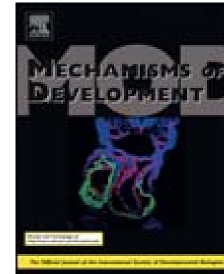
Accepted Manuscript

The global gene expression profile of the secondary transition during pancreatic development



Stefanie J. Willmann, Nikola S. Mueller, Silvia Engert, Michael Sterr, Ingo Burtscher, Aurelia Raducanu, Martin Irmler, Johannes Beckers, Steffen Sass, Fabian J. Theis, Heiko Lickert

PII: S0925-4773(15)30037-X
DOI: doi: 10.1016/j.mod.2015.11.004
Reference: MOD 3386

To appear in:

Received date: 19 June 2015
Revised date: 26 November 2015
Accepted date: 27 November 2015

Please cite this article as: Willmann, Stefanie J., Mueller, Nikola S., Engert, Silvia, Sterr, Michael, Burtscher, Ingo, Raducanu, Aurelia, Irmler, Martin, Beckers, Johannes, Sass, Steffen, Theis, Fabian J., Lickert, Heiko, The global gene expression profile of the secondary transition during pancreatic development, (2015), doi: 10.1016/j.mod.2015.11.004

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

The global gene expression profile of the secondary transition during pancreatic development

Stefanie J. Willmann*[1,5], Nikola S. Mueller*[2], Silvia Engert[1], Michael Sterr[1], Ingo Burtscher[1], Aurelia Raducanu[1], Martin Irmler[3], Johannes Beckers[3,4,5], Steffen Sass[2], Fabian J. Theis[2,6], Heiko Lickert[1,5,7,**]

[1] Institute of Diabetes and Regeneration Research, Institute of Stem Cell Research, Helmholtz Zentrum München, Business Campus Garching, Parkring 11, 85748 Garching, Germany

[2] Institute of Computational Biology, Helmholtz Zentrum München, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany

[3] Institute of Experimental Genetics, Helmholtz Zentrum München, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany

[4] Chair of Experimental Genetics, Center of Life and Food Sciences Weihenstephan, Technische Universität München, Germany

[5] German Center for Diabetes Research (DZD), Ingolstädter Landstraße 1, 85764 Neuherberg, Germany

[6] Department of Mathematics, Technische Universität München, Boltzmannstraße 3 85748 Garching, Germany

[7] Faculty of Medicine, Technische Universität München, Ismaninger Str. 22, 81675 München, Germany

[*]these authors contributed equally to this work

[**]corresponding author. Contact details: Heiko Lickert, Institute of Diabetes and Regeneration Research, Parkring 11, 85748 Garching, phone +49 89 3187-3760, heiko.lickert@helmholtz-muenchen.de

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Abstract

Pancreas organogenesis is a highly dynamic process where neighboring tissue interactions lead to dynamic changes in gene regulatory networks that orchestrates endocrine, exocrine, and ductal lineage formation. To understand the spatio-temporal regulatory logic we have used the Forkhead transcription factor Foxa2-Venus fusion (FVF) knock-in reporter mouse to separate the FVF^{+} pancreatic epithelium from the FVF^{-} surrounding tissue (mesenchyme, neurons, blood, and blood vessels) to perform a genome-wide mRNA expression profiling at embryonic days (E) 12.5-15.5. Annotating genes and molecular processes suggest that FVF marks endoderm-derived multipotent epithelial progenitors at several lineage restriction steps, when the bulk of endocrine, exocrine and ductal cells are formed during secondary transition. In the pancreatic epithelial compartment, we identified most known endocrine and exocrine lineage determining factors and diabetes-associated genes, but also unknown genes with spatio-temporal regulated pancreatic expression. In the non-endoderm-derived compartment, we identified many well-described regulatory genes that are not yet functionally annotated in pancreas development, emphasizing that neighboring tissue interactions are still ill defined. Pancreatic expression of over 635 genes was analyzed with the mRNA *in situ* hybridization Genepaint public database. This validated the quality of the profiling data set and identified hundreds of genes with spatially restricted expression patterns in the pancreas. Some of these genes are also targeted by pancreatic transcription factors and show active chromatin marks in human islets of Langerhans. Thus, with highest spatio-temporal resolution of a global gene expression profile during secondary transition, our study enables to shed light on neighboring tissue interactions, developmental timing and diabetes gene regulation.

Keywords: Pancreas, endocrine, exocrine, ductal, organogenesis, Foxa2.

ACCEPTED MANUSCRIPT

Highlights:

- **High-resolution spatio-temporal gene expression profile of pancreas development**
- **Cell type-specific profile reveals known and many novel developmental regulators**
- **Differential expression in pancreatic compartments reveal details about tissue interactions**
- ***In situ* mRNA expression analysis validates tissue specificity and spatial expression**
- **The established resource is of unique value to study mouse and human development and diabetes**

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Introduction

Diabetes is an epidemic disease and caused 4.9 million deaths worldwide in 2014 (IDF Diabetes Atlas 2014). Type 1 (T1D) and type 2 diabetes (T2D) is triggered by autoimmune destruction of insulin-producing β-cells or by acquired insulin resistance with steady decline of functional β-cell mass, respectively. Current treatments significantly improve life quality of patients, however, they do not provide full glycemic control leading to long-term micro- and macrovascular complications. New treatment strategies aiming at restoration of β-cell mass could normalize blood glucose control and eliminate disease complications (Bonner-Weir et al., 2005).

The insulin-producing β-cells are part of the endocrine pancreas, the islets of Langerhans, which also consists of glucagon-producing α-cells, somatostatin-producing δ-cells, ghrelin-producing ε-cells and pancreatic polypeptide-producing PP-cells (In't Veld and Marichal, 2010). These endocrine cells secrete hormones into the blood stream to regulate nutrient metabolism and glucose homeostasis. The exocrine compartment of the pancreas produces digestive enzymes from its acinar cells, which are then drained via a ductal system into the duodenum.

The Forkhead transcription factor Foxa2 regulates endoderm formation and epithelialization during gastrulation (Burtscher and Lickert, 2009). Foxa1 and Foxa2 are crucial upstream regulators of the pancreatic and duodenal homeobox 1 (Pdx1) transcription factor (Gao et al., 2008). Pancreas development starts with the patterning of the foregut endoderm and the specification of the pre-pancreatic field marked by Pdx1 (Jensen et al., 2004; Pan and Wright 2011; Zorn and Wells, 2009). By E9.5, the first visible pancreatic epithelial buds emerge from the dorsal and ventral foregut and expand into the surrounding mesenchyme. Growth and expansion of the pancreatic buds are associated with the first wave of differentiation and appearance of glucagon-positive cells during the so-called primary transition (E9.5-12.5). By the end of the primary transition, these two buds rotate and fuse together and active growth and epithelial remodeling shape the future pancreas. During the secondary transition (E12.5-15.5), multipotent pancreatic progenitors become committed to the ductal, endocrine and acinar lineages. Several transcription factors such as Pdx1, Ptf1a,

ACCEPTED MANUSCRIPT

Hnf1a, Hnf1b, Hnf4a, Nkx6.1, and Sox9, as well as signaling molecules, such as FGF, EGF, Wnt, Bmp, Shh and Notch have been shown to coordinate commitment and differentiation of the pancreatic lineages concomitant with epithelial branching morphogenesis (Puri and Hebrok, 2010; Pan and Wright, 2011; Raducanu and Lickert, 2012; Pagliuca and Melton, 2013; Shih et al., 2013; Migliorini et al., 2014). Inductive instructions relayed from the surrounding mesenchyme as well as epithelial polarization contribute to the patterning of the pancreatic epithelium into distinct trunk and tip domains. The tip domain, at least for a certain period of time, is the supplier of multipotent progenitors that can generate the endocrine pool as well as ductal and acinar cells (Zhou et al., 2007). Endocrine lineage segregation is initiated by the induction of Neurogenin 3 (Ngn3), which marks the cells that are committed to an endocrine fate and will start migrating out from the ductal epithelium into the surrounding mesenchyme where they cluster together to form the islets of Langerhans (Gradwohl et al., 2000; Schwitzgebel et al., 2000). Cell delamination, asymmetric cell division and epithelial-mesenchymal transition (EMT) have been proposed to govern the endocrine precursors' delineation from the trunk epithelium, however, the mechanisms of pancreatic lineage segregation and neighboring tissue interactions are only beginning to be understood.

To systematically profile pancreas development, we captured the spatio-temporal global gene expression of the endoderm-derived epithelial and non-endodermal compartment during the secondary transition. Pancreatic organs were isolated, dissociated and sorted for epithelial (Foxa2-Venus fusion positive, FVF^+) and non-endodermal (FVF^-) populations and subjected for mRNA profiling on four consecutive days between E12.5-15.5. Extensive statistical analyses, using principal component analysis and linear regression modeling, clearly identified two distinct tissue compartments solely based on their transcriptional profile. Subsequent bioinformatical analyses, using pathway analysis of the profiles, underpin the importance of the established spatio-temporal genome-wide expression resource. We demonstrate how the provided data resource can be mined to advance mechanistic understanding of the secondary transition, neighboring tissue interactions, and diabetes gene regulation. For further validation, we analyzed and classified mRNA expression patterns using the Genepaint database (www.genepaint.org) of 635 regulated genes in embryos and pancreata at E14.5. This enabled us to identify almost

ACCEPTED MANUSCRIPT

all known regulators of pancreas development and to propose gene regulatory networks (GRNs) for genome-wide association study (GWAS)-annotated diabetes genes. We further analyzed publicly available chromatin-immunoprecipitation sequencing (ChIP-seq) data sets of human islets of Langerhans (Pasquali et al., 2014; Morán et al., 2012) and analyzed the regulation of newly identified genes by pancreatic transcription factors. Importantly, many potential regulators of development and disease were identified in the endodermal and non-endodermal compartment of the pancreas, which provides a high spatio-temporal resolution of the GRNs involved in pancreatic lineage allocation for future functional interrogation.

Results

Foxa2 marks the pancreatic epithelium during secondary transition

To dissect the GRNs in different compartments of the pancreas during secondary transition, we utilized our recently generated knock-in FVF reporter mouse (Burtscher et al., 2013). Immunohistochemistry on pancreatic cryosections combined with laser-scanning confocal microscopy analysis was performed at E12.5-15.5 (Figure 1). We used Pdx1 as a marker for epithelial (Pdx1^{low}) and endocrine progenitors (Pdx1^{high}), as well as Cd49f (Itga6; Integrin alpha 6) as a marker for the exocrine (Cd49f^{high}) and endocrine (Cd49f^{low}) lineage. Note that prior to pancreatic lineage segregation Cd49f is expressed at low levels in the whole epithelium (Sugiyama et al., 2007). The analysis illustrates extensive FVF co-localization with Pdx1 in the ductal, endocrine and exocrine compartment of the pancreas (**Figure 1a**), in line with the notion that *Pdx1* is a downstream target of Foxa1 and Foxa2 (Gao et al., 2008). This idea is further highlighted by the fact that both Pdx1 and FVF proteins are strongly upregulated when progenitor cells commit to the endocrine lineage and leave the ductal epithelium (**Figure 1b**, arrowheads). The FVF negative population (FVF^{-}) represents the surrounding non-endodermal tissues, mainly mesenchymal cells, peripheral neurons, blood, and blood vessels. Together, these results emphasize that the FVF knock-in reporter mouse can be utilized to generate a spatio-temporal transcriptional profile of the distinct pancreatic tissue compartments during the secondary transition.

Global gene expression analysis of the secondary transition

ACCEPTED MANUSCRIPT

To characterize global gene expression changes during pancreatic lineage allocation, we focused our analysis on the secondary transition between E12.5-15.5 (**Figure 2a**). We isolated pancreata from heterozygous FVF mice, generated single cell suspensions using collagenase treatment and used fluorescence-activated cell sorting (FACS) to separate the FVF^{-} and FVF^{+} cell populations from the pancreas at consecutive developmental stages (**Figure S1).** After total RNA isolation, samples were subjected to Affymetrix analysis and only microarrays with high quality data were normalized using robust multi-chip averages and used further for statistical analysis (normalized expression available at **Table S1**, and raw data available at **the** Gene Expression Onmibus, GSE66856). Global description of the dataset was obtained by applying principal component (PC) analysis (**Figure 2b**). The first (spatial) PC explained 68.7% of total variance of the dataset showing that the endoderm-derived FVF^{+} and non-endoderm-derived FVF^{-} cell populations are clearly distinct tissue compartments. The second (temporal) PC explained 13.3% of total variance and correlated with the different developmental stages of pancreas development. Note that the global temporal gene expression changes occur between E13.5-15.5 and that the non-endodermal compartment undergoes more subtle changes (smaller range in the second temporal PC), when compared to the endodermal compartment (**Figure 2b**). Taken together, these results demonstrate dynamic changes in GRNs of the different compartments of the pancreas during lineage allocation.

Spatio-temporal analysis of the pancreatic tissue populations

As a next step we first analysed expression changes between the FVF^{-} and FVF^{+} cell compartments individually within each developmental stage (**Figures 2c and S2a**). Differential expression analysis (FDR adjusted p-value < 0.01) was performed using a linear model, which uses an estimated variance for each gene accounting for low samples sizes of two and three for pairwise comparison (Ritchie et al., 2015). From E12.5-15.5 an increasing number of genes were differentially regulated between the FVF^{-} and FVF^{+} populations (**Figures 2c and S2a, Table S1**). Up to 7346 genes were significantly regulated between FVF^{-} and FVF^{+} populations at E14.5. For the comparison between FVF- and FVF+ populations at E13.5, the lower number of regulated genes was due to overall higher p-values (**Figure S2a**), which can be influenced by the larger variance in the second (temporal) PC of E13.5 FVF- samples

ACCEPTED MANUSCRIPT

when compared to respective E13.5 FVF+ samples (**Figure 2b**, dashed box). Thus, we cannot exclude a comparable number of genes actually being regulated at E13.5.

Functional pathway analysis of respective differentially regulated genes was performed (**Figure 2d**). Pathways that were differentially regulated between the two compartments over time were summarized to understand temporal regulation of cellular and molecular processes associated with pancreas organogenesis and lineage commitment. Type 2 diabetes mellitus, protein digestion and absorption, pancreas secretion and maturity onset diabetes of the young (MODY) are key pathways for the physiology and function of the mature pancreas and are already enriched in the FVF^{+} pancreatic epithelial population during development. The pathway analysis encompassed axon guidance in both populations, which is consistent with previously recognized similarity between neuronal and pancreatic progenitors (Arensbergen et al., 2010; Schwartz et al., 2013) and might implicate a role of these class of molecules in coordinated epithelial plexus remodeling and branching morphogenesis. Differential and dynamic expression of extracellular matrix (ECM) and cell adhesion (CAM, tight junction, adherent junction) molecules indicate morphological remodeling of the epithelium into the different pancreatic lineages and formation of the endocrine lineage by EMT or delamination (Gouzi et al., 2011; Pan and Wright, 2011). From E14.5 onwards, metabolic and amino-acid-related biosynthesis pathways were differentially regulated between both populations, while the populations reflect similar expression regarding adherent and tight junctions. Interestingly, amino acid, sugar and nucleotide metabolism seem differentially regulated in the different compartments in the pancreas while progenitor cells exit proliferation and differentiate into the lineages. Differences in complement and coagulation cascades are also detectable, illustrating that besides the mesenchymal component, the FVF^{-} population also contains vascular endothelial and blood cells. Taken together, our spatio-temporal analysis supports the described differential regulation of key pathways that are involved in development, metabolism and morphogenesis of the pancreas (Pan and Wright, 2011).

Temporal progression in tissue formation in endoderm and non-endoderm compartment of the pancreas

To assess dynamic transcriptional changes within the progenitor and differentiating pancreatic populations in both compartments we analyzed temporal regulation using

ACCEPTED MANUSCRIPT

the above described linear model and testing between embryonic stages in each population. The number of differentially regulated genes (FDR adjusted p-value < 0.01) increased during development, peaking in 2196 and 246 genes in FVF$^+$ and FVF$^-$ population between E14.5 and E15.5, respectively (**Figures 3a, b**). Note that the transcriptional regulation over time is one order of magnitude higher in the FVF$^+$ than the FVF$^-$ compartment, emphasizing greater changes in the GRN of the differentiating pancreatic epithelium. The non-endodermal tissue was transcriptionally less heterogeneous during secondary transition compared to the pancreatic epithelium, suggesting that the mesenchyme, vessels and blood cells undergo less dramatic changes during morphogenesis. However little is known about the dynamic of pancreatic mesenchyme during secondary transition and the identified genes that do undergo regulation might be of interest for future studies.

Hierarchical clustering of temporal gene regulation identified clusters with characteristic molecular profiles. Clusters active in FVF$^+$ epithelial progenitors at E12.5-14.5 contain many Notch pathway components (*Hey1, Notch2, Notch3, Jag1, Hes1*) likely essential for progenitor maintenance and differentiation (**Figure 3c**) (Horn et al., 2012, Jensen et al., 2000, Apelqvist et al., 1999). Interestingly, many of the early active genes are associated with transcriptional regulation (*Hhex, Sox11, Gata6, Tcf7l1, NeuroD1, Onecut3, Tbx6, Tbx19*), miRNA processing (*Drosha*), or chromatin regulation (*Smarca2, Smarcc1, Smarce1, Smarcd1, Hdac1*), maybe important to maintain or differentiate the progenitor population. Moreover, GLI-Kruppel family member *Gli3*, the main repressor for Hedgehog (Hh) signaling, is downregulated at E14.5-15.5 and Hh ligand *Ihh* is upregulated at E15.5, indicating a switch in Hh pathway activity. Genes that are indicative for endocrine (*Ins1, Ins2, Insrr, Pyy, Npy, Iapp, Sst, Pcsk1-3, Gipr, Sytl1, Sytl4, Glp1r, Gip, Ghr, Syt7, Syt8, Syt13, Ghrl, Slc2a2*) and exocrine differentiation (*Amy2a5, Serpina1a, 1b, 1d, 1e, Mist1, Nr5a2*) first appear at E14.5, showing that lineage allocation occurs at this stage in pancreas development. Interestingly, clusters with gene activation at E15.5 in FVF$^+$ differentiating cells contain many members of the Wnt- (*Wnt1, 2, 3a, 5a, 6, 10a, 10b and Fzd9, 10*) and Fgf-signaling (*Fgf6, 8, 15, 20, 23*) as well as genes known to have an important function in mid-hindbrain boundary formation (*Wnt1, Fgf8, Otx2, En1*), suggesting that these pathways are involved in endocrine lineage allocation and trunk-tip patterning.

ACCEPTED MANUSCRIPT

In the FVF$^-$ non-endodermal compartment, generally fewer genes seem to be regulated as compared to the FVF$^+$ endodermal epithelial compartment (**Figure 3b, d**). The miRNA pathway (*Drosha, Ago1* and several miRNAs) and chromatin regulation (*Smarcc2, Smarca4, Med12, Med23*) seems similarly regulated at early stages of pancreas development. At later stages between E14.5-15.5, different ligands of the Wnt-, Tgfβ- and Fgf-signaling cascades are activated (*Wnt4, Fgf7, Gdf10*), suggesting that autocrine and paracrine signaling takes place between the two tissue compartments. Results of temporal progression analysis implicate that during pancreatic progenitor differentiation, morphogenesis and lineage allocation the major developmental regulatory signaling pathways are differentially regulated to remodel the epigenetic and genetic landscape of progenitor cells.

Mechanistic understanding of pancreatic development

To illustrate, how the transcriptional profile of the secondary transition can be interrogated, we posed questions advancing the understanding of further mechanisms governing early pancreatic development. First we were interested if epithelial cells during secondary transition undergo a classical EMT. This is still controversial, as a valid alternative option is the delamination of ductal epithelial progenitors committing to the "pseudo-epithelial" endocrine cells that form the islet of Langerhans (Kesavan et al., 2014). To investigate timing and location of EMT during secondary transition we analyzed expression of the master transcription factors (TFs) *Snai1/2*, *Twist1* and *Zeb1/2* (**Figure 3e**). *Snai1/2* and *Zeb1/2* are highly expressed in the non-endodermal mesenchymal compartment between E12.5 and E15.5. Surprisingly, in the endodermal compartment we observe only a slight and transient up-regulation of *Snai2* and *Zeb2* at E13.5 suggesting that classical EMT is not a main driver of endocrine lineage allocation.

Another intriguing question is when and how the *Insulin* promoter is induced in the endocrine population and are GWAS or MODY gene products involved in the regulation? We first analyzed the expression of annotated T1D and T2D GWAS hits during pancreas development (**Figure S3**). Interestingly, a great majority of potential diabetes disease genes is expressed during pancreas development in the non-endodermal or endodermal compartment with a subset correlating with the onset of

ACCEPTED MANUSCRIPT

insulin expression. The *Insulin1* and *2* genes become highly expressed in the FVF^{+} epithelial compartment at E14.5 (**Figure 3f**). To understand the transcriptional regulation, we interrogate gene expression of *Ins2* promoter regulating TFs. In addition to known and MODY TFs, we studied TF binding sites on the *Ins2* promotor (-342 to +50 bp relative to the transcriptional start site) (TSS) (Whelan et al., 1989). In a homology-based screen, we identified a TF module of the two TF families MAF and AP1 factors as well as RxR heterodimer binding sites. When we analyzed the mRNA expression of the respective genes, we noticed correlated and reciprocal expression patterns to the Insulin genes (**Figure 3f**). Interestingly, the retinoic acid (RA) and steroid nuclear receptor Rxrg that forms heterodimers with the HNF steroid receptors and the major TF of the canonical Wnt signaling pathway Tcf7l2 are reciprocally expressed to the Insulin genes between E12.5-15.5. This suggests that RA and Wnt signaling either maintain progenitor cell populations or suppress Insulin gene expression during pancreas development. Interestingly, the Vitamin D receptor (Vdr) and Rxrg were predicted to directly regulate Ins2 (**Figure 3f**) and were shown to be implicated in insulin-resistance (Sung et al., 2012), suggesting that homo- and heterodimer formation might positively or negatively regulate the expression of the Insulin genes. The high expression of MODY genes (*Pdx1, HNF1β, Hnf4a, and NeuroD1*) during secondary transition suggests that MODY genes have an impact on endocrine lineage formation already at the pancreatic progenitor state. Therefore, GRNs and TF programs that are important for human pancreas development and disease can be analyzed using control and patient-derived iPSCs differentiated into pancreatic progenitors (Weedon et al., 2014).

Gene candidate validation with* in situ *public database

To validate our dataset and analyze the gene expression in the pancreas, we focused on the strongly regulated genes (absolute fold change > 2 between FVF^{-} and FVF^{+}) and filtered for the gene ontology (GO) terms transcriptional factors, signaling pathway, extracellular/located at plasma membrane and cilium-related genes (**Figure 4a**; details in Methods section). We manually curated and classified localization patterns of these 635 genes into epithelial, tip, mesenchyme expression patterns and no expression (**Figure 4b; Tables S2 and S3**) as proxies for predicted function and lineage. For the *in situ* profiles we explored the public gene expression database

ACCEPTED MANUSCRIPT

(www.GenePaint.org), which is a digital atlas and provides *in situ* expression patterns of the mouse at E14.5. Known pancreatic factors were used to determine the specific localization patterns (**Figure 4c**). The epithelial localization pattern was characterized by the pancreatic epithelial markers *Pdx1*, *Foxa2* and endocrine markers *Ngn3*, *NK6 homeobox 1* (*Nkx6.1*) and *Islet 1* (*Isl1*). The tip localization pattern encompasses the typical exocrine progenitors as *nuclear receptor subfamily 5, group A, member 2* (*Nr5a2*), *one cut domain, family member 1* (*Onecut1/Hnf6*), *basic helix-loop-helix family, member a15* (*Mist1/Bhlha15*), *chymotrypsin-like elastase family member 1* (*Cela1*) and the *serpine peptidase inhibitor, clade A, member 10* (*Serpina10*). The mesenchymal localization pattern contained genes, such as *odd-skipped related 1* (*Osr1*), *secreted frizzled-related protein 1* (*Sfrp1*), *fibroblast growth factor 9* (*Fgf9*), *wingless-type MMTV integration site family, member 5A* (*Wnt5a*) and *platelet derived growth factor C* (*Pdgfc*). Several of the tissue-characteristic genes were also significantly regulated in time and space **(Figure 4d).** Taken together, these data illustrate the high quality of the data set in terms of FVF$^-$ and FVF$^+$ compartment separation and spatio-temporal regulation of GRNs.

Genes from the GO filtering were analyzed separately with respect to the manually curated *in situ* localization pattern and differential regulation in the endodermal and non-endodermal compartment. The *in situ* profile revealed unique distribution in the localization pattern in the epithelium at E14.5 (**Figure 5a**). Irrespective of temporal regulation in FVF$^+$ and FVF$^-$, all epithelium genes were upregulated in FVF$^+$ when compared to FVF$^-$ compartment at all timepoints (**Figure S4a**). Expressions of genes coding for secreted hormones, such as *islet amyloid polypeptide* (*Iapp*) and *insulin 2* (*Ins2*) were induced during secondary transition in the FVF$^+$ endodermal compartment and have a specific epithelial *in situ* pattern in the pancreas at E14.5. The RNA expression profile shows relatively stable expression of TF in the secondary transition (*Pax6, Nkx2-2, Isl1*), whereas genes coding for secreted factors or hormones in differentiated endocrine cells are strongly regulated (*Scgn, Ins2, Iapp, Ghrl*). Most importantly these factors are strongly upregulated during endocrine differentiation and remain expressed in adult islets of Langerhans (Pan and Wright, 2011).

ACCEPTED MANUSCRIPT

The *in situ* pattern for the "tip" highlighted consistently peptidases, lipases and amylases next to exocrine progenitor markers (**Figure 5b**). Again tip genes were all found to be consistently upregulated in FVF$^+$ when compared to FVF$^-$ compartment at all timepoints (**Figure S4b**). The *Carboxypeptidase 1* (*Cpa1*) has very specific *in situ* pattern in acinar cells at the ceiling of the pancreas and shows steady expression during secondary transition. *Zhou et al* already showed that *Cpa1* mRNA is detected at E9.5 with protein expression at E10.5 marking the multipotent progenitors (MPC). Similarity in the *in situ* expression pattern is obvious for the protease *Chymotrypsin-like elastase family member 1* (*Cela1*), the *Amylase 2a5* (*Amy2a5*), and the *Ectodysplasin-A receptor-associated adapter protein* (*Edaradd*) in the acinar compartment of the pancreas at E14.5. All of these genes show increased expression intensity towards the end of secondary transition. Interestingly, the Serpina family (1a, 1b, 1d, and 1e) is a syn-expression group that clusters together on chromosome 12 and is collectively downregulated towards the end of secondary transition, suggesting that these negative regulators of endopeptidases have a role in pancreas differentiation or morphogenesis. Indeed, null-mutants for the alpha1-antitrypsin (A1AT) Serpina1a die prior to E8.5, indicating a critical role in regulation of *in utero* development (Wang et al., 2011).

The non-endodermal *in situ* pattern reflected a specific pattern in the mesenchyme around the pancreas epithelium (**Figure 5c**) and genes were all found to be consistently downregulated in FVF$^+$ when compared to FVF$^-$ compartment at all timepoints (**Figure S4c**). As a gene with strong temporal expression we found *thrombospondin 2* (*Thbs2*) with an *in situ* expression in the surrounding mesenchyme of the pancreas. Thbs2 is a homotrimeric glycoprotein that mediates cell-to-cell and cell-to-matrix interactions and homozygous knock-out mutants die due to marked structural and functional abnormalities in a variety of connective tissues (Kyriakides et al., 1998). Additionally, *Lumican* (*Lum*) reflects a similar dynamic expression pattern in the non-endodermal compartment and knock-out mice have disorganized extracellular matrix (Chakravarti et al., 1998). These data suggests that genes strongly regulated in the non-endoderm compartment are essential for ECM modulation and cell-cell and cell-matrix interactions. Interestingly *insulin-like growth factor binding protein 4* (*Igfbp4*) and *Igfbp5* show stable expression in the non-endodermal compartment verified by the *in situ* pattern. On the other hand, insulin-like growth

ACCEPTED MANUSCRIPT

factor 1 (Igf1) is strongly upregulated at E14.5-15.5 (**Figure 3c**), suggesting activation of the Igf1 pathway to enhance growth and proliferation during the end of secondary transition.

Novel candidates shaping transcriptional landscape of secondary transition

As a next step we examined factors, which were previously not annotated as functional genes in pancreas development or disease. The criteria for these factors comprised strong regulation during secondary transition (absolute fold change > 2 between FVF^- and FVF^+) as well as a regional strong and specific pancreatic *in situ* expression pattern at E14.5 (**Figure 6; Tables S2 and S3**).

For the "epithelial" *in situ* pattern we identified genes with unique expression in the pancreas (**Figure 6a**) and validated expression with independent *in situ* stainings as well as qPCR (**Figure 7**). The novel pancreatic factor *growth factor receptor bound protein 7* (*Grb7*) had steady expression during pancreas development with typical epithelial *in situ* pattern (**Figure 7**). Further, we observed decreasing mRNA expression and a weak *in situ* pattern for the novel *transmembrane protein 171* (*Tmem171*) during the secondary transition. As a new putative pancreatic gene *von Willebrand factor A domain containing 5B2* (*Vwa5b2*) was identified to be upregulated in FVF^+ with *in situ* expression in the epithelium, only (**Figure 7**). As a potential endocrine factor we discovered *Synaptotagmin 13* (*Syt13*) with a similar *in situ* pattern compared to *Ngn3* (**Figure 4c, Figure S5**) and increasing expression during the ongoing secondary transition (**Figure 7**). Interestingly, Syt13 was shown to be absent in Ngn3 knockout mice (Juhl et al., 2008). Syt13 is a member of the Synaptotagmin family, known to play a role in vesicle-mediated transport and vesicle exocytosis (Xu et al., 2009).

For novel pancreatic genes with the *in situ* expression profile processed as "tip" the RNA expression of *solute carrier family 38* (*Slc38a3*) and *RNA binding motif protein 47* (*Rbm47*) declines in the epithelium during secondary transition as illustrated by the weak pancreatic *in situ* expression patterns (**Figure 6b**). Nevertheless, the putative pancreatic gene *Claudin-10* (*Cldn10*) shows increasing expression during pancreas development with unique *in situ* expression in the acinar compartment of the

ACCEPTED MANUSCRIPT

pancreas. As a novel pancreatic gene we found the gene *2210010C04Rik* with rising expression during secondary transition and clear *in situ* pattern at the ceiling of the pancreas, a gene which likely codes for a peptidase.

The "mesenchymal" localization contains an additional *Synaptotagmin* (*Syt6*), *Tensin-like C1 domain-containing phosphatase* (*Tenc1*) and *Sushi repeat-containing protein X-linked 2* (*Srpx2;* **Figure 7**) and *Interleukin 11 receptor alpha chain 2* (*Il11ra2*) (**Figure 6c**). All the non-endodermal genes show constant expression over time and clear *in situ* expression pattern in close proximity to the pancreas epithelium. Interestingly, Tenc1 deficient mice show abnormal assembly and maturation of the basement membrane in the kidney, likely due to a specific role of Tenc1 in the integrin pathway (Uchio -Yamada K. et al., 2013). It might be interesting in the future to analyze the function of the vesicle fusion regulators of the Synaptotagmin family or any other of these novel and interesting genes during pancreas development.

Transcription factor binding and epigenetic signatures in human islets at known and novel loci

Finally, we were seeking evidence for the regulation and function of known and newly identified developmental genes in the human islets of Langerhans as a first step of translation to human biology and disease. Therefore, we used the ChIP-seq data from adult islets (Pasquali et al. 2014, Morán et al. 2012) and compared histone modifications and transcription factor occupancy at the human homologues of our identified genes (**Figure 8**). The *PDX1* locus, serving as example for a typical pancreatic gene, is marked by the histone modifications H3K4me3, H3K4me1 and H3K27ac, which are associated with active promoters, active enhancers and chromatin accessibility, respectively. In addition, the promoter is bound by several transcription factors (FOXA2, NKX2.2, NKX6.1, MAFB, and PDX1), known to be important to establish and maintain the neuroendocrine cell fate, that co-occupy several sites (**Figure 8a**). Interestingly, the novel pancreatic genes *SYT13, SCGN* and *LRP11* show a similar pattern of active chromatin marks and their gene bodies were bound by a similar set of transcription factors (FOXA2, NKX2.2, NKX6.1 and PDX1) that co-occupy multiple sites in the promoter and at intronic enhancers (**Figure 8b-d**). In contrast, the homologues loci of genes, expressed in the acinar or mesenchymal compartment, such as *AMY2B* (**Figure S6a**) and *SFRP1* (**Figure S6b**),

ACCEPTED MANUSCRIPT

show no signs of activity at the chromatin level and are not bound by any of the evaluated pancreatic transcription factors in islets of Langerhans as expected. Thus, the ChIP-seq data from adult human islets highlights the quality of our data set and the relevance of the identified novel pancreatic genes also for human islet biology.

Discussion

In this study we have separated the endodermal from the non-endodermal compartment of the pancreas during secondary transition when multipotent progenitor cells commit to the endocrine, exocrine and ductal lineages. The overall goal was to capture genome-wide expression changes in multipotent progenitor cells and the neighboring tissue to better annotate the cellular and molecular processes associated with pancreas organogenesis. Neighboring tissue interactions are important to orchestrate differentiation and morphogenesis of the pancreas and our extensive statistical analyses clearly identified distinct spatio-temporal regulation of GRNs in both tissue compartments. Our results indicate that the overall expression landscape is less heterogeneous in the non-endoderm compartment. Although many important chromatin regulators, transcription factors, and signaling molecules are differentially expressed in this compartment, the function of most of these factors is not yet functionally annotated in pancreas development. One reason is certainly the lack of conditional gene targeting information in the non-endodermal compartment due to the lack of pancreas mesenchyme specific Cre-driver lines. The enrichment analysis of molecular processes and the transcriptional profile in the epithelium, however, clearly identified known factors involved in pancreas development and disease, but also provided additional information on spatio-temporal gene regulation and discovered novel genes that are differentially regulated and regionally expressed. Thus, our transcriptional profile of the different pancreas compartments during secondary transition is a rich resource to explore molecular processes, when differentiation and morphogenesis is actively taking place. In the further discussion we highlight several interesting details we observed during our exploration of the comprehensive data set.

Foxa2 marks multipotent pancreatic progenitors

Foxa2 is a pioneer factor that is expressed and functionally important for early endoderm and pancreas development (Burtscher and Lickert, 2009; Gao et al., 2008).

ACCEPTED MANUSCRIPT

Our Foxa2 and Pdx1 co-localization study revealed that both factors are co-expressed to a large extent in the entire epithelium at the beginning of the secondary transition and get upregulated when progenitor cells leave the ductal epithelium to commit to the endocrine lineage. The uniform and ubiquitous expression of several key transcription factors and MODY genes that regulate cell lineage determination in the FVF$^+$ population, suggests that Foxa2 marks multipotent epithelial progenitors while they commit to ductal, exocrine and endocrine fate (Pan and Wright, 2011). This is in line with conditional deletion of Foxa1 and Foxa2 in the Pdx1$^+$ pancreas epithelium, which leads to complete loss of *Pdx1* expression and pancreatic hyperplasia, emphasizing the role of Foxa2 as a pioneer factor to activate a pancreatic gene regulatory program (Lee et al., 2002).

Tissue remodeling and patterning during secondary transition

The high-resolution gene expression profile captured characteristics of autonomous and non-autonomous determination of the pancreatic epithelium and the surrounding non-endodermal compartment. Almost all known critical regulators of pancreas development were identified in the mRNA gene expression profiling and were verified by our Genepaint *in situ* analysis to be expressed in the pancreatic epithelium at E14.5. However, the distinct tissue compartments clearly showed different temporal gene expression dynamics associated with distinct molecular functions. The pathway analysis of both tissue populations expectedly identified extracellular matrix-receptor interactions correlated with regulation of cell adhesion molecules, focal adhesion, and actin cytoskeleton regulation, indicating that during secondary transition large tissue remodeling is taking place including branching morphogenesis and tip-trunk patterning. Interestingly, different microRNAs are downregulated in the non-endodermal population at E15.5. With increasing evidence that microRNAs participate in pancreatic development by regulation of pluripotency, differentiation and morphogenesis (Hinton et al., 2012; Wei et al., 2013), it is important to further explore microRNA function in both, the endodermal and non-endodermal compartment.

Furthermore, pathway analyses identified a link between pancreas development and axon guidance. The striking similarities between neuronal and neuroendocrine lineage formation is exemplified by the large overlap of shared transcription factor networks

ACCEPTED MANUSCRIPT

that regulate differentiation. However, axon guidance molecules have not been extensively studied during pancreas development, but likely play a role during branching morphogenesis or islet neogenesis, where coordinated cell movements occur. Interestingly, we also identified several genes known to have an important function in mid-hindbrain boundary formation activated sequentially in FVF$^+$ differentiating cells at E14.5-15.5. Most prominently, the *En1*, *Otx2*, *Six3*, *Pax6* transcription factors known to be important for mid-hindbrain patterning, as well as the Fgf, Shh and Wnt signaling components get strongly upregulated at E14.5-15.5 (Wurst and Bally-Cuif, 2001). These results suggest that during tip-trunk patterning of the pancreatic epithelium conserved molecular machineries generate tissue boundaries.

Our gene expression resource captured a short wave of EMT (epithelial-mesenchymal-transition) transcription factor expression around E12.5-13.5, which could be accountable for endocrine lineage formation. If ductal progenitors give rise to endocrine cells by EMT or delamination is still controversial, however, accumulating evidence suggests the later scenario (Kesavan et al., 2014). Our results show that EMT TFs are far higher expressed in the surrounding non-endodermal compartment and are only transiently activated at low levels in the pancreas epithelium. Since we analyze a large number of cells per sample, the low signal of EMT TF expression could be explained by only a fraction of endocrine progenitors undergoing EMT, a hypothesis, which is supported by the transient expression of Ngn3 in a minor subpopulation of the ductal epithelial cells (Gradwohl et al., 2000; Schwitzgebel et al., 2000). We speculate that EMT TF expression is only necessary for endocrine cells to delaminate from the ductal epithelium, but not to undergo a classical epithelial-mesenchymal transition, where cells transform to a mesenchymal fate. Instead, ductal progenitors very likely stay in a pseudo-epithelial state when they leave the epithelium and directly contribute to the islets of Langerhans (Pictet and Rutter, 1972).

It has long been proposed that epithelial-mesenchymal cross talk is important for epithelial proliferation, differentiation and morphogenesis (Golosow and Grobstein, 1962). Specifically, if the mesenchyme is stripped away from early pancreatic epithelial buds at E11.5, the epithelium stops proliferating and by “default”

ACCEPTED MANUSCRIPT

differentiates into the endocrine lineage (Gittes et al., 1996). These results suggest that mitogenic factors as well as signals that maintain multipotent progenitors are provided from the mesenchyme. Our analysis captured gene expression changes in both compartments during secondary transition. Interestingly, we find TGFβ-, Nodal/Activin-, Bmp-, Fgf-, Wnt-, and Notch-signaling mainly regulated in the pancreas epithelium, suggesting that autocrine self-regulatory epithelial interactions are the main driver for pancreatic epithelial patterning and differentiation. This idea is further supported by the fact that pancreatic buds without the mesenchyme placed under the kidney capsule formed clusters of mature islets (Gittes et al., 1996). However, in the non-endoderm compartment we have also identified *Fgf7*, *Wnt4*, *Gdf10*, and *Pgf* being significantly regulated, offering a few signaling pathways that act in a paracrine fashion. Indeed Fgf1, 7, and 10 were suggested to signal via Fgf receptor 2b to stimulate epithelial growth, differentiation and morphogenesis (Miralles et al., 1999). Importantly, a multitude of transcription factors and signaling molecules identified in both compartments are certainly important for pancreas development, but so far not functionally annotated. Further mining and characterization of the pancreatic transcriptional profile will help to precisely explore repetitively used signal cascades, such as Notch signaling that is essential for progenitor maintenance (Apelquist et al., 1999; Fujikura et al., 2006; Cras-Méneur et al., 2009) and endocrine differentiation (Horn et al., 2012). In the future, this will allow to better device *in vitro* differentiation protocols for the generation of β-cells from pluripotent stem cells.

Understanding pancreas development to unravel pathomechanisms of diabetes

Recent genome-wide association studies have identified a multitude of genes associated with type 1 (T1D) and type 2 (T2D) diabetes (Marullo et al., 2014; Evangelou et al., 2014). For a long time it is known that monogenic forms of diabetes, such as maturity onset diabetes of the young (MODY) and neonatal diabetes are caused by mutations in developmental regulatory genes that control pancreas development (Schwitzgebel, 2014). Defects can either cause β-cell dysfunction when the genes regulate cell homeostasis, glucose sensing or insulin secretion, or cause a reduction in β-cell number when the genes act early during β-cell development. The pathogenesis of T1D is less well understood, but might also be caused by defects in β-

ACCEPTED MANUSCRIPT

cell development or function. We have compared our gene expression profile to T1D and T2D GWAS hits and found that the great majority of genes are either expressed in the non-endodermal or endodermal compartment of the pancreas. Many of the T2D GWAS genes important for glucose sensing, insulin secretion and metabolic regulation of the β-cell, such as *Pcsk1*, *Slc2a2*, *Slc30a8*, *Gck*, *Gipr*, *G6pc2*, etc. are expressed during late stage differentiation of β-cells and the expression correlates with the sharp onset of *Ins1* and *Ins2* mRNA expression at E14.5. The expression of many MODY genes, such as *PDX1*, *HNF1b*, *HNF4a*, *HNF4b*, *NEUROD1*, *PAX4* etc. are exclusively expressed in the pancreas epithelium throughout secondary transition, suggesting that they are actively involved in β-cell differentiation and mutations lead to congential defects in β-cell development and islet neogenesis. Interestingly, the mouse homologues gene for one of the major T2D GWAS hits, *Tcf7l2*, shows reciprocal expression to *Ins1* and *Ins2* mRNA gene expression. During pancreas development the function of Tcf7l2 is still controversial. Whereas the whole body knock-out suggests that Tcf7l2 is not necessary for β-cell development (Boj et al., 2012), conditional deletion of Tcf7l2 in Pdx1 progenitors or Ins1 positive cells lead to a decreased β-cell mass and function (da Silva Xavier et al., 2012; Mitchell et al., 2014). The reciprocal expression of Tcf7l2 to *Ins1* and *Ins2* mRNA might either suggest that Tcf7l2 suppresses Insulin gene activation or that Wnt/β-catenin signaling maintains multipotent progenitors proliferation. The latter hypothesis is in line with results obtained from the adult pancreas where Wnt/β-catenin signaling via the downstream transcription factor Tcf7l2 is suggested to regulate Glp1-mediated β-cell proliferation and SDF1-mediated β-cell survival (Liu and Habener, 2008). However, TCF7L2 was also suggested to be a master regulator of insulin production and processing (Zhou et al., 2014). The reciprocal expression of several of the predicted TCF7L2-regulated transcriptional target genes (*ISL1*, *PCSK1*, *SLC30A8*) in our data set, suggests differences in mouse and human pancreas development or distinct target gene sets in development and adult β-cells. Taken together, careful analysis of GWAS genes and their associated gene regulatory networks may shed light on the pathomechanisms of diabetes.

Conclusion

ACCEPTED MANUSCRIPT

Our comprehensive study of the global gene expression profile of secondary transition during pancreas development dissected developmental mechanisms and subsequently suggested novel pancreatic factors. Further mining of this valuable resource might reveal mechanisms of pancreas development, neighboring tissue interactions and pathomechanisms of diabetes.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Figure Legends

Figure 1. FVF marks the pancreatic epithelium during pancreas development

a. Characterisation of FVF pancreatic progenitors during secondary transition (E12.5-15.5) with pancreatic epithelial marker Pdx1 and endocrine (low) /exocrine (high) marker Cd49f showing an overlay of FVF and Pdx1 in the pancreatic epithelium. Magnification of boxed area shown in b. Scale bars 25µm.

b. Higher magnification of E15.5 with FVF, Pdx1 and Cd49f. Notably, $Pdx1^{high}$ expression correlates with FVF^{high} expression (arrowheads). Scale bar 25µm.

Figure 2. Workflow of the experimental setup and temporal differences of the two tissue populations

a. Scheme of the experimental setup: The FVF mice were used for mating and the plug positive mice utilized to prepare the embryonic pancreas during secondary transition. Successively we accomplished a single cell suspension comprised of FVF^+ and FVF^- cells representing the epithelium and the surrounding non-endodermal compartment. After FACs sorting and total mRNA isolation the global gene expression array illustrated 2 distinct cell populations.

b. Principal component analysis of genome-wide profiling. The first component (PC1) clearly distinguishes between FVF^+ and FVF^- samples. PC2 aligns samples according to developmental stages. Arrows are drawn in increasing number of embryonic days from respective sample means in PC1 and PC2.

c. and d. Differential gene expression analysis of both populations per development day. Histogram of genes significantly regulated (adjusted p-value < 0.01). *Number at E13.5 is lower due to lower samples sizes (n=2 each). The lower panel heatmap shows pathway enrichments of spatial tissue differences for the respective timepoints. KEGG pathway enrichment significance is color-coded (red values indicate high probability ($\log10(p)<1$ with $p<0.1$). Pathways are hierarchically clustered by their temporal pathway enrichment pattern.

Figure 3. Temporal expression changes

Significant expression changes during the specific embryonic stages within the epithelium (**a.**) and non-endodermal tissue (**b.**). Histograms (upper panels) indicate number of differentially expressed genes of one developmental day to the next. Note the different y-axis scaling up 200 or 2000 genes. Heatmaps (**c. and d.**) of temporally

ACCEPTED MANUSCRIPT

regulated genes within the FVF^+ and FVF^- tissue populations are hierarchically clustered by their relative gene expression (z-scales to mean expression per row) across samples. Histogram on colour-scale summarizes number of expression values shown in the heatmap.

e. Heatmap visualizes relative expression values of EMT transcription factors (TF).

f. Heatmap visualizes relative expression values of known, MODY and predicted TF genes. TF binding sites of predicted and conserved TFs in the *Ins2* promoter are shown in the lower box.

Figure 4. GO term and Genepaint *in situ* analysis of known pancreatic factors

a. Representative GO terms used for the identification of known and novel pancreas-related genes with >2 fold expression difference between both tissue populations. The terms include extracellular/ located at plasma membrane (291 genes), signaling (138 genes), transcription factors (141 genes), cilium-related (34 genes) and unknown (31 genes). In total, we found 635 non-redundant genes. The heatmaps reflect the percentage of genes with GO terms either in the non-endodermal, epithelial population and the contrast of these two cell populations.

b. These genes had been further analysed by Genepaint *in situs* and classified manually to their different localization pattern. The patterns demonstrate pancreatic epithelium, tip and pancreatic mesenchyme.

c. Known pancreatic genes sorted for their typical Genepaint *in situ* pattern. For the pattern pancreatic epithelium representative genes include *Foxa2*, *Pdx1*, *Ngn3*, *Nkx6.1* and *Isl1*. In addition the tip fraction is reflected by the genes *Nr5a2*, *Onecut1/Hnf6*, *Mist1*, *Cela1* and *Serpina10*. For the mesenchymal pattern we found *Osr1*, *Sfrp1*, *Fgf9*, *Wnt5a* and *Pdgfc*.

d. Heatmap of known pancreatic factors from c. Genes are hierarchically clustered by relative gene expression (z-scales to mean expression per row) across samples.

Figure 5. Temporal and spatial analysis of known pancreatic genes including their Genepaint *in situs.*

a-c. The upper panels describe temporal expression changes of all classified genes in the indicated compartment (average fold-change relative to E12.5). Interesting genes are marked and stated by their names.

ACCEPTED MANUSCRIPT

a. *In situ* profiling of established pancreatic factors presented for the pancreatic epithelium pattern by well described pancreatic marker as *Ins2*, *Nkx2.2*, *Pax6* and *Iapp*.

b. For the typically tip pattern we depicted intrinsic exocrine progenitors as *Cpa1*, *Cela1*, *Amy2a5* and *Edaradd*.

c. In addition, a mesenchymal pattern is illustrated by the genes *Thbs2*, *Lum*, *Igfbp4* and *Igfbp5*.

Figure 6. Temporal and spatial analysis of novel pancreatic genes including their Genepaint *in situs.*

a-c. The upper panels describe temporal expression changes of all classified genes in the indicated compartment (average fold-change relative to E12.5). Interesting genes are marked and stated by their names.

The genes described are manually depicted genes previously unknown to play a role in pancreatic lineage allocations and are either in the GO terms or classified as unknown.

With the advantage of the *in situ* database Genepaint, unidentified pancreatic related genes are arranged into the different patterns due to their unique *in situ* expression pattern.

a. For the epithelium, expression changes during secondary transition with unique epithelial *in situ* pattern illustrates the unidentified pancreatic related genes *Grb7*, *Tmem171*, *Vwa5b2* and *Syt13*.

b. In the section of the tip there are typical expression patterns for the novel genes *Slc38a3*, *Rbm47*, *Cldn10* and *2210010C04Rik*.

c. Furthermore the mesenchymal population discloses pancreas specific unfamiliar candidates such as *Syt6*, *Tenc1*, *Srpx2* and *Il11ra2*.

Figure 7. Confirmation of selected candidate gene in pancreatic tissue by whole mount *in situ* hybridization and quantitative PCR

a) Whole mount *in situ* hybridization of the candidate genes Syt13, Scgn, Grb7, Vwa5b2 and Srpx2 at E12.5 (n=3), E13.5 (n=2), E14.5 (n=3) and E15.0 (n= 2). Scale bar = 250µm. b) qPCR of selected candidate genes between E12.5 to E14.5. Y-axis

ACCEPTED MANUSCRIPT

shows the fold change values (mean ± SEM) relative to E12.5. Two-tails unpaired t-test, * significant at $p<0.05$, *** significant at $p<0.001$, ns = not significant.

Figure 8. Transcription factor binding and distribution of histone modifications at loci of known and novel pancreatic factors in adult human islets.

Histone modification and transcription factor binding patterns in the vicinity of the human homologues of known and novel pancreatic factors from publicly available ChIP-seq data.

a-d. In adult human islets *PDX1* is bound by several endocrine specific transcription factors (FOXA2, PDX1, NKX2.2, NKX6.1 and MAFB) and shows an active chromatin configuration, which is indicated by the histone modifications H3K4me3, H3K4me1 and H3K27ac. In line with *PDX1*, the novel pancreatic factors *SYT13, SCGN* and *LRP11* were bound by FOXA2, PDX1, NKX2.2 and NKX6.1 and marked by H3K4me3, H3K4me1 and H3K27ac. Additionally, RNA-seq data shows that *PDX1* and *SYT13* are well expressed in adult human islets. RNA-seq data shows very low expression levels, which are presumably caused by a slight impurity of islet preparations (HI32: 93,8% islet purity).

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Methods

Immunhistochemistry

The pancreas of the different embryonic stages E12.5 until E15.5 had been dissected, fixed for 20min in 4% PFA and after sucrose gradient sectioned into 10µm slices.

The tissue had been permeabilized for 20min (0.1M Glycine, 0.2% TritonX-100) and incubated overnight in blocking solution containing 0.1% Tween-20, 10% FCS, 0.1% BSA and 3% donkey serum in PBS (phosphate buffered saline). The first antibody had been diluted in the blocking solution for 6h at RT, and after 2 washing steps with PBST (0.7% Tween-20) the second antibody had been incubated. After washing with PBST for 5min twice, the tissue sections were mounted with ProLong Gold Antifade with DAPI (Invitrogen).

Antibodies used: Chicken anti-GFP (Aves Lab # GFP-1020 – 1:1000), Rabbit anti-Pdx1 (NEB # 5679 – 1:300) and Rat anti-Cd49f (BD # 555736 – 1:200)

Expression profiling

Total RNA was isolated employing the Rneasy Mini kit (Qiagen) including digestion of remaining gDNA. The Agilent 2100 Bioanalyzer was used to assess RNA quality and only high quality RNA (RIN>7) was used for microarray analysis.

Total RNA (100 ng) was amplified using the WT Expression kit (Ambion) and the WT Terminal labeling and Fragmentation Kit (Affymetrix). 2 µg of amplified cDNA were hybridized on Affymetrix Mouse Gene ST 1.0 arrays containing about 29,000 probe sets. Staining (Fluidics script FS450_0007) and scanning was done according to the Affymetrix expression protocol. In total, 21 samples were analyzed at 4 time points (E12.5, E13.5, E14.5, E15.5) in two different populations (FVF^- and FVF). Due to low quality we removed two samples from further analysis. Raw array data were submitted to Gene Expression Omnibus (GSE66856) and a link for referees has been created:

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=gnubwesiphanjiv&acc=GSE66856

Statistical transcriptome analysis

ACCEPTED MANUSCRIPT

Expression console (Affymetrix) was used for quality control and to obtain annotated normalized RMA gene-level data (standard settings including sketch-quantile normalisation). Principle component analysis was performed on the global transcriptional profiles. The first two principle components explained most of the variance in the data set. To analyze significant gene regulation, we applied a linear regression model with the Limma package (Ritchie et al., 2015). Probeset expression values y_i were modelled to be explained by embryonic days *ed* (E12.5, E13.5, E14.5, E15.5) and tissue population *t* (FVF$^+$, FVF$^-$). Contrasts of regression coefficients were used to determine differential expression. For example temporal regulation of FVF+ between E12.5 and E13.5 was performed using samples (*ed*=E12.5 & *t*=FVF$^+$) versus (*ed*=E13.5 & *t*=FVF$^+$). The significance threshold of Benjamini-Hochberg adjusted p-values < 0.01 was implemented. All analyses were performed in R Statistical software (www.r-project.org).

Functional analysis

Pathway enrichment analysis was performed with KEGG (excluding human disease pathways) and hypergtest packages from R/Bioconductor. Pathways were considered enriched with BH-adjusted p-value < 0.01.

Insulin binding site analysis

To predict novel transcription factors (TF) binding insulin we analysed the Ins2 promoter -342 to +8bp around the transcription start site. We used Genomatix software to search for common TF binding sites (TFBS) using conserved TFBS across vertebrates (Cartharius et al., 2005). Three TF families were common across species: RXR heterodimer binding sites, TATA binding factors and AP1/MAF related factors. For Tata factors, no respective gene was measured on the chip.

Candidate Genes Selection

First, Gene ontology (GO) term enrichment for transcription factors, cilium related, extracellular/located at plasma membrane and signaling molecules arranged the genes into their cluster. Afterwards Genepaint *in situs* of whole mount embryo and Zoom into the region of the pancreas had been performed. Selection of known pancreatic factors disclosed the advantage of typical *in situ* patterns for the different pancreas

ACCEPTED MANUSCRIPT

specific lineages. Using this outlook, novel pancreatic genes had been arranged to their *in situ* pattern and therefore to their predicted lineage in the pancreas.

Processing and visualization of ChIP-seq and RNA-seq data

Sequenced reads from ChIP-seq and RNA-seq experiments were downloaded from ArrayExpress (E-MTAB-1919 and E-MTAB-1294). For a complete list of all processed files see **Table S4**. Reads from ChIP-seq experiments were mapped to the human genome (hg19) using bowtie (v1.0.0) with default mapping settings and SAM format output. Replicate reads from different lanes were combined prior to the alignment. SAM files were then converted to BAM files using samtools (v0.1.18). Reads from the RNA-seq experiment were mapped using the Subread aligner (Liao et al., 2013) in paired-end mode and default setting and the -u , –H and BAM output options. The hg19 index was generated using default settings and the –F option. For all BAM-files, duplicate reads were removed and BAM-files were sorted and indexed using samtools (v0.1.18). Processed BAM-files were then visualized using the R package Gviz.

Whole mount in situ hybridization

ISH probes were generated from FANTOM and I.M.A.G.E full length cDNA clones. For hybridizations organs were processed as described (Lickert et al., 2001) with the following modifications. After rehydration organs were bleached with 3% H_2O_2 in PBT for 20 min in the dark, washed in PBT and incubated with proteinase K (10µg/ml in PBT) for 3 min.

qPCR

For each stage total mRNA from five to ten pancreata (three biological replicates) was isolated using miRNeasy micro kit (Qiagen). cDNA reverse transcription was carried out using “SuperScript® VILO™ cDNA Synthesis Kit” (Invitrogen). For each qPCR reaction 20 ng of cDNA was used. TaqMan® qPCR was assessed by using the following probes according with the manufacturing instruction. Taqman probes: Grb7 (Mm01306734_m1), Vwa5b2 (Mm00725391_m1), Srpx2 (Mm01354530_m1), Syt13 (Mm00600526_m1), Scgn (Mm00520475_m1), Gapdh (Mm99999915_g1). For each biological replicate three technical replicates were analyzed. Relative expression

ACCEPTED MANUSCRIPT

levels were normalized to Gapdh and fold change values were calculated relative to the mean expression level of E12.5.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Author Contributions

SJW, IB, AR, SE performed mouse work. MI, JB performed transcriptome profiling. SJW analysed *in situ* patterns. NSM, SS performed statistical and bioinformatical analyses. MS analyzed ChIP-seq data sets. HL and FJT design the study. SJW and NSM drafted the manuscript. HL wrote the manuscript.

Acknowledgements

This work was funded by Helmholtz Alliance ICEMED and Helmholtz Portfolio Theme 'Metabolic Dysfunction and Common Disease' (grants to HL and JB), the European Research Council (Latent Causes: 259294; grant to FJT). Additionally, the HumEn project has received funding from the European Union's Seventh Framework Programme for Research, Technological Development and Demonstration under grant agreement No. 602587 (http://www.hum-en.eu/). H.L. thanks the Helmholtz Society, German Research Foundation and German Center for Diabetes Research (DZD e.V.) for financial support and the Federal Ministry of Education and Research [Virtual Liver: FKZ 0315752 and LungSys: FKZ 0316042I]. We thank Bianca Vogel for technical support.

References

Apelqvist A., Li H., Sommer L., Beatus P., Anderson DJ., Honjo T., Hrabe de Angelis M., Lendahl U., Edlund H., 1999. Notch signalling controls pancreatic cell differentiation. Nature. 1999 Aug 26;400(6747):877-81.

Boj SF., van Es JH., Huch M., Li VS., José A., Hatzis P., Mokry M., Haegebarth A., van den Born M., Chambon P., Voshol P., Dor Y., Cuppen E., Fillat C., Clevers H., 2012. Diabetes risk gene and Wnt effector Tcf7l2/TCF4 controls hepatic response to perinatal and adult metabolic demand. Cell. 2012 Dec 21;151(7):1595-607.

Bonner-Weir S., Weir GC., 2005. New sources of pancreatic beta-cells. Nat Biotechnol. 2005 Jul;23(7):857-61.

Burtscher I., Barkey W., Lickert H., 2013. Foxa2-venus fusion reporter mouse line allows live-cell analysis of endoderm-derived organ formation. Genesis. 2013 Aug;51(8):596-604.

ACCEPTED MANUSCRIPT

Burtscher I., Lickert H., 2009. Foxa2 regulates polarity and epithelialization in the endoderm germ layer of the mouse embryo. Development. 2009 Mar;136(6)):1029-38.

Cartharius K., Frech K., Grote K., Klocke B., Haltmeier M., Klingenhoff A., Frisch M., Bayerlein M., Werner T., 2005. MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinformatics. 2005 Jul 1;21(13):2933-42.

Chakravarti S., Magnuson T., Lass JH., Jepsen KJ., LaMantia C., Carroll H., 1998. Lumican regulates collagen fibril assembly: skin fragility and corneal opacity in the absence of lumican. J Cell Biol. 1998 Jun 1;141(5):1277-86.

Cras-Méneur C., Li L., Kopan R., Permutt MA., 2009. Presenilins, Notch dose control the fate of pancreatic endocrine progenitors during a narrow developmental window. Genes Dev. 2009 Sep 1;23(17):2088-101.

da Silva Xavier G., Mondragon A., Sun G., Chen L., McGinty JA., French PM., Rutter GA., 2012. Abnormal glucose tolerance and insulin secretion in pancreas-specific Tcf7l2-null mice. Diabetologia. 2012 Oct;55(10):2667-76.

Evangelou M., Smyth DJ., Fortune MD., Burren OS., Walker NM., Guo H., Onengut-Gumuscu S., Chen WM., Concannon P., Rich SS., Todd JA., Wallace C., 2014. A method for gene-based pathway analysis using genomewide association study summary statistics reveals nine new type 1 diabetes associations. Genet Epidemiol. 2014 Dec;38(8):661-70.

Fujikura J., Hosoda K., Iwakura H., Tomita T., Noguchi M., Masuzaki H., Tanigaki K., Yabe D., Honjo T., Nakao K., 2006. Notch/Rbp-j signaling prevents premature endocrine and ductal cell differentiation in the pancreas. Cell Metab. 2006 Jan;3(1):59-65.

Gao N., LeLay J., Vatamaniuk MZ., Rieck S., Friedman JR., Kaestner KH., 2008. Dynamic regulation of Pdx1 enhancers by Foxa1 and Foxa2 is essential for pancreas development. Genes Dev. 2008 Dec 15;22(24):3435-48.

Gittes GK., Galante PE., Hanahan D., Rutter WJ., Debase HT., 1996. Lineage-specific morphogenesis in the developing pancreas: role of mesenchymal factors. Development. 1996 Feb;122(2):439-47.

Golosow N., Grobstein C., 1962. Epitheliomesenchymal interaction in pancreatic morphogenesis. Dev Biol. 1962 Apr;4:242-55.

Gouzi M., Kim YH., Katsumoto K., Johansson K., Grapin-Botton A., 2011. Neurogenin3 initiates stepwise delamination of differentiating endocrine cells during pancreas development. Dev Dyn. 2011 Mar;240(3):589-604.

ACCEPTED MANUSCRIPT

Gradwohl G., Dierich A., LeMeur M., Guillemot F., 2000. neurogenin3 is required for the development of the four endocrine cell lineages of the pancreas. Proc Natl Acad Sci U S A. 2000 Feb 15;97(4):1607-11.

Hinton A., Hunter S., Reyes G., Fogel GB., King CC., 2012. From pluripotency to islets: miRNAs as critical regulators of human cellular differentiation. Adv Genet. 2012;79:1-34.

Horn S., Kobberup S., Jørgensen MC., Kalisz M., Klein T., Kageyama R., Gegg M., Lickert H., Lindner J., Magnuson MA., Kong YY., Serup P., Ahnfelt-Rønne J., Jensen JN., 2000. Mind bomb 1 is required for pancreatic β-cell formation. Proc Natl Acad Sci U S A. 2012 May 8;109(19):7356-61.

In't Veld P. , Marichal M., 2010. Microscopic anatomy of the human islet of Langerhans. Adv Exp Med Biol. 2010;654:1-19.

Jensen J., 2004. Gene regulatory factors in pancreatic development. Dev Dyn. 2004 Jan;229(1):176-200.

Jensen J., Pedersen EE., Galante P., Hald J., Heller RS., Ishibashi M., Kageyama R., Guillemot F., Serup P., Madsen OD., 2000. Control of endodermal endocrine development by Hes-1. Nat Genet. 2000 Jan;24(1):36-44.

Juhl K, Sarkar SA, Wong R, Jensen J, Hutton JC., 2008. Mouse pancreatic endocrine cell transcriptome defined in the embryonic Ngn3-null mouse. Diabetes. 2008 Oct;57(10):2755-61.

Kesavan G., Lieven O., Mamidi A., Öhlin ZL., Johansson JK., Li WC., Lommel S., Greiner TU., Semb H., 2014. Cdc42/N-WASP signaling links actin dynamics to pancreatic β cell delamination and differentiation. Development. 2014 Feb;141(3):685-96.

Kyriakides TR., Zhu YH., Smith LT., Bain SD., Yang Z., Lin MT., Danielson KG., Iozzo RV., LaMarca M., McKinney CE., Ginns EI., Bornstein P., 1998. Mice that lack thrombospondin 2 display connective tissue abnormalities that are associated with disordered collagen fibrillogenesis, an increased vascular density, and a bleeding diathesis. J Cell Biol. 1998 Jan 26;140(2):419-30.

Lee CS., Sund NJ., Vatamaniuk MZ., Matschinsky FM., Stoffers DA., Kaestner KH., 2002. Foxa2 controls Pdx1 gene expression in pancreatic beta-cells in vivo. Diabetes. 2002 Aug;51(8):2546-51.

Liao Y., Smyth GK., Shi W., 2013. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. 2013 May 1;41(10):e108.

Lickert H, Kispert A, Kutsch S, Kemler R.,2001. Expression patterns of Wnt genes in mouse gut development. Mech Dev. 2001 Jul;105(1-2):181-4.

ACCEPTED MANUSCRIPT

Liu Z., Habener JF., 2008. Glucagon-like peptide-1 activation of TCF7L2-dependent Wnt signaling enhances pancreatic beta cell proliferation. J Biol Chem. 2008 Mar 28;283(13):8723-35.

Marullo L., El-Sayed Moustafa JS., Prokopenko I., 2014. Insights into the genetic susceptibility to type 2 diabetes from genome-wide association studies of glycaemic traits. Curr Diab Rep. 2014 Nov;14(11):551.

Migliorini A., Bader E., Lickert H., 2014. Islet cell plasticity and regeneration. Mol Metab. 2014 Jan 22;3(3):268-74.

Miralles F., Czernichow P., Ozaki K., Itoh N., Scharfmann R., 1999. Signaling through fibroblast growth factor receptor 2b plays a key role in the development of the exocrine pancreas. Proc Natl Acad Sci U S A. 1999 May 25;96(11):6267-72.

Mitchell RK., Mondragon A., Chen L., Mcginty JA., French PM., Ferrer J., Thorens B., Hodson DJ., Rutter GA., Da Silva Xavier G., 2015. Selective disruption of Tcf7l2 in the pancreatic β cell impairs secretory function and lowers β cell mass. Hum Mol Genet. 2015 Mar 1;24(5):1390-9.

Morán I., Akerman I., van de Bunt M., Xie R., Benazra M., Nammo T., Arnes L., Nakić N., García-Hurtado J., Rodríguez-Seguí 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., McCarthy MI., Ravassard P., Ferrer J.,2012. Human β cell transcriptome analysis uncovers lncRNAs that are tissue-specific, dynamically regulated, and abnormally expressed in type 2 diabetes. Cell Metab. 2012 Oct 3;16(4):435-48.

Pagliuca FW., Melton DA., 2013. How to make a functional β-cell. Development. 2013 Jun;140(12):2472-83.

Pan FC., Wright C., 2011. Pancreas organogenesis: from bud to plexus to gland. Dev Dyn. 2011 Mar;240(3):530-65.

Pasquali L., Gaulton KJ., Rodríguez-Seguí SA., Mularoni L., Miguel-Escalada I., Akerman I., Tena JJ., Morán I., Gómez-Marín C., van de Bunt M., Ponsa-Cobas J., Castro N., Nammo T., Cebola I., García-Hurtado J., Maestro MA., Pattou F., Piemonti L., Berney T., Gloyn AL., Ravassard P., Gómez-Skarmeta JL., Müller F., McCarthy MI., Ferrer J., 2014. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. Nat Genet. 2014 Feb;46(2):136-43.

Pictet RL., Clark WR., Williams RH., Rutter WJ., 1972. An ultrastructural analysis of the developing embryonic pancreas. Dev Biol. 1972 Dec;29(4):436-67.

Puri S., Hebrok M., 2010. Cellular plasticity within the pancreas-lessons learned from development. Dev Cell. 2010 Mar 16;18(3):342-56.

ACCEPTED MANUSCRIPT

Raducanu A., Lickert H., 2012. Understanding pancreas development for β-cell repair and replacement therapies. Curr Diab Rep. 2012 Oct;12(5):481-9.

Ritchie ME., Phipson B., Wu D., Hu Y., Law CW., Shi W., Smyth GK., 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015 Jan 20.

Schwartz MW., Seeley RJ., Tschöp MH., Woods SC., Morton GJ., Myers MG., D'Alessio D., 2013. Cooperation between brain and islet in glucose homeostasis and diabetes. Nature. 2013 Nov 7;503(7474):59-66.

Schwitzgebel VM., 2014. Many faces of monogenic diabetes. J Diabetes Investig. 2014 Mar 23;5(2):121-33.

Schwitzgebel VM., Scheel DW., Conners JR., Kalamaras J., Lee JE., Anderson DJ., Sussel L., Johnson JD., German MS., 2000. Expression of neurogenin3 reveals an islet cell precursor population in the pancreas. Development. 2000 Aug;127(16):3533-42.

Shih HP., Wang A., Sander M., 2013. Pancreas organogenesis: from lineage determination to morphogenesis. Annu Rev Cell Dev Biol. 2013;29:81-105.

Sugiyama T, Rodriguez RT, McLean GW, Kim SK., 2007. Conserved markers of fetal pancreatic epithelium permit prospective isolation of islet progenitor cells by FACS. Proc Natl Acad Sci U S A. 2007 Jan 2;104(1):175-80

Sung CC., Liao MT., Lu KC., Wu CC., 2012. Role of vitamin D in insulin resistance. J Biomed Biotechnol. 2012;2012:634195.

Uchio-Yamada K., Sawada K., Tamura K., Katayama S., Monobe Y., Yamamoto Y., Ogura A., Manabe N., 2013. Tenc1-deficient mice develop glomerular disease in a strain-specific manner. Nephron Exp Nephrol. 2013;123(3-4):22-33.

van Arensbergen J., García-Hurtado J., Moran I., Maestro MA., Xu X., Van de Casteele M., Skoudy AL., Palassini M., Heimberg H., Ferrer J., 2010. Derepression of Polycomb targets during pancreatic organogenesis allows insulin-producing beta-cells to adopt a neural gene activity program. Genome Res. 2010 Jun;20(6):722-32.

Wang D., Wang W., Dawkins P., Paterson T., Kalsheker N., Sallenave JM., Houghton AM., 2011. Deletion of Serpina1a, a murine α1-antitrypsin ortholog, results in embryonic lethality. Exp Lung Res. 2011 Jun;37(5):291-300.

Weedon MN., Cebola I., Patch AM., Flanagan SE., De Franco E., Caswell R., Rodríguez-Seguí SA., Shaw-Smith C., Cho CH., Lango Allen H., Houghton JA., Roth CL., Chen R., Hussain K., Marsh P., Vallier L., Murray A., 2014. Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. Nat Genet. 2014 Jan;46(1):61-4.

ACCEPTED MANUSCRIPT

Wei R., Yang J., Liu GQ., Gao MJ., Hou WF., Zhang L., Gao HW., Liu Y., Chen GA., Hong TP., 2013. Dynamic expression of microRNAs during the differentiation of human embryonic stem cells into insulin-producing cells. Gene. 2013 Apr 15;518(2):246-55.

Whelan J., Poon D., Weil PA., Stein R., 1989. Pancreatic beta-cell-type-specific expression of the rat insulin II gene is controlled by positive and negative cellular transcriptional elements. Mol Cell Biol. 1989 Aug;9(8):3253-9.

Wurst W., Bally-Cuif L., 2001. Neural plate patterning: upstream and downstream of the isthmic organizer. Nat Rev Neurosci. 2001 Feb;2(2):99-108.

Xu J., Pang ZP., Shin OH., Südhof TC., Synaptotagmin-1 functions as a Ca2+ sensor for spontaneous release. Nat Neurosci. 2009 Jun;12(6):759-66.

Zhou Q., Law AC., Rajagopal J., Anderson WJ., Gray PA., Melton DA., 2007. A multipotent progenitor domain guides pancreatic organogenesis. Dev Cell. 2007 Jul;13(1):103-14.

Zhou Y., Park SY., Su J., Bailey K., Ottosson-Laakso E., Shcherbina L., Oskolkov N., Zhang E., Thevenin T., Fadista J., Bennet H., Vikman P., Wierup N., Fex M., Rung J., Wollheim C., Nobrega M., Renström E., Groop L., Hansson O., 2014. TCF7L2 is a master regulator of insulin production and processing. Hum Mol Genet. 2014 Dec 15;23(24):6419-31.

Zorn A.M., Wells J.M. 2009. Vertebrate endoderm development and organ formation. Annu Rev Cell Dev Biol. 2009; Vol. 25: 221-251

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Abbreviations:

2210010C04Rik - RIKEN cDNA 2210010C04 gene
5330417C22Rik - RIKEN cDNA 5330417C22 gene
A1AT - alpha1-antitrypsin
Ago1 - argonaute RISC catalytic component 1
Amy2a5 - amylase 2a5
Amy2b - amylase, alpha 2B
AP1/JUN - jun proto-oncogene
Bmp - bone morphogenetic protein
CAM - cell adhesion molecule
CD49f - Integrin subunit alpha 6 (Itga6)
Cela1 - Chymotrypsin-like elastase family member 1
ChIP-seq - chromatin-immunoprecipitation sequencing
Cldn10 - claudin 10
Cpa1 - Carboxypeptidase 1
Drosha - drosha, ribonuclease type III
E - embryonic day
ECM - extracellular matrix
Edaradd - Ectodysplasin-A receptor (EDAR)-associated adapter protein
EGF - epidermal growth factor
EMT - epithelial-mesenchymal transition
En1 - engrailed homeobox 1
FACS - Fluorescence Activated Cell Sorting
FGF - Fibroblast growth factor
Fgf1 - Fibroblast growth factor 1
Fgf10 - fibroblast growth factor 10
Fgf15 - fibroblast growth factor 15
Fgf20 - fibroblast growth factor 20
Fgf23 - fibroblast growth factor 23
Fgf6 - fibroblast growth factor 6
Fgf7 - fibroblast growth factor 7
Fgf8 - fibroblast growth factor 8
Fgf9 - fibroblast growth factor 9
Foxa2 - Forkhead-Box-Protein A2
FVF- - Foxa2-Venus negative
FVF+ - Foxa2-Venus positive
Fzd9 - frizzled homolog 9
Fzd10 - frizzled homolog 10
G6pc2 - glucose-6-phosphatase, catalytic, 2
Gata6 - GATA binding protein 6
Gck - glucokinase
Gdf10 - growth differentiation factor 10
Ghr - growth hormone receptor
Ghrl - ghrelin/obestatin prepropeptide
Gip - gastric inhibitory polypeptide
Gipr - gastric inhibitory polypeptide receptor
Gli3 - GLI-Kruppel family member GLI3
Glp1- glucagon-like peptide 1

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Glp1r - glucagon-like peptide 1 receptor
GO - gene ontology
Grb7 - growth factor receptor bound protein 7
GRNs - gene regulation networks
GWAS - genome-wide association study
H3K27ac - Histone 3 lysine 27 acetylation
H3K4me1 - Histone 3 lysine 4 monomethylation
H3K4me3 - Histone 3 lysine 4 trimethylation
Hdac1 - histone deacetylase 1
Hes1 - hairy and enhancer of split 1
Hey1 - hairy/enhancer-of-split related with YRPW motif 1
Hh - Hedegehog
Hhex - hematopoietically expressed homeobox
High mag - higher magnitude
HNF - hepatocyte nuclear factor
Hnf1a - hepatocyte nuclear factor 1, alpha
Hnf1b - hepatocyte nuclear factor1, beta
Hnf4a - hepatocyte nuclear factor 4, alpha
Hnf4b - hepatocyte nuclear factor4, beta
Iaap - islet amyloid polypeptide
Igf1 - insulin-like growth factor 1
Igfbp4/5 - insulin-like growth factor binding protein 4/5
Ihh - Indian hedgehog
Il11ra2 - interleukin 11 receptor, alpha chain 2
Ins1 - Insulin 1
Ins2 - Insulin 2
Insrr - insulin receptor-related receptor
iPSC - induced pluripotent stem cells
Isl1 – ISL LIM homeobox 1
Jag1 - jagged1
Lum - Lumican
MAF - v-maf avian musculoaponeurotic fibrosarcoma oncogene homolog
MAFB - v-maf musculoaponeurotic fibrosarcoma oncogene family, protein B
Med12 - mediator complex subunit 12
Med23 - mediator complex subunit 23
Mist1/ Bhlha15 - Basic Helix-Loop-Helix Family, Member A15
MODY - maturity-onset diabetes of the young
MPC - multipotent progenitors
NeuroD1 - neurogenic differentiation 1
Ngn3 - Neurogenin 3
Nkx2-2 - NK2 homeobox 2
Nkx6.1 - NK6 homeobox 1
Npy - neuropeptide Y
Nr5a2 - nuclear receptor subfamily 5
Onecut1/Hnf6 - one cut homeobox 1
Onecut3 - one cut domain, family member 3
Osr1 - odd-skipped related 1
Otx2 - orthodenticle homeobox 2
Pax4 - paired box 4
Pax6 - paired box 6

ACCEPTED MANUSCRIPT

PC - principal component
Pcsk1-3 - proprotein convertase subtilisin/kexin type 1-3
Pdgfc - platelet derived growth factor C
Pdx1 - pancreatic duodenal transcription homeobox 1 factor
Pgf - placental growth factor
Ptf1a - pancreas specific transcription factor, 1a
Pyy - peptide YY
RA - retinoic acid
Rbm47 - RNA binding motif protein 47
Rxrg - Retinoic acid receptor gamma
Scgn - secretagogin, EF-hand calcium binding protein
Sdf1 - stromal cell-derived factor 1
Serpina10 - Serpine peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 10
Serpina1a - serine (or cysteine) preptidase inhibitor, clade A, member 1A
Serpina1b - serine (or cysteine) preptidase inhibitor, clade A, member 1B
Serpina1d - serine (or cysteine) preptidase inhibitor, clade A, member 1D
Serpina1e - serine (or cysteine) preptidase inhibitor, clade A, member 1E
Sfrp1 - secreted frizzled-related protein 1
Shh - sonic hedgehog
Six3 - sine oculis-related homeobox 3
Slc2a2 - solute carrier family 2 (facilitated glucose transporter), member 2
Slc30a8 - solute carrier family 30 (zinc transporter), member 8
Slc38a3 - solute carrier family 38, member 3
Smarca2 - SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2
Smarca4 - SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4
Smarcc1 - SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily c, member 1
Smarcc2 - SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily c, member 2
Smarcd1 - SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 1
Smarcd1 - SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 1
Smarce1 - SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily e, member 1
Snai1/2 - snail family zinc finger 1/2
Sox11 - SRY (sex determining region Y)-box 11
Sox9 – SRY (sex determining region Y)-box 9
Srpx2 - sushi-repeat-containing protein, X-linked 2
Sst - somatostatin
Syt13 - Synaptotagmin 13
Syt6 - Synaptotagmin 6
Syt7 - Synaptotagmin 7
Syt8 - Synaptotagmin 8
Sytl1 - synaptotagmin-like 1
Sytl4 - synaptotagmin-like 4
T1/2D - type 1/ 2 diabetes

ACCEPTED MANUSCRIPT

Tbx19 - T-box 19
Tbx6 - T-box 6
Tcf7l1 - transcription factor 7-like 1
Tcf7l2 - transcription factor 7-like 2
Tenc1 - tensin like C1 domain-containing phosphatase
TF - transcription factor
TFBS - transcription factor binding site
Tgfß - transforming growth factor, beta 1
Thbs2 - thrombospondin 2
Tmem171 - transmembrane protein 171
TSS - transcriptional starting site
Twist1 - twist family bHLH transcription factor 1
Vdr - vitamin D receptor
Vwa5b2 - von Willebrand factor A domain containing 2B2
Wnt - wingless
Wnt1 - wingless-type MMTV integration site family, member 1
Wnt10a - wingless-type MMTV integration site family, member 10A
Wnt10b - wingless-type MMTV integration site family, member 10B
Wnt2 - wingless-type MMTV integration site family, member 2
Wnt3a - wingless-type MMTV integration site family, member 3A
Wnt4 - wingless-type MMTV integration site family, member 4
Wnt5a - wingless-type MMTV integration site family, member 5A
Wnt6 - wingless-type MMTV integration site family, member 6
Zeb1/2 - zinc finger E-box binding homeobox 1/2

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

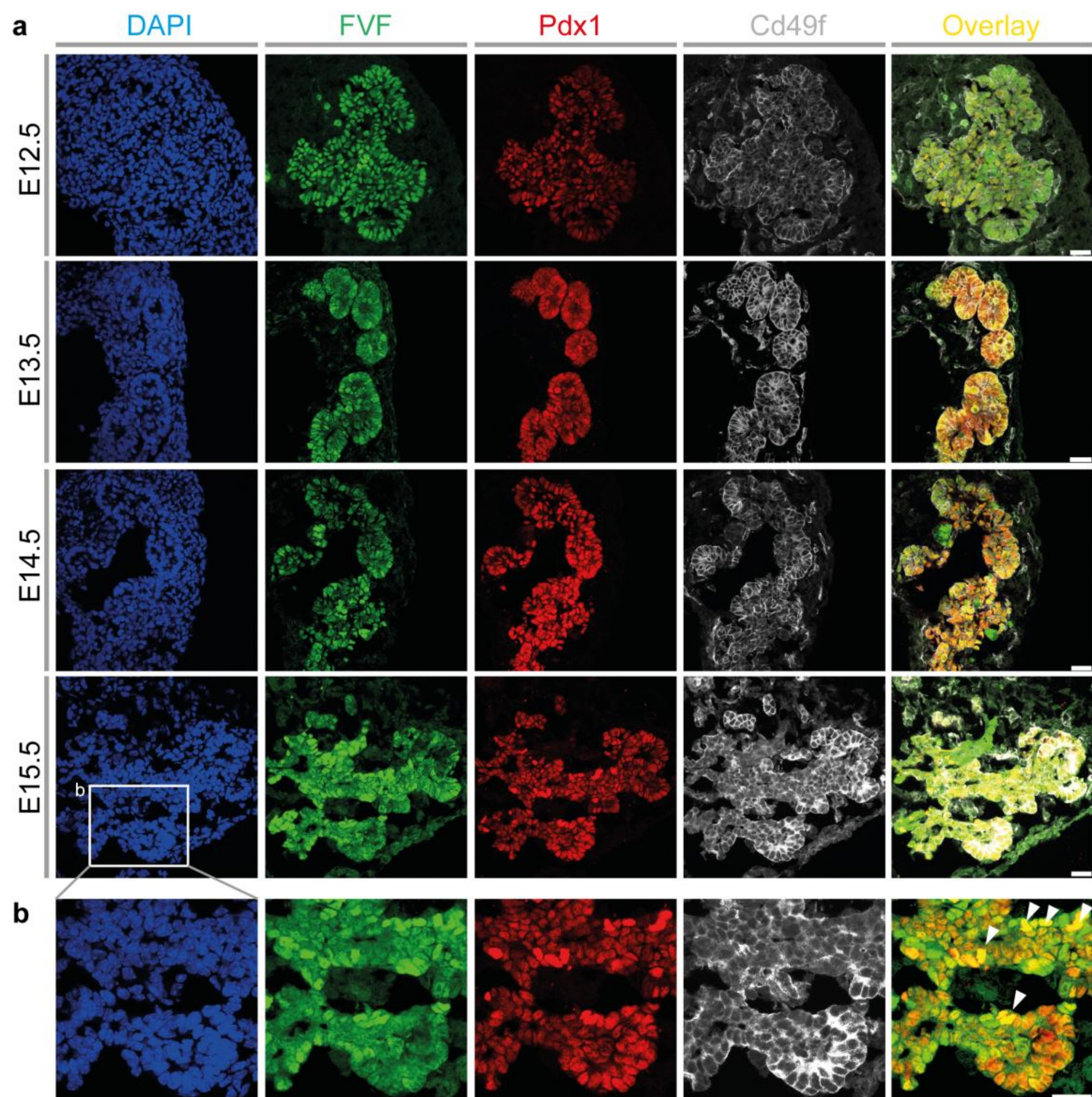


Figure 1

ACCEPTED MANUSCRIPT

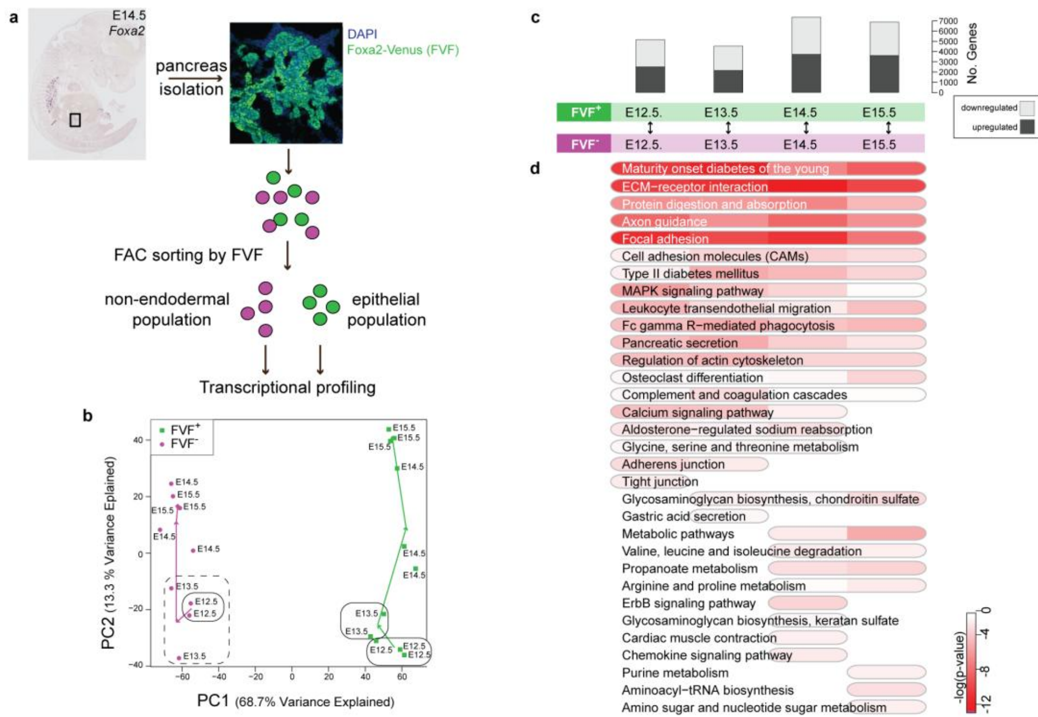


Figure 2

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

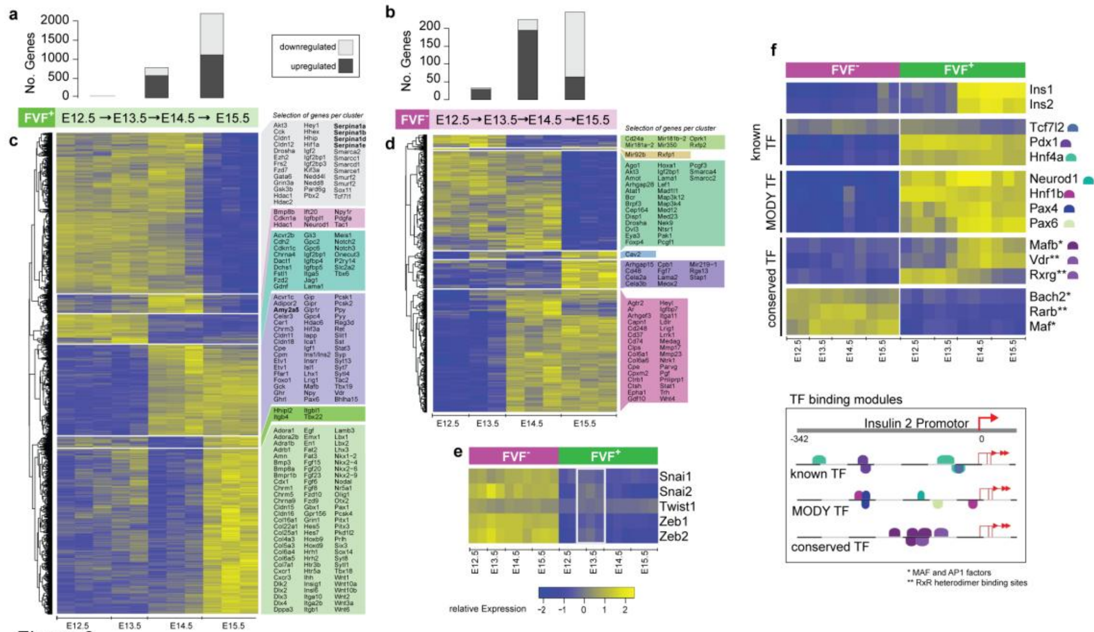


Figure 3

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

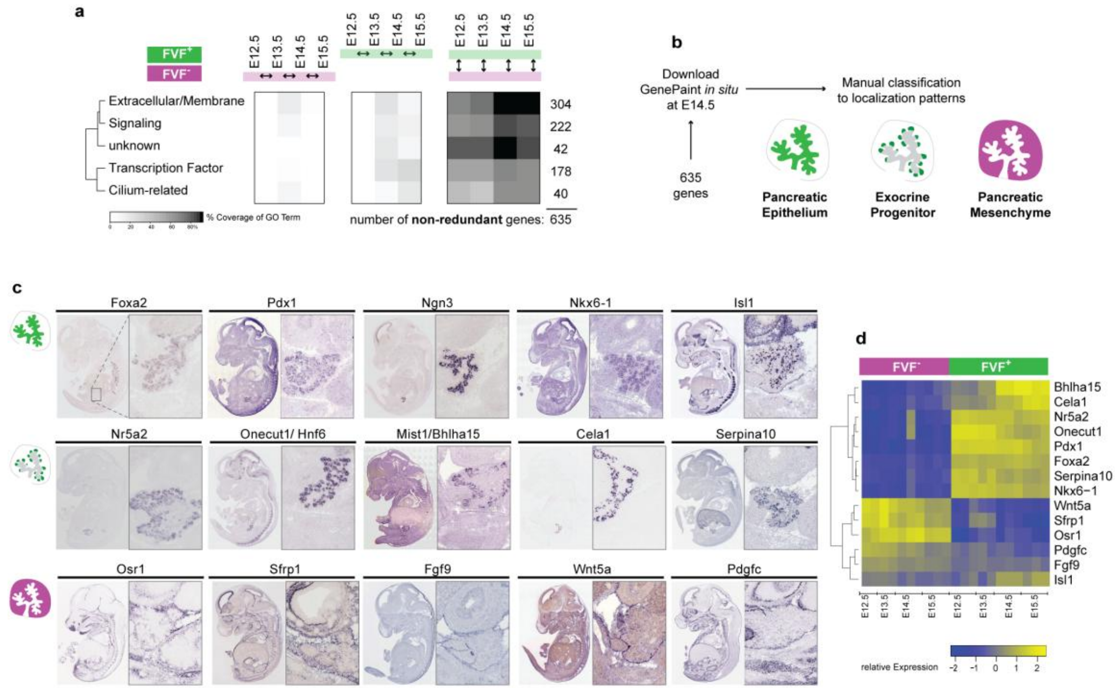


Figure 4

ACCEPTED MANUSCRIPT

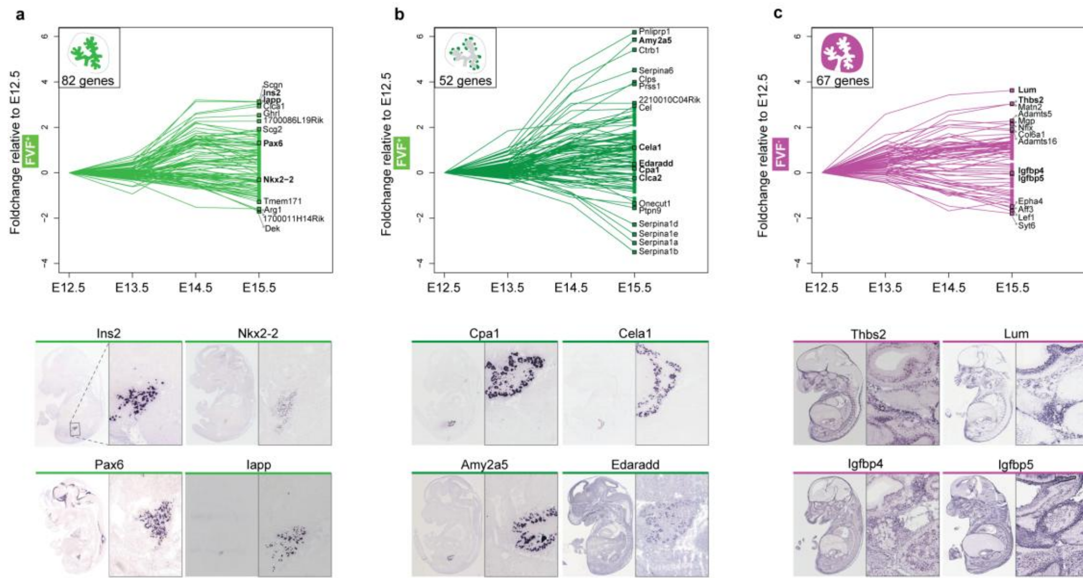


Figure 5

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

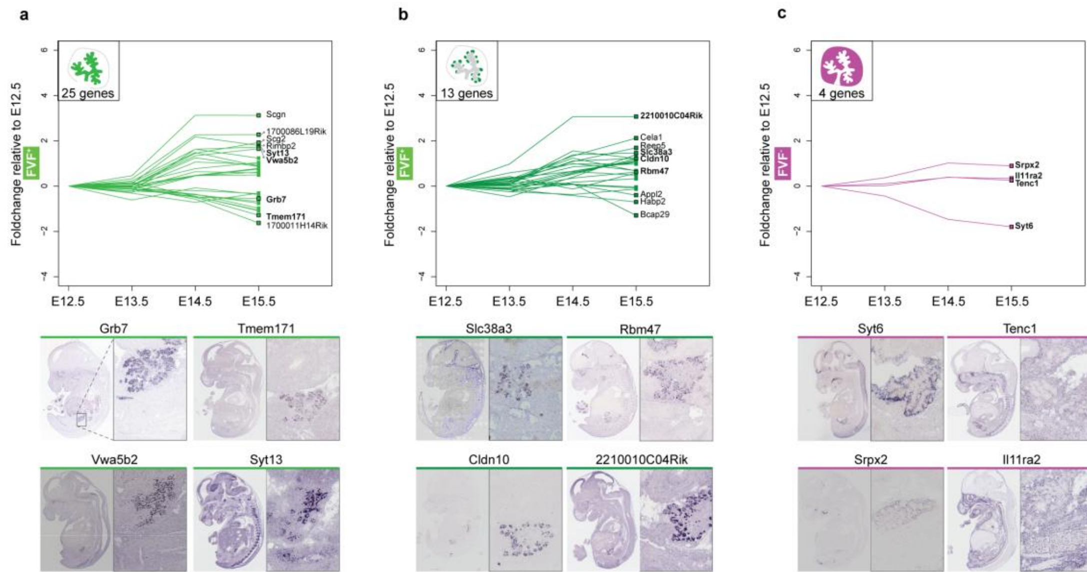


Figure 6

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

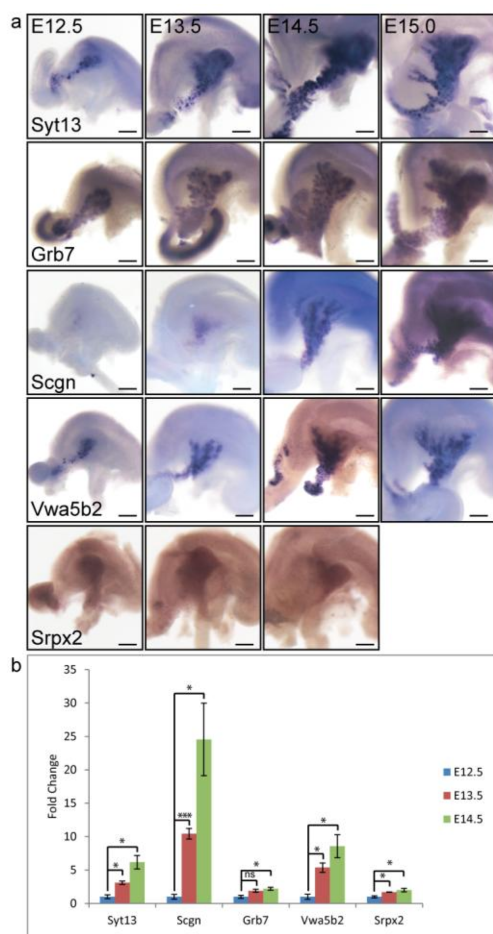


Figure 7

ACCEPTED MANUSCRIPT

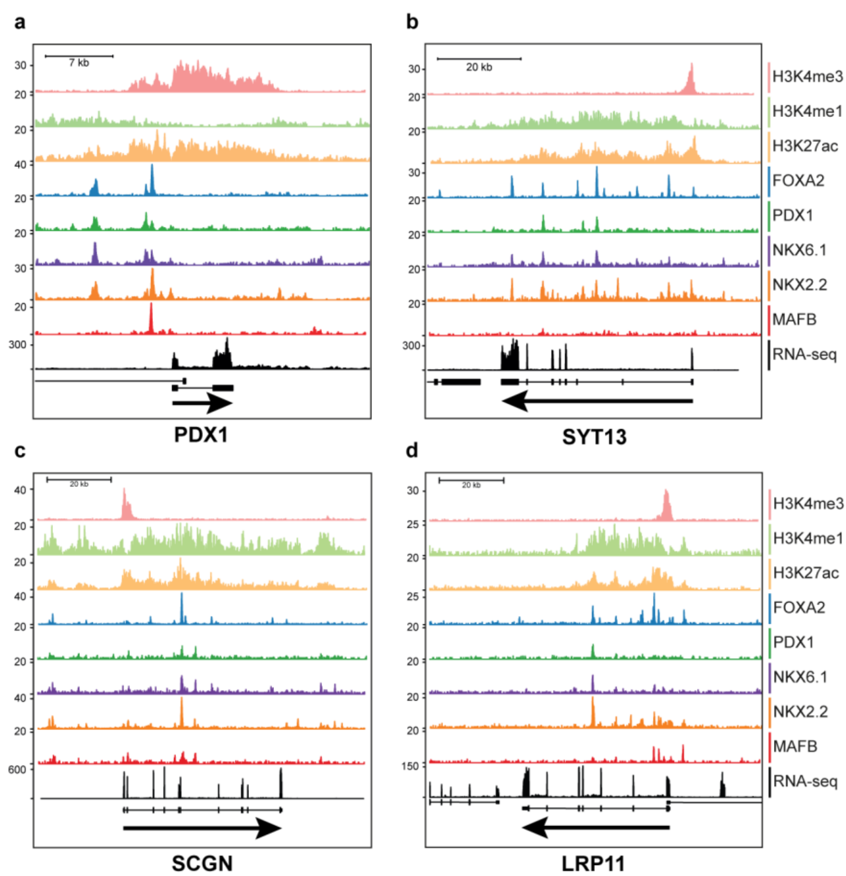


Figure 8

ACCEPTED

ACCEPTED MANUSCRIPT

Highlights:

- **High-resolution spatio-temporal gene expression profile of pancreas development**
- **Cell type-specific profile reveals known and many novel developmental regulators**
- **Differential expression in pancreatic compartments reveal details about tissue interactions**
- ***In situ* mRNA expression analysis validates tissue specificity and spatial expression**
- **The established resource is of unique value to study mouse and human development and diabetes**

ACCEPTED MANUSCRIPT