrate the important role of *cdx4* for delta-cell differentiation.

Our transcriptomic analyses show that embryonic pancreatic cells express both *myt1a* and *myt1b* paralogs, with *myt1b* being expressed at much higher levels. This was confirmed by ISH revealing expression of *myt1b* in the dorsal pancreatic bud of zebrafish embryos (S7 Fig) while the paralog *myt1a* was barely detectable (data not shown). The adult RNA-seq data indicate that the

myt1a/b genes are expressed in alpha, beta and delta cells. Simultaneous inactivation of the two zebrafish *myt1* genes was performed by multiplex CRISPR/Cas9 mutagenesis through injection of four guide RNA (two different guides targeting each gene, see M&M). Injected embryos revealed a significant decrease of *glucagon* expression and no significative effect on *insulin* expression (S7 Fig). To confirm these data obtained in F0 embryos, a line harboring a null mutation in *myt1b* (*myt1b*^{ulg029}) was raised. Alpha-cell mass was decreased in the *myt1b*^{ulg029} -/- embryos compared to the wild type and heterozygous siblings (Fig9).

Discussion

In this study, we have explored the transcriptomic landscape of the major pancreatic cell types in zebrafish, thereby defining the transcriptomic signature of acinar and endocrine cells, as well as of the three major endocrine cells subtypes alpha, beta and delta. By this analysis, we identified many novel cell-type specific markers with still unknown pancreatic function. By comparing the endocrine and exocrine transcriptomic signatures from zebrafish, human and mouse, we could define an evolutionary conserved signature for these two pancreatic tissues, pinpointing genes and pathways that likely represent key players in the pancreas of vertebrates. Consistent with this notion, more than half of all transcription factors in these endocrine or exocrine conserved signatures are known regulators of pancreatic cell differentiation, such as the transcription factors *neuroD, isl1, pax6, insm1, ptf1a, rbpjl* among others. *Myt1* is part of the endocrine conserved signature and we show here that inactivation of the *myt1b* gene in zebrafish leads to a decrease of alpha cell mass in 2dpf embryos. Also, *cdx4*, which is selectively expressed in delta cells, is also necessary for their differentiation.

The endocrine signature conserved among the three vertebrate species reveals that several signaling pathways regulating hormone secretion and cellular homeostasis are commonly used by pancreatic cells from fish to mammals. For example, the evolutionary conserved signature comprises the RNA binding protein *Elavl4/Hud* and the ionotropic glutamate receptor *Gria2*, shown to regulate hormone synthesis and secretion in rodents [40, 52, 53]. Similarly, the glucose response system regulating insulin and glucagon secretion in mammals seems to be also used in zebrafish as various components of this pathway are found in the conserved endocrine genes, such as the ATP sensitive K+ channels *abcc8* and *kcnj11* and the voltage-dependent Ca++ channel *cacna2d2*. The use of this pathway in zebrafish is further strengthened by the very high expression of the glucose transporter glut2 (*slc2a2*) and the glucokinase (*gck*) in zebrafish beta-endocrine cells. Also, some GPCR receptors which have been reported to control the activity of pancreatic islet cells in mice, such as the Sstr3 (somatostatin receptor) [54], CasR [55] or Celsr3 [56] are included in the conserved endocrine signature. Thus, this list of conserved endocrine-enriched genes strongly suggests that the function of several signaling

pathways controlling the formation and activity of pancreatic cells has been maintained from fish to humans. This is consistent with many studies using zebrafish as model for pancreas development and diabetes that revealed conserved physiological regulations (reviewed in [34, 35]). The conserved endocrine and exocrine pancreatic signatures also pinpoint many novel candidate regulatory genes that deserve special attention for future studies. For instance, the transcription factors *lmo1* and *npas4a*, or the signaling factors *gpr158*, *rgs7* and *rgs17,* all selectively expressed in pancreatic endocrine cells in zebrafish, mice and human, more than likely play an important role in pancreatic cells.

In the present study, we identified the human and murine endocrine- and exocrine-enriched genes by comparing transcriptomic data from purified islets with data obtained from whole pancreas as well as from one human pancreatic sample enriched in acinar tissue. While this approach identified almost all expected endocrine and exocrine markers, some markers may have been missed due to low purity of tissue preparations. For example, the human *MNX1* and murine *Esr1* (*Estrogen receptor alpha*) genes were not enriched in the human or murine endocrine samples, respectively, while these two genes were reported to be important for endocrine beta-cells in both species [57–60]. These misclassifications may be due to the use of RNA-seq data obtained from whole pancreas instead of highly purified exocrine cells; such potential errors will be corrected when RNAseq data from purified ductal and acinar cells will be available. The recent single cell transcriptomic data reported for human and murine pancreas bypass the need for purified cells and will probably constitute an useful resource for doing such interspecies comparison. However, murine acinar cells have not been captured in these single-cell studies preventing the identification of acinar-enriched and endocrine-enriched genes. When such data will be available, it will interesting to perform a global analysis of all pancreatic single cell transcriptomic studies and compared with the zebrafish pancreatic data. In the meantime, genes classified in our list of conserved “ZM” or “ZH” endocrine genes should be also considered as they indeed include *mnx1* and *esr1*, indicating that important genes fall in these two categories. For proof, c*dx4,* classified as endocrine-enriched in zebrafish and mice (i.e. “ZM”), was found here to be crucial for endocrine delta-cell differentiation as many novel delta-cell makers were drastically reduced in the

cdx4 mutant zebrafish embryos. Our findings warrant future analyses on the murine (and human) *Cdx4* gene to decipher its expression and function during pancreas development in mammals. Interestingly, the loss of delta-cells with the concomitant increase of beta-cells observed in the zebrafish *cdx4* mutant suggests that Cdx4 could act on the balance of delta- versus beta-cells by determining the fate of endocrine precursors.

The present study also provides a comprehensive list of new markers of the zebrafish endocrine alpha-, beta- and delta-cell subtypes. An overall view indicates that the alpha- and beta-cells are slightly more similar to each other compared to delta-cells, as shown by the PC analysis (Fig 1B and Fig 5C). This is also supported by the Venn diagram (Fig 5A) showing that i) the number of genes with enriched expression in both alpha- and beta-cells and not in delta-cells ("apha-beta" enriched genes) is higher than those of the "alpha-delta" enriched and "beta-delta" enriched genes (123 genes versus 28 and 68 genes, respectively) and, ii) the number of delta-cell markers is higher than the alpha- and beta-cell marker (192 versus 70 and 73 genes, respectively). These observations are consistent with a recent study in zebrafish larva showing the higher capacity of alpha-cells to transdifferentiate to beta-cells compared to delta-cells [61]. If this is also true in adult zebrafish must be verified as it has been demonstrated in mice that the transdifferentiation competence of alpha-cells or delta-cells toward β-cells is age-dependent [62].

An unexpected observation of our study is the low conservation of the alpha- and beta-cell transcriptomic signatures between zebrafish, mice and human. Indeed, only 20 and 3 genes define the conserved alpha-cell and beta-cell signatures, respectively. Such low conservation cannot be attributed to a low quality of cell samples or RNA-seq data because, for the 3 species data sets, there is a very high enrichment of alpha- and beta-cell markers in the corresponding libraries confirming the good purity of alpha- and beta-cell preparations. Also, it is noteworthy that this low conservation is not only observed between zebrafish and the two mammalian species but also between human and mouse (see Venn diagram in Fig 8). Differences between human and mouse alpha- and beta-enriched genes have also been previously noticed [12, 21]. Several non-exclusive explanations can be proposed for this apparent

low conservation of the alpha- and beta-signatures in both distant and closer species. First, as the alpha- and beta-cells have relatively similar transcriptomes as shown on Fig.1A, the identity of these two cell subtypes could rely on the action of only few regulatory genes. This hypothesis is supported by the presence of *Arx* among the few conserved alpha-cell genes, which is the key determinant of alpha-cell identity and its inactivation in mice is sufficient to transdifferentiate alpha-cells to beta-cells [63]. Similarly, Pdx1, one of the three conserved beta-cell genes, is required to maintain beta-cell identify and repress the alpha-cell program [64]. This first hypothesis is also supported by the ability of alpha- and beta-cells to transdifferentiate into each other [65, 66]. A second explanation can be the relatively stringent threshold that we used to select differentially expressed genes (4-fold enrichment, adjusted-P<0.05). For example, the *mnx1* gene, known to be involved in beta-cell differentiation [67, 68], is enriched in beta-cells in the three vertebrate species but was not classified as beta-enriched in human and zebrafish due to a rather low beta- versus alpha-cell enrichment (3-fold in human, 1.9-fold in zebrafish versus 70-fold in mice). These differences in enrichment between human and mouse *Mnx1* is confirmed in the recent single cell RNA-seq data [18–21]. Similarly, the *prohormone convertase 1* gene *pcsk1* required for insulin hormone maturation was only 2.7-fold enriched in zebrafish beta-cells versus alpha-cells. It is possible that small differences in expression levels may be sufficient for some genes to give a different physiological response. A third explanation of the low cell subtype conservation could stem from functional switches occurring between homologous genes able to perform the same function. For example, *Nkx6.1* is expressed in beta-cells in mice and human but not in zebrafish where its paralog *nkx6.2* instead of *nkx6.1* is selectively expressed in mature beta-cells [45]. *nkx6.2* may fulfill the function of *Nkx6.1* as Nkx6.1 and Nkx6.2 have equivalent biological activities [69]. Another example is the Zn++ transporter *scl30a8*, which is strongly expressed in beta-cells in mammals but not in zebrafish, which instead strongly expresses the homologous Zn++ transporter *scl30a2*. Such functional switches may be difficult to identify as they can occur between genes of the same superfamily but belonging to different subclasses. For example, we have shown that, in zebrafish, the role of the bHLH Neurog3 as pancreatic endocrine cell fate determinant is fulfilled by two other bHLH factors Ascl1b and Neurod1 [70]. A functional switch is also observed for the specification of the secretory cell of the intestine, played by

Atoh1 in mice and by Ascl1a in zebrafish [71]. The capacity of homologous factors to fulfill the same function is also supported by the observations that many null mutations can be compensated by a homologous gene [72–74]. Such compensations seem to occur for the murine *Myt1* gene as the *Myt1* KO mice only display very mild pancreatic defects with no decrease of endocrine cells [75] while more drastic defects are observed by expression of a dominant negative Myt1-Eng protein [1, 76]. Furthermore, adult murine endocrine pancreatic cells express the homologous *Myt1l* and *Myt3/St18* genes [77] and their expression increases in the *Myt1* KO mice [75]. In zebrafish, *myt1a*, *myt1l* and *myt3/st18* are expressed at much lower level compared to *myt1b*; this probably explains the decrease in alpha-cell mass observed in the single *myt1b* mutant. Further experiments will be required to determine whether some compensation occurs through *myt1a* and if the defects are more drastic in the double myt1a/b mutant.

The stronger expression of glucagon receptor in the beta-cells of the three vertebrate species highlights the importance of paracrine regulations between alpha- and beta-cells and indicates a role of glucagon on beta-cell physiology as previously reported in rodents [78] and in zebrafish for beta-cell regeneration [61]. Interestingly, our transcriptomic analyses highlight *fev* as an endocrine conserved gene as well as an alpha-cell conserved gene, suggesting a role of this transcription factor in alpha-cells. Glucose tolerance test in *Fev* knockout mice previously revealed a slower response apparently due to a reduction in insulin production while the level of most crucial transcription factors of beta-cells were unchanged [79]. As our transcriptomic analysis indicates that *fev* is expressed more than 20-fold higher in alpha-cells versus beta-cells in zebrafish, mice and human, this argues for an important function in alpha-cells and warrants further phenotypic analyses.

Conclusions

The present transcriptomic analysis of the distinct zebrafish pancreatic cell types identifies novel pancreatic regulatory genes and thereby constitutes a valuable resource for future studies of the pancreas not only in zebrafish but also in mammals giving interesting clues on genes and signaling pathways active in these cells. The comparison of the present zebrafish pancreatic RNA-seq data with

the recent single cell pancreatic transcriptomic data or with future data of pure pancreatic cells obtained not only from human and mouse but also from other species will help to better define gene regulatory networks controlling pancreas ontogeny and physiology in vertebrates. This will be useful for studies aimed at understanding dysfunction of pancreatic endocrine cells in human diseases like diabetes and to design novel drugs or therapies.

Materials and Methods

Generation of the tg*(Insulin:GFP)*$^{ulg021\ Tg}$ *line*

The *(Insulin:GFP)* transgene was generated by first cloning a 897 pb PCR fragment, amplified with O97 and O98 primers (S10 Table), that includes 745 bp of the insulin promoter, the exon 1 and intron 1 and 7 bp of exon 2 just upstream of the ATG of the insulin ORF, into the gateway vector pCR8/GW/TOPO to produce pMV90-G2a plasmid. Second, a triple LR recombination using p5E-MCS, pMV90-G2a and p3E-EGFP-PA inserted into pDestTol2pA2, provided by the tol2kit [80] generated the transgene *Tg(insulin:GFP)* that has been introduced into AB embryos by coinjection with the Tol2 transposase to generate the *Tg(insulin:GFP)*$^{ulg021\ Tg}$ line.

Preparation of zebrafish pancreatic cells by FACS:

The endocrine cells were prepared using the zebrafish transgenic lines *Tg(Insulin:GFP)*^{ulg021Tg}, *Tg(gcga:GFP/ins: mCherry)ia1*, and *Tg(st22:GFP)* to isolate beta, alpha and delta cells respectively [28, 81]. Acinar and ductal cells were isolated by using *Tg(ptf1a:GFP)* [27] and *Tg(nkx6.1:GFP)*^{ulg004Tg} [29], respectively. The endocrine tissue was dissected under an epifluorescence stereomicroscope before dissociation by enzymatic treatment 30min at 28°C with 1X Tryple Select (Life Techologies), 40ug/mL Proteinase K (Roche) and 10ug/mL Collagenase IV (Life Technologies), combined with mechanical

disruption by pipetting every 5 minutes. For acinar cell preparations, pancreata were digested with a mix of collagenase IV (500 µg/ml), collagenase P (350 µg/ml) and dispaseII (1 mg/ml) for 15 min at 28° with mechanical disruption every 5 minutes. After dissociation, cells were washed twice with 1x PBS containing 1% BSA and pelleted at 4°C for 5 min at 300g and immediately sorted by FACS. Cells were selected based on GFP expression using FACS Aria II by two consecutive sorting steps: the first sorting was done in the “yield” mode and the second in the “purity” mode. Purity was estimated by FACS Aria II after cell sorting (more than 99% purity of GPF+ cells) and by fluorescence microscopy (from 95-99% purity). Each replicate sample was prepared from 4 adult zebrafish. About 10000-20000 endocrine cells, 20000-40000 acinar cells and 2000-10000 ductal cells were obtained after FACS and used for library preparations.

RNA extraction, cDNA amplification, library preparation and Sequencing:

Total RNA was extracted from FACS sorted cells using the RNeasy plus micro kit (Qiagene). RNA from endocrine and acinar cells was eluted in 10uL with a concentration of 100-400 pg/uL. RNA integrity was assessed by a capillar electrophoresis using Agilent RNA 6000 pico chip (Agilent technologies), the RIN value for each sample was from 8 to 10. The Smarter Ultra low RNA input kit (clontech)[82] was used to for the synthesis and amplification of cDNA synthesis using up to 10ng of total RNA following the manufacturer's instructions and performing no more than 12 cycles of PCR in order to minimize amplification biases. The quality of cDNA was verified by 2100 High Sensitivity DNA assay (Agilent technologies). Truseq DNA Illumina libraries were prepared and sequenced to obtain around 90 millions of reads (100 bp paired-end reads) per library using the Hiseq 2000 Illumina sequencer.

RNA-seq data analysis

Sequences were trimmed in order to remove adaptors and low quality bases. Trimmed reads were mapped in to the genome (Zv9, Ensembl genome version 75, ensembl.org) using Tophat v.2.0.9 [83]. Tophat's options were set according to the library features (-r 220 --mate-std-dev 82 --segment-length 18) and the option --min-intron-length was set up to 30 nucleotides according to the intron length

described by [84] . For the mouse and human data sets, raw data were downloaded from the public databases : human pancreas (4 samples from Fagerberg et *al*. [37] and 1 acinar cells enriched sample from Morán et *al*. [9] : the E-MTAB-1294 dataset), human islets (3 samples from Nica et *al*. [46] and 4 samples from Morán et *al*. [9]), murine pancreas (2 samples from Holmstrom et *al*. [36]) and murine islets (3 samples from Morán et *al*. [9]). The alpha-cells and beta-cells RNA-seq datasets were obtained from Benner et al. for mouse (2 samples of each) [12] and from Blodgett et al. for human (6 samples of each) [48]. Reads were mapped in to the mouse GRCm38 and the human GRCh37 genomes (Ensembl genome version 75, ensemble.org) using default options. Gene expression was measured from the mapped reads by using HT-seq-count [85]. Principal component analysis (PCA), using princomp function of R, was calculated for the whole data set using the variance stabilization transformation (VSD) values obtained by DESeq R package from the gene expression values [86]. For differential expression (DE) analysis, we used the R package DESeq2 [87]. DESeq2 employs shrinkage estimation for dispersions and fold of change (FC); it uses Wald test for significance with posterior adjustment of p-values using the procedure of Benjamini and Hochberg (giving adjusted P-values). Genes differentially expressed were selected with an adjusted P-value < 0.05 and a FC>4. For the comparison of the acinar, ductal and endocrine cell data, in-silico endocrine data sets were simulated by the combination of alpha, beta and delta raw RNA-seq data, the endocrine data set #1 being obtained by the combination of the alpha1, beta1 and delta1 data sets (same process for endocrine dataset#2 and #3). The RNA-seq raw data have been deposited on ENA (http://www.ebi.ac.uk/ena) under the accession number PRJEB10140.

Comparison of the human, murine and zebrafish pancreatic transcriptomes

The predicted orthologs among Zebrafish, Mouse and Human were obtained from Ensembl [88]. The informations from the 3 species were retrieved using Biomart tool (http://www.ensembl.org/biomart). Two different orthology tables were generated (tables available upon request). The first table contains all zebrafish, murine and human genes with 1-1-1 orthology relationship. The orthology table2 comprises all 1-1-1 orthologs as well as the genes presenting 1-many-many orthology relationships and thus

include all duplicated genes (paralogs) which are notably found in the zebrafish genome. The interspecies PC analysis was performed using only the genes with a 1-1-1 orthology relationship. The genes presenting an endocrine-enriched or an exocrine-enriched expression were identified in mouse and in human by using DESeq2 software selecting genes with at least 4-fold higher expression in islet dataset or in whole pancreas dataset (adjusted P-value<0.05). The interspecies comparison of endocrine- and exocrine-enriched genes was performed using the orthology table2.

GO enrichment analysis

Tissue- and cell type-enriched genes were converted and uploaded to DAVID bioinformatics resource [89]. Using the Functional annotation tool, we run a GO enrichment analysis for endocrine and acinar enriched genes as well as for the alpha-, beta- and delta-enriched genes. Enriched processes identified by GOTerm_BP_FAT with a p-value < 0.1 were selected.

In situ hybridization

Antisense RNA probe for the different genes were prepared as described by Thisse et al 2008 [90], except for *fev* [91]. Briefly, primers were designed to amplify a part of the transcript that is used as a template to synthesize the probe. The reverse primer at the 5' contains the minimal promoter sequence for T3 RNA polymerase (5'-AATTAACCCTCACTAAAGGGAG-3'), templates were amplified by RT-PCR using the set of primers shown in Supplementary table 10. Whole mount *in situ* hybridization and fluorescent *in situ* hybridization (WISH and FISH) were performed as it is described in [92] applying some modification to this protocol. Briefly, larvae of 3 dpf or older were incubated during 20 minutes in Methanol and 3% H_2O_2 at RT, prior to dehydration. Antisense probe hybridization was performed using 10-50ng of DIG- and DNP-probes in hybridization buffer containing 5% Dextran sulfate (MW: 500000) at 65C overnight. Antibodies were pre-absorbed on homogenized larvae (mix of different developmental stages) for 2h at RT and then diluted to 1/3000 DIG-AP, 1/1500 DIG-HRP, 1/800 DNP-HRP (PerkinElmer).

Inactivation of *myt1a* and *myt1b* genes by multiplex CRISPR/cas9 mutagenesis.

Mutations in the *myt1a* and *myt1b* genes were generated by multiplex CRISPR/Cas9 technology essentially as described previously [93, 94]. The nls-zCas9-nls mRNA was synthesized by transcription of the plasmid pT3TS-nCas9n (Addgene). CRISPR guide RNAs were selected using CRISPR design (http://crispr.mit.edu/) and chopchop (https://chopchop.rc.fas.harvard.edu/) softwares to target the beginning of Myt1a and Myt1b coding regions (guides 1) and the regions coding for the first zinc finger domain (guides 2). The following target sites were used: GCCAAGACGCAGATGATAAGCGG and GATGGTTTAGGCCATGTCAGTGG for *myt1a*; GTCTGAGGGAGGGCCGGCAGCGG and TGCCATTGCATCCTGGAGTGGGG for *myt1b* (PAM motifs are underlined). The DNA templates were prepared by annealing and filling 2 oligonucleotides containing the T7 promoter sequence and the target sequences as previously described [94]. After synthesis and purification of gRNA, fertilized zebrafish eggs were injected with about 1 nl of a solution containing 50 ng of the four gRNA and 300 ng of nls-zCas9-nls mRNA. The efficiency of mutagenesis was verified by genotyping using Heteroduplex Migration Assays after amplification of targeted genomic sequences. A few injected embryos were fixed in PFA at 48 hpf for phenotypic analysis and the others were raised until adulthood. Founder fish transmitting a germline mutation in myt1b were outcross with wild type fish; F1 fish harbouring a 5 bp insertion in *myt1b* (*myt1b*^{ulg039} allele) causing a frameshift in the coding sequences were incrossed to generate *myt1b-/-* embryos and +/- , +/+ siblings which were analysed by immunohistochemistry.

Immunohistochemistry

Expression of insulin and glucagon was analyzed by immunofluorescence on whole-mount zebrafish embryos. After overnight fixation in 2% PFA at 4°C, 1 hour incubation in PBS 1% Triton X-100 for and 2 hours at RT in blocking solution (4% BSA, 10% DMSO, 0.3% Triton X-100 in PBS), embryos were incubated with the primary antibodies anti-Mouse Glucagon 1/200 (Sigma, G2654) and Anti-Guinea Pig Insulin 1/300 (MP, 64714). After washing, embryos were incubated with the secondary antibodies Anti-Mouse Alexa 568 or 633 (Invitrogen, A-11004 and A-21052) and Anti-Guinea Pig Alexa 568 or 633 (Invitrogen, A-11075 and A-21105) diluted 1/300. After washing and mounting, embryos were scanned

with a Leica SP5 confocal microscope and images were analyzed using Imaris 7.2.3 software. Alpha- and beta- cell mass was measured using Imaris based on 3D reconstitution of Fluorescence signal obtained by immunofluorescence; the same confocal and Imaris parameter settings were used for wild-type and mutant embryos. To compare the number of alpha cells and beta cells in wild type and mutants, glucagon+ and insulin+ cells were counted every 6 µm optical section throughout the whole principal islet. Expression of GFP and mCherry in the pancreas of Tg(gcga:GFP);(ins:NTR-mCherry) adult transgenic fish as well as of Tg(sst2:GFP) adult fish was analysed by immunofluorescence on cryosections (S1 fig). The antibodies used are the same than described above and also include the anti-Rabbit Somatostatin 1/300 (Dako, a0564) and Anti- Rabbit Alexa 568 or 633 1/300 (Invitrogen, A-11011 and A-21070). Expression of GFP in beta-cells was also verified for the Tg(insulin:GFP) larvae (S1 Fig).

Declarations

Ethic approval:

All animal work has been conducted according to national guidelines and all animal experiments described herein were approved by the ethical committee of the University of Liège (protocol numbers 1328).

Acknowledgments :

We are very grateful to Francesco Argenton for the transgenic line Tg(*gcga:GFP*), to Zhen Li and Zhiyuan Gong for the transgenic line Tg(*st22:GFP*), to Steven Leach for the transgenic line Tg(ptf1a:GFP) and to Leonard Zon for the *kgg/cdx4* mutant line. We thank the following GIGA technical platforms: GIGA-Zebrafish (H. Pendeville), GIGA-Cell Imaging and Flow Cytometry platform (S. Ormenese and S. Raafat), GIGA-Genotranscriptomic (B. Hennuy, W. Coppieters and L. Karim) and GIGA-Immunohistochemistry (C. Humblet and E. Dortu). E. T-S. was supported by WBI, Becas Chile and Leon Fredericq fund, A.L. by FRIA, K.P. by WBI, M.L.V, I.M, and B.P, are Chercheur qualifié FNRS. This work was funded by the FNRS-FRS, the Belgian State's “Interuniversity Attraction Poles” Program (SSTC, PAI) and the “Fonds Speciaux” from ULg.

Availability of Data:

The RNA-seq raw data have been deposited on ENA (http://www.ebi.ac.uk/ena) under the accession number PRJEB10140. The gene expression values are given in Supplemental Tables 1 and 2.

Authors' contributions.

ET-S, MLV, IM and BP designed the experiments and ET-S, AL, AB, KP, DB and IM performed the experiments. ET-S, MLV, IM and BP wrote the manuscript. All authors read and approved the article.

Competing interest.

The authors declare that they have no competing interest.

Abbreviations:

hpf: hours post fertilization; dpf: days post fertilization; WISH : Whole-mount in situ hybridization; FISH: Fluorescent *in situ* hybridization; ins: insulin; gcga: glucagon a; sst2: somatostatin 2; ZMH: genes with conserved expression pattern in Zebrafish, Mice and Human; ZM: genes with conserved expression pattern in Zebrafish and Mice; ZH: genes with conserved expression pattern in Zebrafish and Human;

References:

1. Gu G, Wells JM, Dombkowski D, Preffer F, Aronow B, Melton DA: **Global expression analysis of gene regulatory pathways during endocrine pancreatic development.** *Development* 2004, **131**:165–179.

2. Gunton JE, Kulkarni RN, Yim S, Okada T, Hawthorne WJ, Tseng Y-H, Roberson RS, Ricordi C, O'Connell PJ, Gonzalez FJ, Kahn CR, O'Connell PJ, Gonzalez FJ, Kahn CR: **Loss of ARNT/HIF1beta mediates altered gene expression and pancreatic-islet dysfunction in human type 2 diabetes.** *Cell* 2005, **122**:337–349.

3. Dorrell C, Schug J, Lin CF, Canaday PS, Fox a. J, Smirnova O, Bonnah R, Streeter PR, Stoeckert

CJ, Kaestner KH, Grompe M, Stoeckert Jr. CJ, Kaestner KH, Grompe M, Stoeckert CJ, Kaestner KH, Grompe M: **Transcriptomes of the major human pancreatic cell types**. *Diabetologia* 2011, **54**:2832–2844.

4. Martens G a., Jiang L, Hellemans KH, Stangé G, Heimberg H, Nielsen FC, Sand O, van Helden J, Gorus FK, Pipeleers DG: **Clusters of conserved beta cell marker genes for assessment of beta cell phenotype**. *PLoS One* 2011, **6**.

5. Blum B, Hrvatin SSŠ, Schuetz C, Bonal C, Rezania A, Melton DA: **Functional beta-cell maturation is marked by an increased glucose threshold and by expression of urocortin 3**. *Nat Biotechnol* 2012, **30**:261–264.

6. Eizirik DL, Sammeth M, Bouckenooghe T, Bottu G, Sisino G, Igoillo-Esteve M, Ortis F, Santin I, Colli ML, Barthson J, Bouwens L, Hughes L, Gregory L, Lunter G, Marselli L, Marchetti P, McCarthy MI, Cnop M: **The human pancreatic islet transcriptome: expression of candidate genes for type 1 diabetes and the impact of pro-inflammatory cytokines.** *PLoS Genet* 2012, **8**:e1002552.

7. Ku GM, Kim H, Vaughn IW, Hangauer MJ, Myung Oh C, German MS, McManus MT: **Research resource: RNA-Seq reveals unique features of the pancreatic β-cell transcriptome.** *Mol Endocrinol* 2012, **26**:1783–92.

8. Benitez CM, Qu K, Sugiyama T, Pauerstein PT, Liu Y, Tsai J, Gu X, Ghodasara A, Arda HE, Zhang J, Dekker JD, Tucker HO, Chang HY, Kim SK: **An Integrated Cell Purification and Genomics Strategy Reveals Multiple Regulators of Pancreas Development**. *PLoS Genet* 2014, **10**.

9. Moran I, Akerman I, van de Bunt M, Xie R, Benazra M, Nammo T, Arnes L, Nakic N, Garcia-Hurtado J, Rodriguez-Segui S, Pasquali L, Sauty-Colace C, Beucher A, Scharfmann R, Van Arensbergen J, Johnson PR, Berry A, Lee C, Harkins T, Gmyr V, Pattou F, Kerr-Conte J, Piemonti L, Berney T, Hanley N, Gloyn AL, Sussel L, Langman L, Brayman KL, Sander M, et al.: **Human beta cell transcriptome analysis uncovers lncRNAs that are tissue-specific, dynamically regulated, and abnormally expressed in type 2 diabetes**. *Cell Metab* 2012, **16**:435–448.

10. Bramswig NC, Everett LJ, Schug J, Dorrell C, Liu C, Luo Y, Streeter PR, Naji A, Grompe M, Kaestner KH: **Epigenomic plasticity enables human pancreatic alpha to beta cell**

reprogramming. *J Clin Invest* 2013, **123**:1275–1284.

11. Nica AC, Ongen H, Irminger J: **Cell-type , allelic and genetic signatures in the human pancreatic beta cell transcriptome Cell-type , allelic and genetic signatures in the human pancreatic beta cell transcriptome**. 2013:1554–1562.

12. Benner C, van der Meulen T, Cacéres E, Tigyi K, Donaldson CJ, Huising MO, Caceres E, Tigyi K, Donaldson CJ, Huising MO: **The transcriptional landscape of mouse beta cells compared to human beta cells reveals notable species differences in long non-coding RNA and protein-coding gene expression.** *BMC Genomics* 2014, **15**:620.

13. Li J, Klughammer J, Farlik M, Penz T, Spittler A, Barbieux C, Berishvili E, Bock C, Kubicek S: **Single-cell transcriptomes reveal characteristic features of human pancreatic islet cell types**. *EMBO Rep* 2016, **17**:178–187.

14. Ackermann AM, Wang Z, Schug J, Naji A, Kaestner KH: **Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes**. *Mol Metab* 2016, **5**:233–244.

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

16. Wang YJ, Schug J, Won K-J, Liu C, Naji A, Avrahami D, Golson ML, Kaestner KH: **Single cell transcriptomics of the human endocrine pancreas**. *Diabetes* 2016.

17. Adriaenssens AE, Svendsen B, Lam BYH, Yeo GSH, Holst JJ, Reimann F, Gribble FM: **Transcriptomic profiling of pancreatic alpha, beta and delta cell populations identifies delta cells as a principal target for ghrelin in mouse islets.** *Diabetologia* 2016, **59**:2156–65.

18. Baron M, Veres A, Wolock SL, Faust AL, Gaujoux R, Vetere A, Ryu JH, Wagner BK, Shen-Orr SS, Klein AM, Melton DA, Yanai I: **A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure**. *Cell Syst* 2016:1–15.

19. Muraro MJ, Dharmadhikari G, Grün D, Groen N, Dielen T, Jansen E, van Gurp L, Engelse MA, Carlotti F, de Koning EJP, van Oudenaarden A: **A Single-Cell Transcriptome Atlas of the Human**

Pancreas. *Cell Syst* 2016:1–10.

20. Segerstolpe Å, Palasantza A, Eliasson P, Andersson E-M, Andréasson A-C, Sun X, Picelli S, Sabirsh A, Clausen M, Bjursell MK, Smith DM, Kasper M, Ämmälä C, Sandberg R: **Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes.** *Cell Metab* 2016, **24**:593–607.

21. Xin Y, Kim J, Okamoto H, Ni M, Wei Y, Adler C, Murphy AJ, Yancopoulos GD, Lin C, Gromada J: **RNA Sequencing of Single Human Islet Cells Reveals Type 2 Diabetes Genes**. *Cell Metab* 2016, **24**:608–615.

22. Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A, Kircher M, Albert FW, Zeller U, Khaitovich P, Grützner F, Bergmann S, Nielsen R, Pääbo S, Kaessmann H: **The evolution of gene expression levels in mammalian organs.** *Nature* 2011, **478**:343–8.

23. Rebeiz M, Patel NH, Hinman VF: **Unraveling the Tangled Skein: The Evolution of Transcriptional Regulatory Networks in Development**. *Annu Rev Genomics Hum Genet* 2015, **16**:103–131.

24. Necsulea A, Kaessmann H: **Evolutionary dynamics of coding and non-coding transcriptomes**. *Nat Rev Genet* 2014, **15**:734–748.

25. Pishesha N, Thiru P, Shi J, Eng JC, Sankaran VG, Lodish HF: **Transcriptional divergence and conservation of human and mouse erythropoiesis.** *Proc Natl Acad Sci U S A* 2014, **111**:4103–8.

26. Shay T, Jojic V, Zuk O, Rothamel K, Puyraimond-Zemmour D, Feng T, Wakamatsu E, Benoist C, Koller D, Regev A, ImmGen Consortium the I: **Conservation and divergence in the transcriptional programs of the human and mouse immune systems.** *Proc Natl Acad Sci U S A* 2013, **110**:2946–51.

27. Godinho L, Mumm JS, Williams PR, Schroeter EH, Koerber A, Park SW, Leach SD, Wong ROL: **Targeting of amacrine cell neurites to appropriate synaptic laminae in the developing zebrafish retina**. *Development* 2005, **132**:5069–5079.

28. Li Z, Wen C, Peng J, Korzh V, Gong Z: **Generation of living color transgenic zebrafish to trace**

somatostatin-expressing cells and endocrine pancreas organization. *Differentiation* 2009, **77**:128–134.

29. Ghaye AP, Bergemann D, Tarifeño-Saldivia E, Flasse LC, Von Berg V, Peers B, Voz ML, Manfroid I: **Progenitor potential of nkx6.1-expressing cells throughout zebrafish life and during beta cell regeneration.** *BMC Biol* 2015, **13**:70.

30. Murtaugh LC: **Pancreas and beta-cell development: from the actual to the possible**. *Development* 2007, **134**:427–438.

31. Ku GM, Pappalardo Z, Luo CC, German MS, McManus MT: **An siRNA Screen in Pancreatic Beta Cells Reveals a Role for Gpr27 in Insulin Production**. *PLoS Genet* 2012, **8**.

32. Ruiz de Azua I, Scarselli M, Rosemond E, Gautam D, Jou W, Gavrilova O, Ebert PJ, Levitt P, Wess J: **RGS4 is a negative regulator of insulin release from pancreatic ?-cells in vitro and in vivo**. *Proc Natl Acad Sci U S A* 2010, **107**:7999–8004.

33. Cripps RM, Olson EN: **Control of Cardiac Development by an Evolutionarily Conserved Transcriptional Network**. *Dev Biol* 2002, **246**:14–28.

34. Kinkel MD, Prince VE: **On the diabetic menu: Zebrafish as a model for pancreas development and function**. *BioEssays* 2009, **31**:139–152.

35. Kimmel RA, Meyer D: **Zebrafish pancreas as a model for development and disease.** *Methods Cell Biol* 2016, **134**:431–61.

36. Holmstrom SR, Deering T, Swift GH, Poelwijk FJ, Mangelsdorf DJ, Kliewer S a., Macdonald RJ: **LRH-1 and PTF1-L coregulate an exocrine pancreas-specific transcriptional network for digestive function**. *Genes Dev* 2011, **25**:1674–1679.

37. Fagerberg L, Hallström BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, Habuka M, Tahmasebpoor S, Danielsson A, Edlund K, Asplund A, Sjöstedt E, Lundberg E, Szigyarto CA-K, Skogs M, Takanen JO, Berling H, Tegel H, Mulder J, Nilsson P, Schwenk JM, Lindskog C, Danielsson F, Mardinoglu A, Sivertsson Å, von Feilitzen K, Forsberg M, Zwahlen M, Olsson I, Navani S, et al.: **Analysis of the Human Tissue-specific Expression by Genome-wide Integration of Transcriptomics and Antibody-based Proteomics**. *Mol Cell Proteomics* 2014, **13**:397–406.

38. Moran VA, Perera RJ, Khalil AM: **Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs**. *Nucleic Acids Res* 2012, **40**:6391–6400.

39. Pagliuca FW, Melton DA: **How to make a functional β-cell.** *Development* 2013, **140**:2472–83.

40. Inagaki N, Kuromi H, Gonoi T, Okamoto Y, Ishida H, Seino Y, Kaneko T, Iwanaga T, Seino S: **Expression and role of ionotropic glutamate receptors in pancreatic islet cells.** *FASEB J* 1995, **9**:686–691.

41. Villasenor A, Wang Z V, Rivera LB, Ocal O, Asterholm IW, Scherer PE, Brekken RA, Cleaver O, Wilkie TM: **Rgs16 and Rgs8 in embryonic endocrine pancreas and mouse models of diabetes.** *Dis Model Mech* 2010, **3**:567–80.

42. Zhang W, Morris QD, Chang R, Shai O, Bakowski M a, Mitsakakis N, Mohammad N, Robinson MD, Zirngibl R, Somogyi E, Laurin N, Eftekharpour E, Sat E, Grigull J, Pan Q, Peng W-T, Krogan N, Greenblatt J, Fehlings M, van der Kooy D, Aubin J, Bruneau BG, Rossant J, Blencowe BJ, Frey BJ, Hughes TR: **The functional landscape of mouse gene expression.** *J Biol* 2004, **3**:21.

43. Hale MA, Swift GH, Hoang CQ, Deering TG, Masui T, Lee Y-K, Xue J, MacDonald RJ: **The nuclear hormone receptor family member NR5A2 controls aspects of multipotent progenitor cell formation and acinar differentiation during pancreatic organogenesis.** *Development* 2014, **141**:3123–33.

44. Jiang Z, Song J, Qi F, Xiao A, An X, Liu N, Zhu Z, Zhang B, Lin S: **Exdpf is a key regulator of exocrine pancreas development controlled by retinoic acid and ptf1a in zebrafish.** *PLoS Biol* 2008, **6**:e293.

45. Binot A-C, Manfroid I, Flasse L, Winandy M, Motte P, Martial JA, Peers B, Voz ML: **Nkx6.1 and nkx6.2 regulate α- and β-cell formation in zebrafish by acting on pancreatic endocrine progenitor cells**. *Dev Biol* 2010, **340**:397–407.

46. Nica AC, Ongen H, Irminger J-CC, Bosco D, Berney T, Antonarakis SE, Halban PA, Dermitzakis ET: **Cell-type, allelic, and genetic signatures in the human pancreatic beta cell transcriptome**. *Genome Res* 2013, **23**:1554–1562.

47. Devos N, Deflorian G, Biemar F, Bortolussi M, Martial JA, Peers B, Argenton F: **Differential**

expression of two somatostatin genes during zebrafish embryonic development. *Mech Dev* 2002, **115**:133–137.

48. Blodgett DM, Nowosielska A, Afik S, Pechhold S, Cura AJ, Kennedy NJ, Kim S, Kucukural A, Davis RJ, Kent SC, Greiner DL, Garber MG, Harlan DM, dilorio P: **Novel Observations From Next-Generation RNA Sequencing of Highly Purified Human Adult and Fetal Islet Cell Subsets**. *Diabetes* 2015, **64**:3172–3181.

49. Tian G, Sandler S, Gylfe E, Tengholm A: **Glucose- and hormone-induced cAMP oscillations in α- and β-cells within intact pancreatic islets.** *Diabetes* 2011, **60**:1535–43.

50. Davidson AJ, Ernst P, Wang Y, Dekens MPS, Kingsley PD, Palis J, Korsmeyer SJ, Daley GQ, Zon LI: **cdx4 mutants fail to specify blood progenitors and can be rescued by multiple hox genes**. *Nature* 2003, **425**:300–306.

51. Kinkel MD, Eames SC, Alonzo MR, Prince VE: **Cdx4 is required in the endoderm to localize the pancreas and limit beta-cell number**. *Development* 2008, **135**:919–929.

52. Bertrand G, Gross R, Puech R, Loubatières-Mariani MM, Bockaert J: **Evidence for a glutamate receptor of the AMPA subtype which mediates insulin release from rat perfused pancreas.** *Br J Pharmacol* 1992, **106**:354–359.

53. Lee EK, Kim W, Tominaga K, Martindale JL, Yang X, Subaran SS, Carlson OD, Mercken EM, Kulkarni RN, Akamatsu W, Okano H, Perrone-Bizzozero NI, de Cabo R, Egan JM, Gorospe M: **RNA-binding protein HuD controls insulin translation.** *Mol Cell* 2012, **45**:826–35.

54. Mergler S, Singh V, Grötzinger C, Kaczmarek P, Wiedenmann B, Strowski MZ: **Characterization of voltage operated R-type Ca2+ channels in modulating somatostatin receptor subtype 2- and 3-dependent inhibition of insulin secretion from INS-1 cells**. *Cell Signal* 2008, **20**:2286–2295.

55. Squires PE, Jones PM, Younis MYG, Hills CE: **Chapter Ten - The Calcium-Sensing Receptor and β-Cell Function**. In *Vitamins & Hormones*. *Volume 95*. Edited by Litwack G. Academic Press; 2014:249–267. [*The Pancreatic Beta Cell*]

56. Cortijo C, Gouzi M, Tissir F, Grapin-Botton A: **Planar Cell Polarity Controls Pancreatic Beta Cell Differentiation and Glucose Homeostasis**. *Cell Rep* 2012, **2**:1593–1606.

57. Flanagan SE, De Franco E, Lango Allen H, Zerah M, Abdul-Rasoul MM, Edge JA, Stewart H, Alamiri E, Hussain K, Wallis S, de Vries L, Rubio-Cabezas O, Houghton JAL, Edghill EL, Patch A-M, Ellard S, Hattersley AT: **Analysis of transcription factors key for mouse pancreatic development establishes NKX2-2 and MNX1 mutations as causes of neonatal diabetes in man.** *Cell Metab* 2014, **19**:146–54.

58. Jennings RE, Berry AA, Strutt JP, Gerrard DT, Hanley NA: **Human pancreas development**. *Development* 2015, **142**.

59. Kilic G, Alvarez-Mercado AI, Zarrouki B, Opland D, Liew CW, Alonso LC, Myers MG, Jonas J-C, Poitout V, Kulkarni RN, Mauvais-Jarvis F: **The Islet Estrogen Receptor-α Is Induced by Hyperglycemia and Protects Against Oxidative Stress-Induced Insulin-Deficient Diabetes**. *PLoS One* 2014, **9**.

60. Yuchi Y, Cai Y, Legein B, De Groef S, Leuckx G, Coppens V, Van Overmeire E, Staels W, De Leu N, Martens G: **Estrogen Receptor α Regulates Beta Cell Formation During Pancreas Development and Following Injury**. *Diabetes* 2015:db141798.

61. Ye L, Robertson MA, Hesselson D, Stainier DYR, Anderson RM: **glucagon is essential for alpha cell transdifferentiation and beta cell neogenesis**. *Development* 2015, **142**:1407–1417.

62. Chera S, Baronnier D, Ghila L, Cigliola V, Jensen JN, Gu G, Furuyama K, Thorel F, Gribble FM, Reimann F, Herrera PL: **Diabetes recovery by age-dependent conversion of pancreatic δ-cells into insulin producers.** *Nature* 2014, **514**:503–7.

63. Courtney M, Gjernes E, Druelle N, Ravaud C, Vieira A, Ben-Othman N, Pfeifer A, Avolio F, Leuckx G, Lacas-Gervais S, Burel-Vandenbos F, Ambrosetti D, Hecksher-Sorensen J, Ravassard P, Heimberg H, Mansouri A, Collombat P: **The inactivation of Arx in pancreatic α-cells triggers their neogenesis and conversion into functional β-like cells.** *PLoS Genet* 2013, **9**:e1003934.

64. Gao T, McKenna B, Li C, Reichert M, Nguyen J, Singh T, Yang C, Pannikar A, Doliba N, Zhang T, Stoffers DA, Edlund H, Matschinsky F, Stein R, Stanger BZ: **Pdx1 maintains β-cell identity and function by repressing an α-cell program**. *Cell Metab* 2014, **19**:259–271.

65. Thorel F, Nepote V, Avril I, Kohno K, Desgraz R, Chera S, Herrera PL: **Conversion of adult**

pancreatic alpha-cells to beta-cells after extreme beta-cell loss. *Nature* 2010, **464**:1149–1154.

66. Spijker HS, Ravelli RBG, Mommaas-Kienhuis AM, van Apeldoorn AA, Engelse MA, Zaldumbide A, Bonner-Weir S, Rabelink TJ, Hoeben RC, Clevers H, Mummery CL, Carlotti F, de Koning EJP: **Conversion of mature human β-cells into glucagon-producing α-cells.** *Diabetes* 2013, **62**:2471–80.

67. Wendik B, Maier E, Meyer D: **Zebrafish mnx genes in endocrine and exocrine pancreas formation**. *Dev Biol* 2004, **268**:372–383.

68. Pan FC, Brissova M, Powers AC, Pfaff S, Wright CVE: **Inactivating the permanent neonatal diabetes gene Mnx1 switches insulin-producing β-cells to a δ-like fate and reveals a facultative proliferative capacity in aged β-cells.** *Development* 2015, **142**:3637–48.

69. Nelson SB, Schaffer AE, Sander M: **The transcription factors Nkx6.1 and Nkx6.2 possess equivalent activities in promoting beta-cell fate specification in Pdx1+ pancreatic progenitor cells**. *Development* 2007, **134**:2491–2500.

70. Flasse LC, Pirson JL, Stern DG, Von Berg V, Manfroid I, Peers B, Voz ML: **Ascl1b and Neurod1, instead of Neurog3, control pancreatic endocrine cell fate in zebrafish**. *BMC Biol* 2013, **11**:78.

71. Flasse LC, Stern DG, Pirson JL, Manfroid I, Peers B, Voz ML: **The bHLH transcription factor Ascl1a is essential for the specification of the intestinal secretory cells and mediates Notch signaling in the zebrafish intestine.** *Dev Biol* 2013, **376**:187–97.

72. Rossi A, Kontarakis Z, Gerri C, Nolte H, Hölper S, Krüger M, Stainier DYR: **Genetic compensation induced by deleterious mutations but not gene knockdowns.** *Nature* 2015, **524**:230–3.

73. McIntyre DC, Rakshit S, Yallowitz AR, Loken L, Jeannotte L, Capecchi MR, Wellik DM: **Hox patterning of the vertebrate rib cage**. *Development* 2007, **134**.

74. Hummler E, Cole TJ, Blendy JA, Ganss R, Aguzzi A, Schmid W, Beermann F, Schütz G: **Targeted mutation of the CREB gene: compensation within the CREB/ATF family of transcription factors.** *Proc Natl Acad Sci U S A* 1994, **91**:5647–51.

75. Wang S, Zhang J, Zhao A, Hipkens S, Magnuson MA, Gu G: **Loss of Myt1 function partially**

compromises endocrine islet cell differentiation and pancreatic physiological function in the mouse. *Mech Dev* 2007, **124**:898–910.

76. Wang S, Hecksher-Sorensen J, Xu Y, Zhao A, Dor Y, Rosenberg L, Serup P, Gu G: **Myt1 and Ngn3 form a feed-forward expression loop to promote endocrine islet cell differentiation.** *Dev Biol* 2008, **317**:531–40.

77. Tennant BR, Islam R, Kramer MM, Merkulova Y, Kiang RL, Whiting CJ, Hoffman BG: **The transcription factor Myt3 acts as a pro-survival factor in β-cells.** *PLoS One* 2012, **7**:e51501.

78. Brereton H, Carvell MJ, Persaud SJ, Jones PM: **Islet alpha-cells do not influence insulin secretion from beta-cells through cell-cell contact.** *Endocrine* 2007, **31**:61–5.

79. Ohta Y, Kosaka Y, Kishimoto N, Wang J, Smith SB, Honig G, Kim H, Gasa RM, Neubauer N, Liou A, Tecott LH, Deneris ES, German MS: **Convergence of the insulin and serotonin programs in the pancreatic β-cell.** *Diabetes* 2011, **60**:3208–16.

80. Kwan KM, Fujimoto E, Grabher C, Mangum BD, Hardy ME, Campbell DS, Parant JM, Yost HJ, Kanki JP, Chien CB: **The Tol2kit: a multisite gateway-based construction kit for Tol2 transposon transgenesis constructs**. *Dev Dyn* 2007, **236**:3088–3099.

81. Zecchin E, Filippi A, Biemar F, Tiso N, Pauls S, Ellertsdottir E, Gnügge L, Bortolussi M, Driever W, Argenton F: **Distinct delta and jagged genes control sequential segregation of pancreatic cell types from precursor pools in zebrafish**. *Dev Biol* 2007, **301**:192–204.

82. Ramsköld D, Luo S, Wang Y-C, Li R, Deng Q, Faridani OR, Daniels G a, Khrebtukova I, Loring JF, Laurent LC, Schroth GP, Sandberg R: **Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells.** *Nat Biotechnol* 2012, **30**:777–82.

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

84. Moss SP, Joyce DA, Humphries S, Tindall KJ, Lunt DH: **Comparative Analysis of Teleost Genome Sequences Reveals an Ancient Intron Size Expansion in the Zebrafish Lineage**. *Genome Biol Evol* 2011, **3**:1187–1196.

85. Anders S, Pyl PT, Huber W: **HTSeq—a Python framework to work with high-throughput sequencing data**. *Bioinformatics* 2014:btu638.

86. Anders S, Huber W: **Differential expression analysis for sequence count data**. *Genome Biol* 2010, **11**:R106.

87. Love MI, Huber W, Anders S: **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.** *Genome Biol* 2014, **15**:550.

88. Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, Birney E: **EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates**. *Genome Res* 2009, **19**:327–335.

89. Huang DW, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources**. *Nat Protoc* 2008, **4**:44–57.

90. Thisse C, Thisse B: **High-resolution in situ hybridization to whole-mount zebrafish embryos**. *Nat Protoc* 2008, **3**:59–69.

91. Lillesaar C, Tannhäuser B, Stigloher C, Kremmer E, Bally-Cuif L: **The serotonergic phenotype is acquired by converging genetic mechanisms within the zebrafish central nervous system**. *Dev Dyn* 2007, **236**:1072–1084.

92. Mavropoulos A, Devos N, Biemar F, Zecchin E, Argenton F, Edlund H, Motte P, Martial JA, Peers B: **sox4b is a key player of pancreatic alpha cell differentiation in zebrafish.** *Dev Biol* 2005, **285**:211–23.

93. Jao L-EE, Wente SR, Chen W: **Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system.** *Proc Natl Acad Sci U S A* 2013, **110**:13904–9.

94. Varshney GK, Pei W, LaFave MC, Idol J, Xu L, Gallardo V, Carrington B, Bishop K, Jones M, Li M, Harper U, Huang SC, Prakash A, Chen W, Sood R, Ledin J, Burgess SM: **High-throughput gene targeting and phenotyping in zebrafish using CRISPR/Cas9.** *Genome Res* 2015, **25**:1030–42.

Figure Legend

Fig 1. Global analysis of the zebrafish pancreatic RNA-seq data

(A) Principal component analyses (PCA) of gene VSD (Variance stabilizing transformation) calculated by DESeq package for the 16 zebrafish pancreatic data sets. (B) PCA of gene VSD for beta, alpha and delta cells (9 samples in total). The PC plots show a close clustering of all replicates and distinct clusters for each pancreatic cell type. PCs were calculated using all the 33726 genes annotated on Zv9 version 75 ensembl.

Fig 2. Transcriptomic signatures of zebrafish endocrine, ductal and acinar cells

(A) Venn diagram showing the number of genes with acinar-, endocrine- and ductal-enriched expression obtained with DESeq2 using a cut-off ratio of 4-fold and adjust P values <0.05. (B) Heatmap plot showing the expression pattern of all differentially expressed genes. Genes listed on the right side of the plot are examples of either known markers (black) or new genes with endocrine, acinar and ductal enriched expression discovered by this analysis (red). The three endocrine samples were generated in silico by combining the raw data obtained from alpha-, beta- and delta-cells as described in Materials and Methods. (C and D) Gene ontology (GO) enrichment analysis for the 1853 genes enriched in endocrine cells (C) and the 492 genes enriched in acinar cells (D) displaying the most enriched biological process.

Fig 3. Conservation of the pancreatic endocrine signature among vertebrates.

(A) Principle component analysis (PCA) performed on human and mouse whole pancreas and islet RNA-seq datasets and including the zebrafish endocrine and acinar datasets. The analysis was done using the 9393 genes displaying 1-1-1 orthology relationship between zebrafish, mouse and human using a total of 24 RNA-seq samples. The endocrine data sets of zebrafish, mouse and human cluster along the PC1 axis representing 55% of the variance, indicating a conserved endocrine signature. The human pancreatic sample enriched in acinar tissue [9] clusters with the human whole pancreatic samples [37], due to the very high proportion of acinar cells in pancreas. (B) Venn diagram showing the number of endocrine-enriched genes found only in zebrafish, mouse or human and those displaying conserved endocrine-enrichment in two species or in the three species (shown in intersections). Due to

gene duplications in some species and often in zebrafish, the number of corresponding murine (M) or human (H) orthologous genes is given in brackets in each intersection. The full List of conserved endocrine-enriched genes is given in S4 Table.

Fig 4. Expression of genes in endocrine cells of the dorsal pancreatic bud.

Whole-mount ISH on zebrafish embryos showing endocrine pancreatic expression of some new zebrafish endocrine markers. Genes with conserved endocrine-enriched expression in Zebrafish, Human and Mouse (ZHM), or endocrine-enriched in Zebrafish and Mouse (ZH), or in Zebrafish and Human (HM) are indicated. Z: Gene endocrine enriched only in Zebrafish. N-orth: Endocrine enriched zebrafish gene with no obvious human or mouse ortholog. Arrowheads indicate the location of the dorsal pancreatic bud containing embryonic endocrine cells and insets at the top-right display higher magnification view of the pancreatic bud.

Fig 5. Zebrafish genes differentially expressed in the endocrine cell subtypes

(A) Venn diagram displaying the number of endocrine genes with Alpha-, Beta- and delta-enriched expression identified with DESeq algorithm based on a cut-off ratio of 4-fold and adjusted P-value<0.05. (B) Heatmap plot showing the expression pattern of all differentially expressed genes. Genes listed at the right side of the plot are some examples of either known (black) or new (red) markers identified in this analysis. (C) Principal Component Analysis performed on the 9 zebrafish endocrine RNAseq datasets using the 1853 endocrine enriched genes. Compared to the PC plot of Fig1B performed on all annotated zebrafish genes (33726 genes), this plot shows a tighter clustering of all replicates and better discrimination between the three endocrine cell subtypes.

Fig 6. Expression of *ppdpfb* and *pcsk2* genes in zebrafish beta-cells

Co-labeling by fluorescent *in situ* hybridization (FISH) of *ppdpfb* and *pcsk2* with insulin, glucagon and

somatostatin. (A-C) *ppdpfb* is mainly expressed by beta cells (A, arrows) and in few alpha cells (B, arrows). No expression of *ppdpfb* was observed in delta cells (C). D-F. beta cells specifically expressed *pcsk2* (D, arrows) while no expression was detected in alpha or delta cells (E and F)

Fig 7. Identification of genes with conserved enriched expression in alpha- and beta-cells.

Venn diagrams showing the number of genes presenting enriched expression in alpha-cells (right panel) and beta-cells (left panel) and displaying this enrichment in zebrafish, mice and/or human (shown in intersections). Due to gene duplications in some species and often in zebrafish, the number of corresponding murine (M) or human (H) orthologous genes is given in brackets in each intersection. The alpha- and beta- enriched genes were selected by DESeq2 with fold change>4 and adjusted P<0.05 using 2 murine alpha and beta cell preparations [12], 6 human alpha and beta cell preparations [48] and the 3 zebrafish alpha and beta cell preparation (this study). The full list of conserved beta- and alpha-enriched genes is given in S8 and S9 Tables.

Fig 8. Delta-cell differentiation is disrupted in the cdx4 mutant embryos.

Analysis of wild-type and *cdx4*^{tv205} -/- mutant embryos at 48 hpf by WHISH (A) and FISH (B-D) using delta-cell markers (*sst2, lamc2, slc7a7* and *map3k15*) as well as *gcga* and *insulin*. *cdx4*^{tv205} mutants display a loss of *sst2* (A), *lamc2* (B) and *slc7a7* (C) pancreatic expression and an increase of beta cells. *Map3k15* expression is strongly reduced in the pancreatic islet (yellow arrows) while not affected in the presumptive pronephric glomeruli (white arrows)(D). No obious effect is observed on *gcga* expression. Nuclei are stained with DAPI (grey staining).

Fig 9. Reduced glucagon expression in the *myt1b* zebrafish mutant embryos.

(A-D) Glucagon expression was analyzed by immunofluorescence in F2 *myt1b*^{ulg019} homozygous mutant larvae and in sibling +/+ (WT) and +/- larvae at 2dpf. The volume of the alpha cell mass was determined in each larva by the imaging software Imaris. The graph shows the quantification measured for all

myt1b-/- larvae and for +/- or +/+ siblings, indicating a slight by statistical significative decrease of alpha cell mass.

Supplementary figure legends

S1 Fig. Expression of the transgenes *Tg(ins:GFP)*, *Tg(gcga:GFP)*, *Tg(sst2:GFP)* and *Tg(ins:NTR:mCherry)* in the endocrine pancreatic cell types.

(A-C) Whole mount immunostaining of 3dpf transgenic larvae *Tg(ins:GFP)* demonstrating the selective expression of GFP in beta-cells. (D-I) Immunostaining of pancreas sections from adult transgenic fish *Tg(gcga:GFP/ins:NTR:mCherry)*. Glucagon+ cells (see arrows in D-F) express high level of GFP and are not labelled by mCherry , while many insulin expressing cells are labelled by mCherry and by GFP (at slightly lower levels)(see arowheads in G-I). These data reveals a leaking expression of the gcga:GFP transgene in beta-cells. (J-L) ISH performed on pancreas section of adult Tg(sst2:gfp) with sst2 probe followed by immunofluorescence using the GFP antibody; the expression of endogenous sst2 gene co-localize with GFP staining confirming the specific expression of the transgene in delta-cells [28].

S2 fig. Identification of genes with evolutionary conserved and enriched expression in pancreatic exocrine cells.

Venn diagram showing the number of exocrine-enriched genes found only in zebrafish, mouse or human and those displaying conserved endocrine-enrichment in two species or in the three species (shown in intersections). Due to gene duplications in some species and often in zebrafish, the number of corresponding murine (M) or human (H) orthologous genes is given in brackets in each intersection. The full list of conserved exocrine-enriched genes is given in S5 Table.

S3 fig . Gene Ontology (GO) enrichment analysis for endocrine cell subtypes.

Left. Bar plot displayig the number of genes constituting enriched GO terms. P-values are denoted on

the bars. Right. Fold of change (in Log2) of genes constituting the most enriched GO terms, (A) GO enrichment for the 70 beta-enriched genes. (B) GO enrichment for the 73 alpha-enriched genes. (C) GO enrichment for the 192 delta-enriched genes. (B-A: pairwise Beta versus Alpha; B-D: pairwise Beta versus Delta).

S4 fig. Expression of *spon1b* gene in beta pancreatic cells of zebrafish embryos.

Co-labeling by FISH of *spon1b* with insulin (*ins*) (A, arrows show colocalization), while no expression was detected in delta cells (B). *ins*: insulin, *sst2*: somatostatin 2, N-orth: Endocrine enriched zebrafish gene with no described ortholog in Human and/or Mouse.

S5 fig. Expression of *pnoca* and *scinlb* genes in alpha pancreatic cells of zebrafish embryos.

Co-labeling by *in situ* hybridization at 30hpf of new discovered genes with cell type specific markers for alpha and beta cells. A and B. *pnoca* is expressed in alpha cells (A, arrows) while no expression was detected in beta cells (B). C and D. *scinlb* was detected specifically in alpha cells (C, arrows) but not detected in beta cells (D). gcga: glucagon a, ins: insulin, Z: no endocrine gene with no conserved expression, N-orth: Endocrine enriched zebrafish gene with no described ortholog in Human and/or Mouse.

S6 fig. Validation of the selective expression of some genes in zebrafish delta-cells.

cdx1b, *cdx4*, *lamc2*, *map3k15* are specifically expressed in delta cells at 24hpf (A, C, E, G, arrows show colocalization with delta cell specific markers, somatostatin 2). No expression was detected for none of the genes in beta cells (B, D, F, H). *ins*: *insulin, sst2: somatostatin 2*, ZHM: Gene expression conserved in Zebrafish, human and Mouse, ZM: Gene expression conserved in Zebrafish and Mouse, Z: Endocrine gene with no conserved expression, N-orth: Endocrine enriched zebrafish gene with no described ortholog in Human and/or Mouse.

S7 fig. Expression and function of myt1b in zebrafish pancreas.

A : Whole-mount ISH showing expression pattern of *myt1b* in zebrafish embryos at 27 hpf. High expression is detected in the dorsal pancreatic bud (indicated by the arrow) and in the central nervous system. B : Quantification of glucagon and insulin expression at 48 hpf in wild-type (non-injected) embryos and in embryos injected with the 4 CRISPR *myt1a/b* guide RNA and Cas9. Graph B shows the volume of all gcga+ cells and all ins+ cells measured in each embryo by the imaging software Imaris (see M&M). This quantification indicates a statistically significant reduction of the volume of alpha cell mass while beta cell mass is not drastically affected in the injected (F0) embryos (results of 1 experiment).

Supplementary table

S1 table. Gene expression levels in alpha-, beta- and delta-cell subtypes.

The expression is given in Normalized counts for the three replicates of alpha-, beta- and delta-cell libraries.

S2 table. Gene expression levels in endocrine, acinar and ductal cells.

The expression is given in Normalized counts for the four replicates of acinar cells, the three replicates of ductal and endocrine cells. The 3 endocrine data sets were obtained by combining the reads obtained with alpha-, beta- and delta-cell libraries as described in Materials and Methods (e.g., endocrine1 data is a mix of alpha1, beta1 and delta1; endocrine2 data is a mix of alpha2, beta2 and delta3, …)

S3 table. Classification of genes according to their enriched expression in endocrine, acinar or ductal cells.

Expression levels were compared for each gene between endocrine, acinar and ductal cells (pair-wise comparison with DEseq2). Classification was performed using the cut-off values of FC>4 and adjusted P value<0.05). The enrichment is given for each gene as Log2 of Fold change and with the adjusted P value.

S4 table. Genes presenting conserved endocrine expression.

The ZMH excel page shows the genes with endocrine enriched expression in Zebrafish, Mice and Human. The ZM, ZH and HM excel pages show genes with endocrine enriched expression in two species (ZM : Zebrafish and Mice; ZH: Zebrafish and Human; HM: Human and Mice). The tables give the expression mean in endocrine cells (Normalized counts) and the enrichment (FC: Fold Change) in the three species.

S5 table. Genes presenting conserved exocrine expression.

The ZMH excel page shows the genes with exocrine enriched expression in Zebrafish, Mice and Human.

The ZM, ZH and HM excel pages show genes with endocrine enriched expression in two species (ZM : Zebrafish and Mice; ZH: Zebrafish and Human; HM: Human and Mice). The tables give the expression mean in endocrine cells (Normalized counts) and the enrichment (FC: Fold Change) in the three species.

S6 table. List of endocrine-enriched genes with an expression level above 100 counts in embryonic endocrine cells at 27 hpf.

List of 911 genes with endocrine enriched expression in adult and showing an expression level above 100 Normalized counts. Gene expression is provided for the three replicates (selection of pancreatic cells at 27 hpf from Tg(pax6:GFP)) as well as the mean expression level.

S7 table. Classification of endocrine-enriched genes according to their expression in alpha, beta and delta endocrine cells.

Gene expression ratio was calculated for all pair-wise comparison (alpha versus beta : A&B; beta versus delta : B&D and alpha versus delta : A&D). The enrichment is given as Log2 Fold change with the adjust value. Classification was done using the cut-off threshold of Log2FC>2 and adjP<0.05.

S8 table. Conserved beta-cell markers.

The ZMH excel page shows the genes with beta-cell enriched expression in Zebrafish, Mice and Human. The ZM, ZH and HM excel pages show genes with beta-cell enriched expression in two species (ZM : Zebrafish and Mice; ZH: Zebrafish and Human; HM: Human and Mice). Beta-cell versus alpha-cell expression ratio is given as Fold change (FC) for each species. The ZMH, ZH and HM genes also detected as enriched in specific pancreatic cell types in the human single cell transcriptomic studies [19–21] are also noted in the columns on the right part.

S9 table. Conserved alpha-cell markers.

The ZMH excel page shows the genes with alpha-cell enriched expression in Zebrafish, Mice and Human. The ZM, ZH and HM excel pages show genes with alpha-cell enriched expression in two

species (ZM : Zebrafish and Mice; ZH: Zebrafish and Human; HM: Human and Mice). Alpha-cell versus beta-cell expression ratio is given as Fold change (FC) for the three species. The ZMH, ZH and HM genes also detected as enriched in specific pancreatic cell types in the human single cell transcriptomic studies [19–21] are also noted in the columns on the right part

S10 table. Primers used for RNA probe synthesis.

List of primers used for the synthesis of the RNA probes for *in situ* hybridization, antisense primers contains the T3 minimal promoter sequence.