Technische Universität München
Fakultät für Medizin

Transcriptional Regulation of Endocrine Cell Formation and Function in Pancreas and Intestine

Michael Sterr

Vollständiger Abdruck der von der Fakultät für Medizin der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender: Prof. Dr. Marc Schmidt-Supprain

Prüfer der Dissertation:
1. Prof. Dr. Heiko Lickert
2. Apl. Prof. Dr. Johannes Beckers
3. Prof. Dr. Eckhard Wolf

Die Dissertation wurde am 3. April 2019 bei der Technischen Universität München eingereicht und durch die Fakultät für Medizin am 11. Februar 2020 angenommen.

Contents

List of Figures

Abbreviations

bHLH	Basic helix-loop-helix
bp	Base pair
bZIP	Basic leucine zipper
CBC	Crypt base columnar cell
ChIP-seq	Chromatin immunoprecipitation sequencing
CS	Carnegie stage
CSLM	Confocal laser scanning microscopy
DCA	Deep count autoencoder network
DE	Definitive endoderm
E	Embryonic day
EC	Enterochromaffin cell
ECL	Enterochromaffin-like cell
EEC	Enteroendocrine cell
Ent	Enterocyte
Ent Pro	Enterocyte progenitor cell
ESC	Embryonic stem cell
FG	Posterior foregut
FHKD	Forkhead
GC	Goblet cell
GO	Gene ontology
GSEA	Gene set enrichment analysis
GT	Gut tube
GWAS	Genome-wide association study
hESC	Human embryonic stem cell
Het	Heterozygous
Hom	Homozygous
ILC2	Group 2 innate lymphoid cell
Insl5	Insulin-like peptide 5
iPSC	Induced pluripotent stem cell
IRF	Interferon regulatory factor
ISC	Intestinal stem cell
kb	Kilobase
lncRNA	Long non-coding RNA
MODY	Maturity onset diabetes of the young
MPC	Multi-potent pancreatic progenitor

NLS ..... Nuclear localization signal
NR ..... Nuclear receptor
PC ..... Paneth cell
PCA ..... Principal component analysis
PCP ..... Planar cell polarity
PCR ..... Polymerase chain reaction
PE ..... Pancreatic endoderm
PP ..... Pancreatic progenitor
PTD ..... Protein trunsduction domain
qPCR ..... quantitative PCR
RFX ..... Regulatory factor X
RNA-seq ..... RNA sequencing
rpm ..... Rotations per minute
RT ..... Room temperature
scRNA-seq ..... Single-cell RNA sequencing
SD ..... Standard deviation
Sec Pro ..... Secretory progenitor cell
SEM ..... Standard error of the mean
SNP ..... Single nucleotide polymorphism
T2DM ..... Type 2 Diabetes Mellitus
T_H2 ..... Type 2 helper T cell
TA ..... Transit amplifying cells
TC ..... Tuft cell
TSS ..... Transcription start site
UTR ..... Untranslated region
wpc ..... Weeks post conception
ZF ..... Zink finger

Abstract

During the past decades, diabetes mellitus gained almost epidemic properties, affecting 425 million worldwide. Hormones secreted by endocrine cells of the pancreas and intestine are essential for the homeostatic control of blood glucose and energy intake, and disturbances of the hormonal balance are implicated in the development of metabolic diseases, such as diabetes. Understanding the developmental programs that govern the formation of pancreatic and intestinal endocrine cells is an important step towards the improvement of diagnosis and treatment of diabetic patients.

Pdx1 is one of the most important regulators of pancreas development. It is expressed throughout pancreatic organogenesis and becomes gradually restricted to β-cells and δ-cells in the mature organ. Mutations in the *PDX1* gene are associated with monogenic and type 2 diabetes, and the homozygous loss causes pancreatic agenesis. While the role of Pdx1 is well established in model organisms, the function of PDX1 during human pancreas and β-cell development is not understood in great detail. In the first part of this thesis, we use *in vitro* differentiated pancreatic progenitors, derived from a novel induced pluripotent stem cell (iPSC) line to identify PDX1 regulated genes during early pancreagenesis. We found that PDX1 has stage-specific functions in pancreatic progenitors and mature β-cells, regulating pancreatic development and β-cell function, respectively. Moreover, analysis of diabetes-associated SNPs in active enhancers of pancreatic progenitors indicated a contribution of early pancreatic development to an increased susceptibility to diabetes. To assess the impact of two common PDX1 missense mutations (C18R and P33T), we analyzed iPSCs generated from heterozygous mutation carriers and corresponding homozygous isogenic control cell lines. We found that both mutations result in the deregulation of important β-cell genes, suggesting an increased predisposition for diabetes in mutation carriers. Our results provide mechanistic insight into pancreas formation and show how mutations of PDX1 affect pancreas and β-cell development and predispose for diabetes.

In response to food intake, enteroendocrine cells of the intestinal epithelium secrete a multitude of hormones, which regulate various aspects of energy homeostasis, including appetite and satiety, digestion, and the release of insulin from pancreatic β-cells. Enteroendocrine cells are therefore important players in the pathogenesis and treatment of metabolic diseases such as diabetes. However, the mechanisms controlling their differentiation from intestinal stem cells are not fully understood. The second part of this thesis addresses the role of the pioneer transcription factor Foxa2 in the initiation of enteroendocrine differentiation in the intestine. Using combination of bulk and single-cell transcriptomics together with ChIP-seq and ATAC-seq we dissect the enteroendocrine lineage decision and show a dose-dependent function of Foxa2. We found that Foxa2 is expressed at low levels in intestinal stem cells, where it regulates metabolic processes, while it is expressed at high levels in enteroendocrine progenitors, to establish the epigenetic premises for the induction of the endocrine cell fate. These results improve our understanding of the mechanisms underlying intestinal endocrinogenesis and define novel roles of Foxa2 in the differentiation of enteroendocrine cells.

Zusammenfassung

Diabetes mellitus betrifft weltweit inzwischen 425 Millionen Menschen und die Zahlen steigen weiter an. Hormone, die von endokrinen Zellen der Bauchspeicheldrüse und des Darms ausgeschüttet werden, sind für die homöostatische Kontrolle des Blutzuckers und der Energiezufuhr von wesentlicher Bedeutung und Störungen dieses Hormonhaushalts führen zur Entwicklung von Stoffwechselerkrankungen wie Diabetes. Das Verständnis der Entwicklungsprogramme, welche die Bildung von endokrinen Zellen in Pankreas und Darm steuern, ist ein wichtiger Schritt zur Verbesserung der Diagnose und Behandlung von Diabetikern.

Pdx1 ist einer der wichtigsten Regulatoren der Pankreasentwicklung und ist während der frühen Entwicklung im gesamten Pankreas exprimiert. Im Laufe der Organentwicklung wird die Expression zunehmend auf β-Zellen und δ-Zellen beschränkt. Mutationen des Gens sind mit monogenem Diabetes und Typ-2-Diabetes assoziiert, und der homozygote Verlust verursacht eine Pankreasagenesie. Während die Rolle von Pdx1 in Modellorganismen gut erforscht ist, ist die Funktion von PDX1 in der Pankreasentwicklung des Menschen nicht im Detail verstanden. Im ersten Teil dieser Arbeit verwenden wir Pankreas-Vorläuferzellen, die aus einer induzierten pluripotenten Stammzelllinie (iPSC) differenziert wurden, um PDX1-regulierte Gene während der frühen Pankreasentwicklung zu identifizieren. Wir fanden heraus, dass PDX1 stadienspezifische Funktionen in Vorläuferzellen und reifen β-Zellen hat, welche die Pankreasentwicklung bzw. die β-Zell-Funktion regulieren. Darüber hinaus zeigte die Analyse von Diabetes-assoziierten SNPs in Pankreas-Vorläuferzellen einen möglichen Zusammenhang zwischen früher Pankreasentwicklung und der Anfälligkeit für Diabetes. Um den Einfluss von zwei häufigen PDX1-Missense-Mutationen (C18R und P33T) zu untersuchen, wurden hetero- und homozygote iPSC Linien untersucht. Beide Mutationen führen zur Deregulierung wichtiger Gene, was auf ein erhöhtes Risiko für Diabetes bei Mutationsträgern hindeutet. Unsere Ergebnisse liefern mechanistische Einblicke in die Pankreasentwicklung und zeigen, wie Mutationen von PDX1 die Pankreas- und β-Zell-Entwicklung beeinflussen und für Diabetes prädisponieren.

In Reaktion auf Nahrungsaufnahme setzen enteroendokrine Zellen des Darmepithels eine Vielzahl von Hormonen frei, die verschiedene Aspekte der Energiehomöostase, einschließlich Sättigung und der Freisetzung von Insulin aus β-Zellen, regulieren. Enteroendokrine Zellen spielen daher eine wichtige Rolle bei der Pathogenese und Behandlung von Stoffwechselerkrankungen wie Diabetes. Die Mechanismen, die ihre Differenzierung aus Darmstammzellen steuern, sind jedoch nicht vollständig verstanden. Der zweite Teil dieser Arbeit befasst sich mit der Rolle des Pioniertranskriptionsfaktors Foxa2 während der Initiierung der enteroendokrinen Differenzierung. Unter Verwendung einer Kombination aus Bulk- und Einzelzelltranskriptomik zusammen mit ChIP-seq und ATAC-seq analysieren wir die Spezifikation enteroendokriner Zellen und zeigen eine dosisabhängige Funktion von Foxa2. In Darmstammzellen ist Foxa2 in geringen Mengen exprimiert und reguliert Stoffwechselprozesse, während die Expression in enteroendokrinen Vorläuferzellen ansteigt, um die epigenetischen Voraussetzungen für die Induktion des endokrinen transkriptionellen Programms zu schaffen. Diese Ergebnisse verbessern unser Verständnis der Mechanismen, die der intestinalen Endokrinogenese zugrunde liegen, und zeigen neue Rollen von Foxa2 bei der Differenzierung enteroendokriner Zellen.

1 Introduction

1.1 Diabetes Mellitus

Diabetes mellitus is a heterogeneous group of metabolic disorders that is characterized by elevated blood glucose levels due to absence or malfunction of insulin-producing β-cells, disturbed insulin action or both. If left untreated, diabetes can damage various organs, leading to disabling and potentially life-threatening complications such as cardiovascular disease, neuropathy, nephropathy, retinopathy and blindness. During the last few decades, the incidence of diabetes has been rapidly increasing. The 425 million worldwide cases of diabetes to date (2017) are predicted to increase by 48% to 629 million cases in 2045 (Figure 1.1). Thus, in addition to the human affliction, diabetes is becoming an increasing economic burden, with an estimated healthcare expenditure of 727 billion USD in 2017. Based on etiology and clinical presentation, diabetes is classified into several sub-types, including type 1 and type 2 diabetes mellitus, gestational diabetes and monogenic diabetes [1, 2].

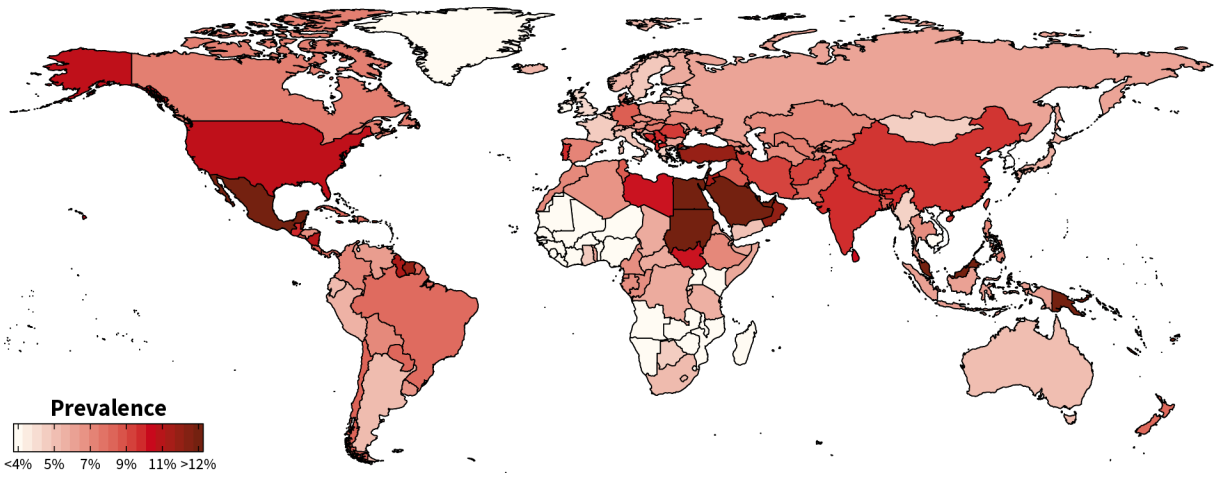

Figure 1.1: The Prevalence of Diabetes. Estimated age-adjusted prevalence of diabetes in adults (20-79 years) in 2017. Data obtained from [1].

Type 1 diabetes mellitus is caused by an autoimmune-mediated destruction and ultimately the almost complete loss of pancreatic β-cells. As a consequence, patients are dependent on the administration of insulin to maintain proper glucose levels [1, 2]. The onset of the disease is multifactorial and, to current understanding, involves a combination of genetic predisposition and environmental triggers such as viral infection, exposure to toxins or dietary factors. However, the exact mechanisms triggering β-cell destruction are not yet fully understood

and disease progression is, to current knowledge, not preventable [3]. The disease usually develops during childhood or adolescence, but can appear at any age. Type 1 diabetes accounts for approximately 7.5% of diabetes cases, though the incidence is increasing worldwide, likely due to environmental factors [3–5].

Type 2 diabetes mellitus (T2DM) is the most common form of diabetes and accounts for approximately 90% of all diabetes cases worldwide [6–8]. In T2DM, hyperglycemia is the result of insufficient insulin secretion and an impaired response to insulin in the peripheral tissues. Disease onset is usually slow and is preceded by a pre-diabetic stage of systemically reduced insulin sensitivity, which causes β-cells to increase their insulin secretion to compensate for insulin resistance [1, 9]. With further disease progression, β-cells eventually become dysfunctional and β-cell mass is lost [9, 10]. T2DM is a multifactorial disease as well, and in addition to genetic susceptibility, the major risk factors include overweight and obesity, unhealthy diet, sedentary life style and age. T2DM most commonly affects older adults, but incidences in children, adolescents and younger adults are increasing due to rising levels of obesity, physical inactivity and poor diet [1, 2, 11, 12].

Monogenic forms of diabetes are with about 1% of cases, far less common and include neonatal diabetes and maturity onset diabetes of the young (MODY). They are all autosomal-dominantly inherited forms of diabetes, each mediated by a single mutation in a key gene of the glucose metabolism. There are eleven recognized forms of MODY, of which MODY2 and MODY3 are the most common, accounting for up to 80% of MODY cases [13, 14]. The latter are caused by mutations in the *GCK* or *HNF1A* genes, respectively. MODY4 is caused by a heterozygous mutation in the *PDX1* gene (Pro63fsdelC) and only a single family carrying this mutation has been identified so far [13, 15, 16]. The mutation in the transactivation domain of PDX1 causes a frame shift and a premature stop and is thus a loss-of function mutation. Since PDX1 is a master regulator of pancreas and β-cell development, homozygous loss of PDX1 function leads to pancreatic agenesis [15].

1.2 The Endocrine Pancreas

The pancreas is the most important organ in the control of blood glucose levels and has an endocrine, as well as an exocrine function. The exocrine part accounts for more than 95% of the organ and is comprised of acinar and ductal cells that form an epithelial branching network. The exocrine acinar cells are arranged in clusters at the tips of the ductal network and secrete digestive enzymes that aid in the breakdown of proteins, carbohydrates, and lipids [17]. The ductal cells produce a bicarbonate-rich fluid and form the channel network, draining the enzymes secreted from acinar cells into the duodenum [18]. The endocrine part of the pancreas consists of clusters of different types of endocrine cells, called islets of Langerhans, which are scattered throughout the pancreas (Figure 1.2). The islets contain five endocrine cell types, α-, β-, δ-, ε- and PP-cells, each of which secretes a specific hormone. The most abundant cell type of the pancreatic islets are β-cells, which secrete insulin in response to elevated blood glucose levels. In turn, insulin promotes glucose clearance by triggering glucose uptake in the peripheral tissues, and glycogen

production in the liver. Conversely, α-cells secrete glucagon when glucose levels are descending in order to induce glycogenolysis and gluconeogenesis. δ-cells secrete somatostatin, which has an inhibitory effect on the secretion of insulin and glucagon from β- and α-cells, respectively [19]. Pancreatic polypeptide is secreted from PP-cells. It is an anorexigenic hormone, regulating food intake and gastric emptying, as well as digestive enzyme secretion from the exocrine pancreas [20–22]. The rarest cell type in the islet, the ε-cells, produce ghrelin, a hormone that is mainly secreted by endocrine cells in the gastic epithelium [23, 24]. Ghrelin increases energy intake and gastrointestinal motility and stimulates gastric acid secretion [24, 25]. In addition, ghrelin has been shown to suppress glucose-stimulated insulin secretion from β-cells, thus modulating systemic insulin levels and glycemia [26]. Together, the hormones secreted from the pancreatic islets, tightly control blood glucose levels and keep them in a narrow range of 4-6 mM [27].

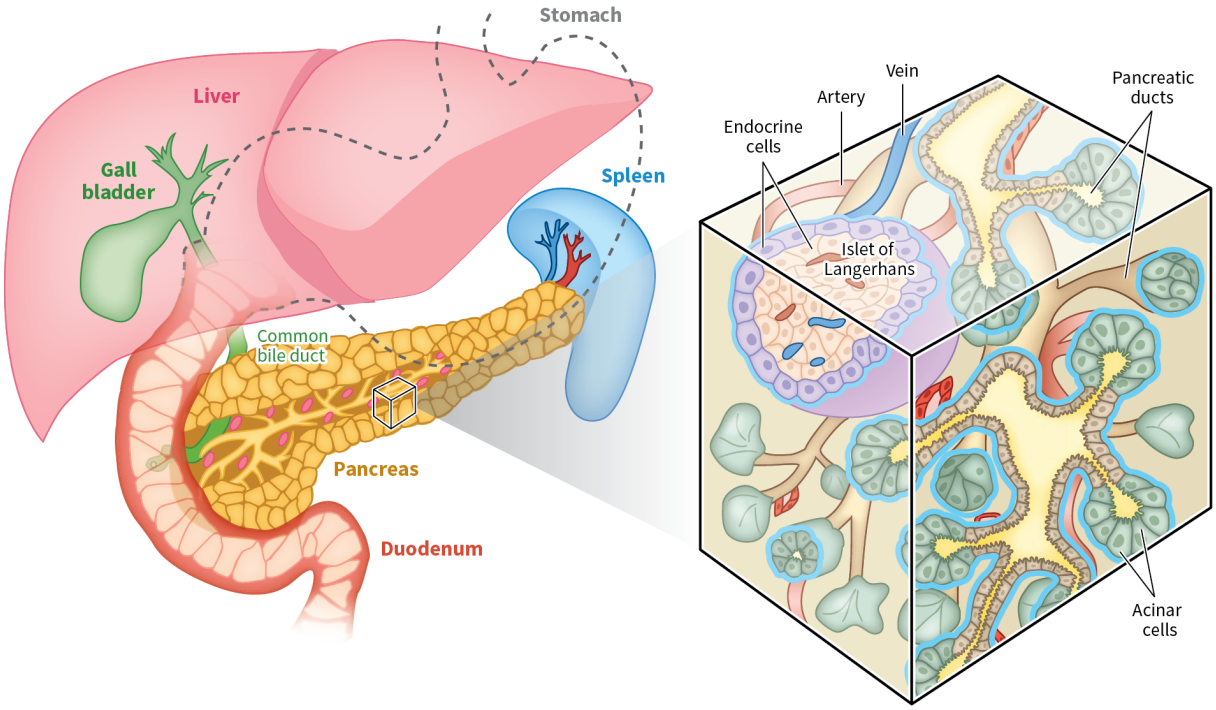

Figure 1.2: The Anatomy of the Pancreas. The pancreas is located behind the stomach and attached to the duodenum. The acinar cells secrete digestive enzymes that are transported to the duodenum via the ductal system. The islets of Langerhans contain different endocrine cells (α-, β-, δ-, ε- and PP-cells) that release hormones, regulating blood glucose levels and energy homeostasis. The pancreatic islets are interveined by a capillary network through which the hormones are transported into the blood stream. Figure modified, with permission, from [17].

1.3 Pancreas Development

All cell types of the pancreas, the acinar, ductal and endocrine cells, are derived from the same lineage of endodermal cells in the distal foregut. In the mouse, around embryonic day 8.5 (E8.5), two pancreatic regions on opposing sides of the foregut are specified [28], and approximately twelve hours later a dorsal and a ventral pancreatic bud are formed [29, 30]. During the first phase

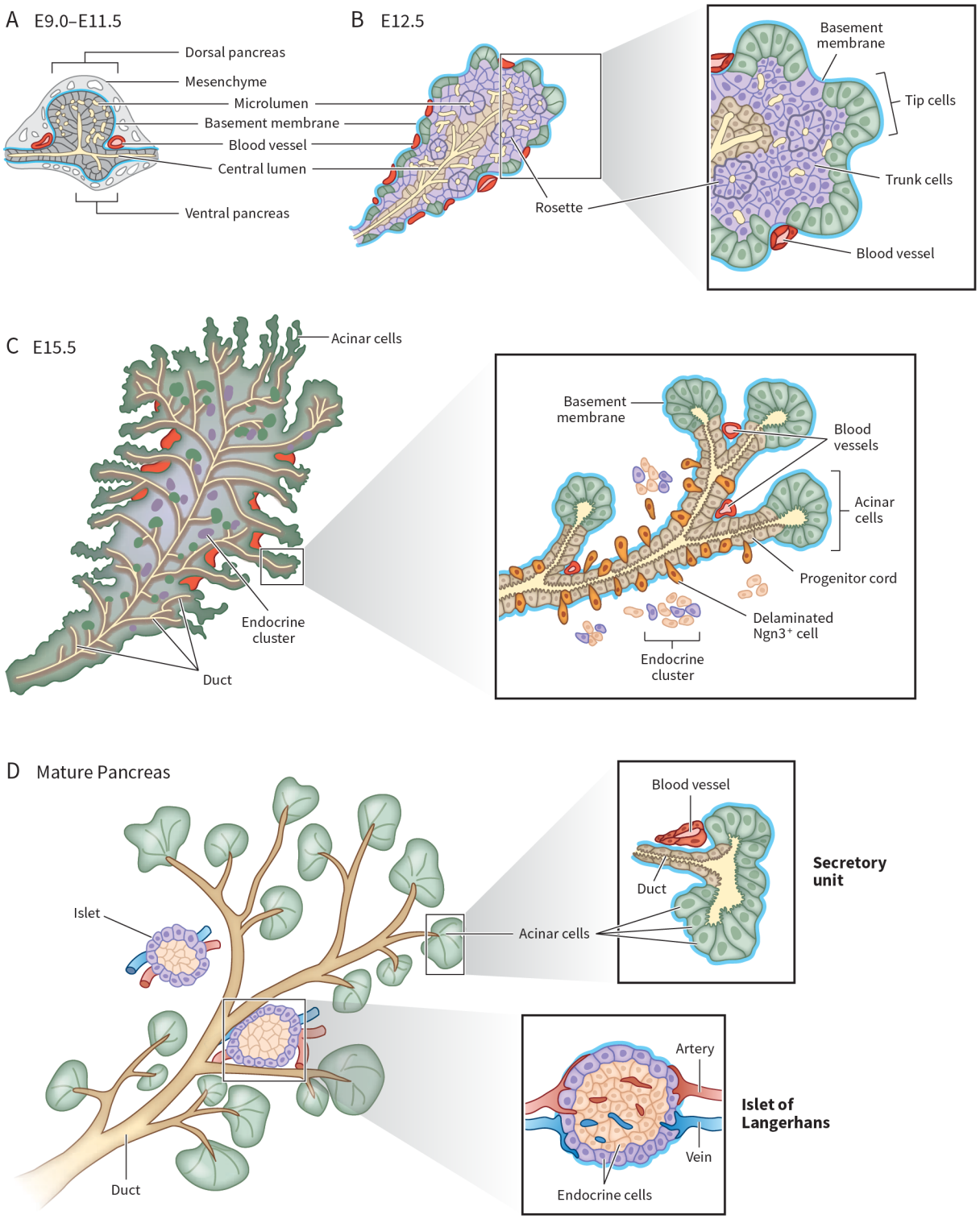

Figure 1.3: Stages of Pancreas Development. Caption on next page.

Figure 1.3: Stages of Pancreas Development. A: At ∼E9 the dorsal and ventral pancreatic buds are beginning to form. At around E10.5, MPCs become polarized and form microlumina. **B:** With the beginning of the secondary transition, tip–trunk patterning is established, while microlumina begin to fuse and form a central plexus region. **C:** The periphery of the developing pancreas is remodeling into ramified epithelium with developing acinar cells at the tips of the bipotent trunk domain that give rise to ductal exocrine and endocrine cells. Differentiating endocrine cells are beginning to form clusters. **D:** In the mature pancreas, the tips of the ductal network form a secretory unit of acinar cells, secreting digestive enzymes. The endocrine clusters have been compacted into islets of Langerhans that are innervated and vascularized to facilitate endocrine function. Figure modified, with permission, from [17].

(primary transition) of pancreas development, driven by signaling cues from the notochord and the surrounding mesenchyme and endothelium [31–33], the two buds, consisting of multi-potent pancreatic progenitors (MPCs), expand alongside the foregut tube and form a multilayered, non-polarized epithelium [17, 34, 35]. As the buds expand, the gut begins to rotate and directs the ventral bud towards the dorsal side. Around E12.5 the two aspects of the pancreas eventually fuse to form a single organ that branches further into the surrounding mesenchyme [29]. At around E10.5, individual cells in the inside of the buds become polarized, begin to rearrange and form microlumina [17, 35–37] (Figure 1.3 A).

During the following secondary transition, which begins around E12.5 the microlumina expand and start to fuse, forming a continuous luminal network. At this stage of branching morphogenesis, the pancreatic epithelium is divided into an expanding stratified core plexus, which is remodeling into ramified epithelium at the periphery [17, 35–37]. The inner plexus also serves as the niche for endocrine progenitor cells that are beginning to emerge. In the periphery, the epithelial branches segregate into tip and trunk domains that are allocated to either acinar or bipotent endocrine–duct progenitors, respectively [38]. While the tips continue to form a secretory unit, consisting of several pyramid-shaped exocrine cells, the bipotent cells of the trunk domain gives rise to ductal exocrine and endocrine cells [17, 34, 35] (Figure 1.3 B–C). In the mature ductal network, the main duct, which is connected to the duodenum via the common bile duct, branches into the entire organ through interlobular, intralobular and intercalated ducts that end in a secretory acinar unit. Besides their function to transport the pancreatic enzymes, secreted by acinar cells, ductal cells themselves also secrete a bicarbonate-rich fluid that helps neutralize gastric acids in the duodenum [39]. Also, during the secondary transition most of endocrine cell specification takes place from progenitors, which derive from the bipotent trunk epithelium [40, 41]. After their specification, endocrine progenitors differentiate and delaminate from the ductal epithelium [34, 42], to eventually form aggregates. These proto-islets expand by proliferation and become vascularized and innervated [43–46]. During this process, which continues postnatally, the endocrine cells mature and islets acquire a compact spherical shape. As result of β-cell proliferation, mature mouse islets consist of 60–80% β-cells [47], which are located at the core, covered by α-, δ- and PP cells. During early postnatal life, β-cells mature into their hormone-producing and glucose-responsive phenotype, a process that completes with the introduction of a carbohydrate-rich diet after weaning [48–52] (Figure 1.3 D).

1.4 Regulation of Pancreas Development

In the mouse, pancreatic development is first evident around E8.5 with the induction of *Pdx1* expression in the primitive gut tube [53, 54]. Twelve hours later, at E9, the pancreatic buds begin to form, and at E9.5 expression of *Ptf1a* is initiated [55]. The expression of *Pdx1* and *Ptf1a* is vital for pancreas formation, as homozygous loss-of-function of *Pdx1* or *Ptf1a* results in pancreatic agenesis. Interestingly, in these animals, the initial buds are still formed but fail to develop, suggesting a potential unidentified upstream regulator of pancreatic bud initiation [56–59]. Loss of *Pdx1* further leads to defects in gastro-duodenal development, as *Pdx1* is widely expressed in the foregut endoderm. During this stage of development, the expression of *Pdx1* and *Ptf1a* is regulated by several transcription factors (TFs), including Sox9, Hnf1β, Foxa1/2 and Cdx2 [60–63]. The exact mechanism by which these factors regulate *Pdx1* and *Ptf1a* expression is not well understood, but it is likely that their cooperative action together with downstream mediators of common signaling cascades is required. There is also evidence that Sox9 and Pdx1 act in a positive cross-regulatory loop and cooperatively govern the pancreatic lineage commitment [63]. Moreover, Sox9 is involved in the regulation of *Hnf1b*, *Onecut1* and *Foxa2*, all of which are expressed in MPCs, highlighting its important role in early pancreas development [62]. The expansion of the MPC pool is also dependent on the expression of several other TFs, such as Gata4/6, Onecut1, Hes1, Prox1 and Mnx1 [17, 34] (Figure 1.4 A). Proper expansion of the MPC pool is a key aspect, determining the size of the adult organ [64].

Induction and early development of the pancreas is governed by a number of signaling cues, originating from the notochord and surrounding mesenchyme and endothelium. However, these signaling cues are far less understood than the network of TFs they control. The notochord releases activin-βB and FGF2 that repress endodermal sonic hedgehog (Shh) signaling, thereby permitting pancreas induction [65]. Signaling from the mesenchyme via FGF10 and its receptor FGFR2 has a pro-proliferative effect on the MPC pool and aids the growth of the pancreatic buds. During bud formation, FGF10, FGFR2 and Sox9 form a feed-forward loop in which *Sox9* expression is positively regulated by FGF10, which in turn regulates the expression of *Fgfr2* to receive FGF10 signals. Perturbation of this signaling loop leads to loss of pancreatic identity and switching to a hepatic fate [66]. Another important pathway in early pancreatic growth is Notch signaling, which is involved in the regulation of the cell cycle and important TFs such as Sox9, and is required for maintenance and proliferation of MPCs. Disruption of the Notch signaling pathway by deletion of the effector RBP-jκ or the downstream target Hes1 leads to early arrest in pancreatic growth and premature endocrine differentiation [67–71].

Notch signaling is also essential for the segregation of the tip and trunk domains at the onset of the secondary transition [38]. Repression of Notch signaling results in excessive tip formation at the expense of the trunk domain, while constitutive activation of the pathway causes expansion of the trunk domain and prevents tip formation, which is at least in part due to the activation of Nkx6.1 [72–74]. The TFs Nkx6.1 and Ptf1a are the master regulators of the tip–trunk patterning and while they are co-expressed in MPCs, their expression becomes restricted to the trunk and tip compartment, respectively [74]. Both TFs act in a mutually repressive mode, where Nkx6.1 induces trunk formation by repressing Ptf1a and the tip commitment. Conversely, Ptf1a

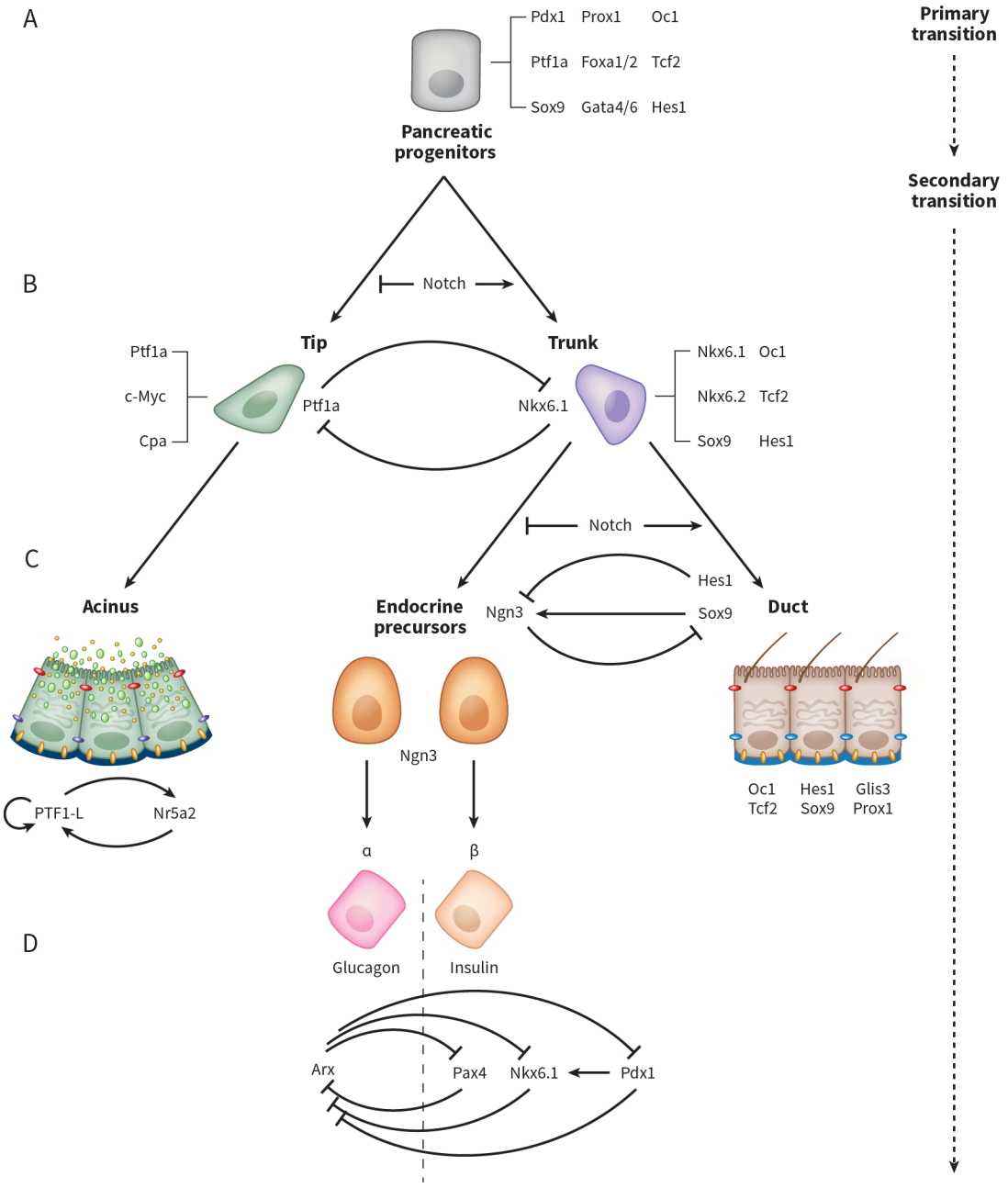

Figure 1.4: Regulation of Pancreas Development. A: During primary transition, multipotent pancreatic progenitors are marked by the expression of several genes, including *Pdx1*, *Ptf1a*, *Sox9*. **B:** Beginning with the secondary transition, tip and trunk domain are established. This process depends on active Notch signaling to induce a trunk fate, while inhibiting the tip fate via a regulatory loop involving Ptf1a and Nkx6.1. Tip and trunk cells can be distinguished by the expression of several marker genes. **C:** Cells of the tip domain become restricted to an acinar fate,established by the synergistic action of the PTF1-L complex and Nr5a2. In the trunk domain, bipotent progenitors give rise to ductal and endocrine cells in a process that is also mediated by Notch signaling activity. While high levels of Notch activity favor a ductal fate decision by repression of endocrine TFs via Hes1, low levels of Notch signaling allow Sox9 to activate the pro-endocrine TF Neurog3 which in turn represses Notch signaling and Sox9 to establish an endocrine cell fate. **D:** After endocrine specification, the different endocrine cell types of the pancreas are segregating. The decision between the α-cell and β-cell fate is regulated by the TFs Arx, which represses the pro-β-cell TFs Pxa4, Nkx6.1 and Pdx1, and Pax4 which inhibits Arx and induces a β-cell program. Figure modified, with permission, from [17].

represses Nkx6.1 and the trunk fate, while inducing the commitment to the tip domain [74]. The committed cells of the tip domain are then characterized by the expression of *Ptf1a, Nr5a2, Cpa1* and *Myc*, whereas trunk cells are marked by the expression of *Nkx6-1, Sox9, Hnf1b, Prox1, Hes1* and *Pdx1* [38, 74–78] (Figure 1.4 B).

The transition of the non-committed MPCs to either tip or trunk fates is the first step in a series of cell fate decisions in the pancreatic development. While the cells of the tip domain become restricted to an acinar fate through the synergistic action of Ptf1a and Rbp-jl, which form the PTF1-L complex, and Nr5a2 [79–83], the bipotent cells of the trunk compartment predominantly yield endocrine and ductal cells. During development, these cells can differentiate into $Sox9^+/Hnf1\beta^+$ duct cells or $Neurog3^+/Pdx1^+$ endocrine progenitors. Notch signaling plays an important role in this lineage decision as well. In the undifferentiated state, trunk cells express Sox9, Hnf1β, Onecut1, Glis3 and Hes1, all of which are regulators of the endocrine determining TF Neurog3 (alias Ngn3) [17]. While Sox9, Hnf1β, Onecut1 and Glis3 promote *Neurog3* expression, Hes1 effectively prevents activation of Neurog3 [67, 68, 84, 85]. Thus, while Notch signaling is active, expression of *Neurog3*, and thereby the endocrine fate, is repressed. Scattered cells in the trunk domain acquire lower levels of Notch activity, leading to reduced Hes1 levels. In these cells, Hes1 mediated inhibition is lacking, which allows Sox9, in cooperation with other TFs, to activate *Neurog3* expression. Subsequently, Neurog3, together with TFs such as Foxa2, activates its own expression and inhibits the transcription of *Sox9* [86–88]. After *Neurog3* expression is established, the endocrine committed cells repress endocrine commitment of neighboring cells by activation of Notch signaling, a process known as lateral inhibition [89] (Figure 1.4 C). Expression of the endocrine master regulator Neurog3 subsequently induces the expression of an endocrine gene program that includes *Neurod1, Insm1, Irx1/2, Rfx6 Pax4* and *Nkx2-2* [90]. While loss of Neurog3 activity results in a complete lack of endocrine cells, ectopic expression of *Neurog3* or its downstream target *Neurod1* under control of the *Pdx1* promoter leads to the formation of hormone-producing cells, highlighting the key function of Neurog3 in the endocrine formation process.

After specification of the endocrine progenitors, $Neurog3^+$ cells delaminate from the ductal epithelium and undergo a series of lineage decision and maturation steps in order to differentiate into the specific endocrine cell types of the adult pancreas. This is achieved through the stepwise activation of a complex regulatory network, governed by TFs that become progressively restricted to the specific endocrine cell types. A well studied example is the decision between α- versus β/δ-cell fates, which is controlled by the TFs Arx and Pax4, respectively. Initially co-expressed in $Neurog3^+$ progenitors, Arx and Pax4 are mutually repressive and their tight balance is lost during differentiation, directing the cell fate in either one of the two lineages [91, 92]. When Arx levels are increasing, Pax4 and other pro-β-cell TFs, such as Nkx6.1 and Pdx1 are inhibited and the α-cell fate is established in cooperation with a set of other TFs, including Pax6, Rfx6, Brn4, Foxa2 and Mafb [93]. In contrast, if the balance tips towards Pax4 expression, Arx is inhibited and cells differentiate into the β/δ-cell direction [94]. During this process, Nxk6.1 and Pdx1 are of particular importance, as both directly repress Arx [91, 92, 95–98] and as development progresses their expression becomes restricted to β-cells and to a subset of δ-cells, respectively. Moreover, during β-cell development, the expression levels of both, *Pdx1* and *Nkx6-1*, gradually

increase, suggesting a dose-dependent function of TFs (Figure 1.4 D). Another important TF during endocrine cell type specification is Nkx2.2, which is involved in both, the repression as well as the induction, of Neurod1 in Neurog3^{+} endocrine progenitors, to induce α-cell or β-cell fate, respectively [99]. This opposing function is likely mediated by cooperating cell type specific TFs. In differentiated endocrine cells, Nkx2.2 is restricted to β-cells, where it is involved in the maintenance of the β-cell identity by the repression of Arx and the α-cell program [100].

After endocrine cells have differentiated into the specific cell types, they undergo a maturation process during which they acquire their glucose-responsive and hormone-producing phenotype. In α- and β-cells this process is driven by two important TFs, MafA and MafB. After specification both cell types express *Mafb*, but in the course of differentiation its expression becomes restricted to α-cells, where it promotes maturation and α-cell identity [101, 102]. In β-cells, on the other hand, it is essential that the expression of *Mafb* is turned off and replaced by *Mafa* which is critical for β-cell maturation and identity [103]. The expression of *Mafa* is regulated by several TFs, including Neurod1, Nxk6.1, Nkx2.2, Foxa2, Rfx6 Pax6 and Glis3. Together with Pdx1, Nkx6.1 and Neurod1, MafA participates in the regulation of insulin expression [104–107]. Moreover, Pdx1 and Nkx2.2 are essential factors for β-cell maturation. Mature β-cells are then marked by the expression of *Ucn3*, *Gck* and *Slc2a2* (Glut2) [35, 108–110]

1.5 The Transcription Factor PDX1

Pancreatic and duodenal homeobox 1 (*Pdx1*) gene, formerly known as insulin promoter factor 1 (*Ipf1*), encodes a homeodomain containing TF and is part of the ParaHox cluster, located on chromosome 13q12.1 in humans and on the syntenic region on chromosome 5qG3 in mice [111, 112]. The gene consists of two exons and its promoter regions contains three highly conserved regulatory regions, which are located between 2800 bp and 1600 bp upstream of the transcription start site (TSS) and termed area I, II and III. In addition, there is a distal enhancer element, called area IV, which is located 6500–6050 bp upstream of the TSS. These regulatory elements facilitate binding of several TFs, including Hnf1β, Foxa1/2, Nkx2.2, Pax6, MafA as well as Pdx1 itself, all of which regulate cell type specific expression of Pdx1 [60, 113, 114]. In humans, PDX1 is a protein of 283 amino acids with an N-terminal transactivation domain, containing three conserved subdomains (A–C) and a central antennapedia-like homeodomain, facilitating DNA binding, as well as protein–protein interaction [115]. The homeodomain itself is comprised of three helices that are flanked by proline-rich sequences and contain a nuclear localization signal (NLS), as well as an antennapedia-like protein transduction domain (PTD), allowing Pdx1 to permeate into cells [116–118]. Moreover, the proline-rich region, N-terminal of the homeodomain, contains an antenna-type hexapeptide that is required for heterodimerization with the TF PBX1 [116, 119]. The C-terminal region of Pdx1 contains a SPOP binding site, interacting with the E3 ubiquitin ligase adaptor protein SPOP, which mediates PDX1 degradation [120, 121]. The C-terminal domain also participates in protein–protein interactions and is important for the full activity of PDX1, as mutations in this region are associated with T2DM and a reduced transactivation capacity [113, 122]. In addition, a PCIF1-interaction module (PCIF1-IM) has been discovered within the C-terminal region of PDX1. The interaction with PCIF1 inhibits PDX1 transactivation

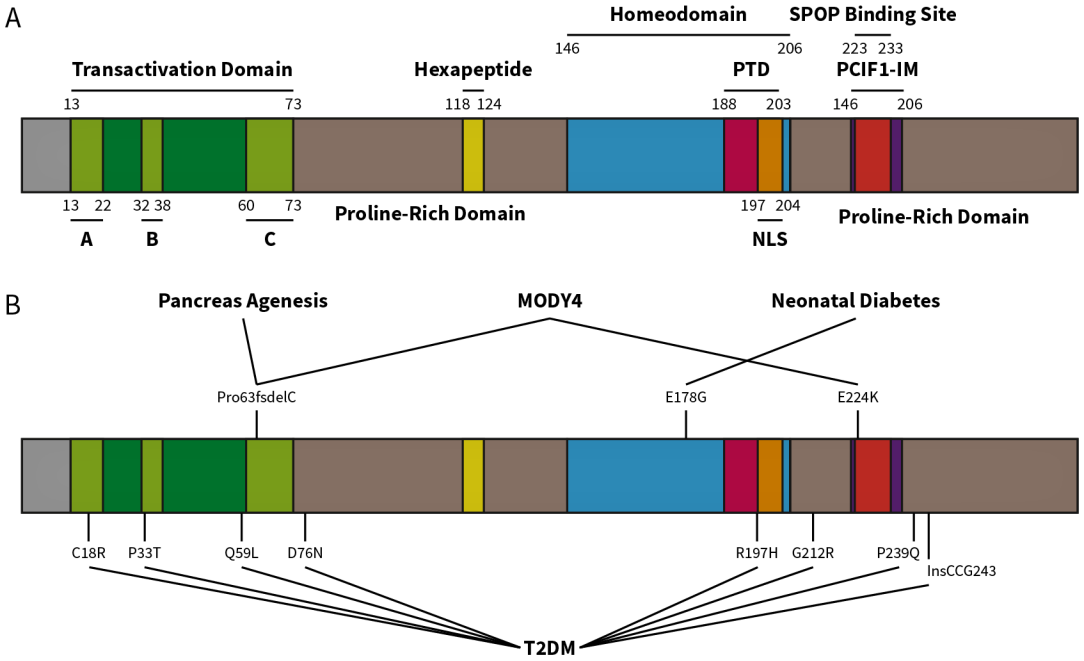

Figure 1.5: Domain Structure and Mutations of PDX1. A: Schematic representation of the domain structure of human PDX1 showing the N-terminal transactivation domain with its three subdomains, the N-terminal proline-rich domain, which contains a hexapeptide, facilitating heterodimerization with PBX1, the DNA binding homeodomain with both, the protein transduction domain (PTD) and the nuclear localization signal (NLS) and the C-terminal proline-rich domain, harboring the PCIF1 interaction module (PCIF1-IM) and the SPOP binding site, which mediates degradation. **B:** Localization of disease relevant mutations of PDX1 that are associated with MODY4, pancreatic agenesis, neonatal diabetes and T2DM.

activity, and mutation of the PCIF1-IM abolished the interaction with PCIF1 [123]. Moreover, several phosphorylation sites have been identified in the Pdx1 protein and phosphorylation of specific sites within the SPOP binding site have been reported to increased Pdx1 degradation [113, 124] (Figure 1.5 A).

During embryonic development, the first *Pdx1* expressing cells can be detected at E8.5. They are located in three domains on the dorsal and ventrolateral sides of the fore–midgut junction and give rise to the pancreatic buds and eventually the entire pancreas [41, 54, 58]. With progressing development, *Pdx1* expression broadens and is evident in the region that forms the stomach, pancreas and duodenum. During later stages of development, Pdx1 expression becomes gradually restricted to β- and δ-cells of the pancreatic islets, the Brunner's glands and the duodenal epithelium, as well as the pyloric glands of the stomach [41, 58, 125, 126]. In addition to its presence in the gastrointestinal tract, *Pdx1* expression is also apparent in the developing brain during neurogenesis [127, 128].

Functionally, Pdx1 is essential for the development of the pancreas, antral stomach and duodenum and the maintenance of the gastro-duodenal junction [56, 58, 126, 129]. In the developing pancreas, Pdx1 acts as master regulator and is a critical factor for the differentiation of all pancreatic lineages [41]. The homozygous loss of Pdx1 function has been shown to causes complete pancreatic agenesis [56]. Moreover, $Pdx1^{-/-}$ mice show defects in the entire posterior foregut regions with disturbances in the gastro-duodenal junction, loss of Brunner's glands and

perturbed enteroendocrine differentiation in the stomach and duodenum [58, 126, 130]. In contrast, heterozygous loss of Pdx1 does not lead to developmental defects, but adult mice become diabetic and β-cells are lost due to apoptosis [131–133]. Similarly, loss of function mutations in humans cause pancreatic agenesis when they occur homozygously or monogenic diabetes (MODY4) in heterozygous carriers [15, 16, 134, 135]. However, not all heterozygous *Pdx1* variants lead to monogenic diabetes. Instead, most mutations are associated with T2DM [110, 136–139] (Figure 1.5 B). In contrast to the rich body of research, which established the important role of Pdx1 in the development, function and homeostasis of the murine pancreas and β-cells, the role of human PDX1 within these processes, and how they may be affected by mutations, is far less understood.

1.6 Pancreas Development in Humans

Our understanding of human pancreas development is still very basic, since the access to human samples is limited. Therefore, much of what is known about pancreas development has been studied in animal models. While this allows the conduction of detailed studies using genetic and other tools not available in humans, animal models cannot be translated directly to human pancreas development, as some significant differences exist. In humans, initiation of the pancreatic endoderm occurs at the Carnegie stage (CS) 9 and a ventral and dorsal bud arises at CS13 [140]. However, unlike in mice, the developing human pancreas does not undergo a primary transition and pancreatic progenitors do not express NKX2.2 [141, 142]. Tip–trunk patterning is established by CS19 at ∼7 weeks post conception (wpc). During the following fetal period, at 8 wpc Neurog3^{+} endocrine progenitors are rapidly emerging and the first embryonic β-cells are beginning to appear [143–145]. Endocrine commitment, based on the expression of Neurog3, peaks from 10–14 wpc and is ending at 26–35 wpc [143]. Starting at 10 wpc, endocrine cells begin to form clusters, and developing islets, containing all endocrine cell types, can be found by ∼12–13 wpc [141, 146]. Unlike in rodent models, the morphology of human islets is changing during development. In the early stages ∼14 wpc, β-cells are arranged in the core of the islet, surrounded by peripheral α-cells. This structure is reminiscent of what is observed in murine and small adult human islets. However, by ∼21 wps, as islets grow, α-cells and β-cells become intermingled within the islet, which seems necessary for endocrine function [147]. The mature glucose-responsive phenotype of human endocrine cells is, as seen in mice, established during a phase of maturation in early postnatal life [148, 149].

1.7 Modelling Human Pancreas Development & Disease with Stem Cells

Detailed knowledge of pancreas development is key to understand pancreas-associated diseases, such as diabetes, so novel therapies for their treatment can be developed. Among the therapeutic routes currently investigated, are the replacement of lost and/or defective β-cells, and the stimulation of endogenous β-cell regeneration [150, 151]. To this end, human embryonic stem cells (hESC), as well as induced pluripotent stem cells (iPSCs) are valuable tools, providing

a promising approach to generate transplantable material, as they can be differentiated into β-like cells. In the ∼20 years that have passed since the first attempts to differentiate pancreatic cells from hESCs, there have been great advancements in the protocols to differentiate functional endocrine cells [152–155]. Today's protocols apply the current knowledge of pancreas development and recapitulate all developmental steps, from definitive endoderm (DE) to gut tube (GT), to posterior foregut (FG), pancreatic progenitors (PP), endocrine progenitors (EP) and eventually to β-like cells, by mimicking the signaling cues, active at each step (Figure 2.1 B). The resulting β-like cells are mostly mono-hormonal and glucose-responsive, although not at the level of primary β-cells [155–157]. In addition to their potential in β-cell-replacement therapy, stem cell differentiations are a useful tool to study the mechanisms of human pancreatic development and disease pathomechanisms, as they help to overcome the limited access to primary human material, as well as the species-specific differences, which are limiting the use of animal models [17, 151]. Monogenic forms of diabetes such as MODY, for example, are best studied in human *in vitro* systems, since animal models fail to recapitulate the human phenotypes [158, 159]. Particularly interesting, to that end, is the use of patient-specific iPSCs that provide a valuable source to study the mechanism leading to disease onset.

1.8 Intestinal Hormones in Glucose & Energy Homeostasis

It is well established that the pancreas plays an important role in maintaining the glucose and energy homeostasis of the body. However, the intestinal endocrine system, as the largest endocrine system of our body, is as important. In the intestinal epithelium, up to 20 different types enteroendocrine cells (EECs) secrete a multitude of hormones, which regulate various physiological responses, essential for the control of energy homeostasis in response to food intake [160–163]. Traditionally, EECs have been classified by the principal hormone they secrete: D-cells (somatostatin, Sst), enterochromaffin cells (ECs) (Serotonin, 5-HT), enterochromaffin-like (ECL) (histamine), G-cell (gastrin, Gast), I-cells (Cholecystokinin, Cck), K-cells (Gastric inhibitory peptide, Gip), L-cells (Glucagon-like peptide 1, Glp1), M-cells (motilin), N-cells (Neurotensin, Nts), S-cells (secretin, Sct) and X-cells (ghrelin, Ghrl) [161]. In recent years, it has been recognized that most EECs produce more than one hormone, making the classification significantly more complex [162, 163]. However, the letter code is well established and is still useful to describe a population of cells that shares the expression of a particular hormone. The secretion of most intestinal hormones rises after food ingestion. However, some hormones such as Sst, motilin and the orexigenic hormones Ghrl and insulin-like peptide 5 (Insl5) are released during the fasting state where they block gastric acid secretion (Sst) or promote food intake (ghrelin, motilin and Insl5). After ingestion of a meal, Sst and Ghrl levels quickly drop to reduce appetite and allow Gast and histamine-mediated acid secretion [161, 164]. Moreover, food intake triggers the release of a variety of hormones, regulating food digestion and absorption, as well as the disposition of absorbed nutrients and satiety. In the duodenum, K-cells secrete Gip, an incretin hormone that increases glucose-stimulated insulin release from pancreatic β-cells [165, 166]. As the food moves further, the release of other hormones, including Cck, secretin, Glp1/2 and Pyy is stimulated. Cck and secretin promote the release of bile acids from

the gallbladder (Cck) and digestive enzymes and bicarbonate from the exocrine pancreas that aid food digestion (Cck & Sct) [167, 168]. Cck, Glp1 and Pyy also exert an anorexigenic effect and mediate postprandial satiety and reduce gastric emptying [161, 167, 169–175]. Together with Gip, Glp1 also amplifies the release of insulin in response to elevated blood glucose levels. This function is known as incretin effect and accounts for up to 70% of the total postprandial insulin secretion, and is therefore an integral part in glucose homeostasis [176]. In the treatment of T2DM, the insulin-releasing effect of Glp1 is exploited and long-acting Glp1 analogues, as well as inhibitors of the dipeptidyl peptidase-4 (Dpp4), which mediates Glp1 degradation, are widely prescribed antidiabetic drugs [27, 161, 177]. Thus, the important role of intestinal hormones in the control of metabolism and energy homeostasis provide interesting therapeutic routes for the treatment of metabolic syndrome or T2DM, and have generated great interest in EEC research.

1.9 Architecture & Cellular Composition of the Intestinal Epithelium

The intestinal epithelium is organized in a crypt and villus structure made of millions of crypt–villus units that maximize the absorptive surface. Each unit consists of a finger-like protrusion of the epithelium, the villus, surrounded by multiple small invaginations into the submucosa, the crypts of Lieberkühn [178]. The villi of the small intestine contain postmitotic, mature cells that serve as a barrier to protect the organism and facilitate nutrient absorption. Underneath the epithelial layer, a network of capillaries and lymph vessels transport the absorbed nutrients into the body. Owing to the direct contact with the luminal content, the villus epithelium is exposed to high levels of mechanical and chemical stress. To cope with these harmful conditions, the intestinal epithelium is constantly renewing. The fast turnover leads to very short lifetimes of mature cells of only 3–5 days, thereby effectively minimizing the exposure of individual cells to their hazardous ambiance [179]. The crypts provide a protected environment, harboring the intestinal stem cells (ISCs) and the rapidly proliferating progenitor cells (transit amplifying cells, TA). The ISCs that are continually dividing, fuel the high turnover, necessary to maintain the epithelial function. They divide approximately once per day to give rise to the highly proliferative TA progenitors cells [180]. TA cells divide 2–3 times and gradually commit to the enterocyte lineage, while being pushed out of the crypt and along the crypt–villus axis [179, 181]. After 3–5 days, when the cells reach the tip of the villus, they are shed into the gut lumen through a complex process that maintains the integrity of the epithelial barrier [182] (Figure 1.6).

The intestinal epithelium contains 6 different mature types of cells, each of which serving a specific function. With up to 80% of the epithelial cells, enterocytes (Ent) constitute the largest cell fraction. They represent the absorptive lineage and are primarily responsible for the uptake of water, sugar, lipids, peptides, vitamins and ions, and also reabsorb unconjugated bile acids [184]. Microfold (M) cells are specialized absorptive cells, covering the Peyer's patches. They sample the content of the intestinal lumen and pass antigens to the underlying lymphoid follicles of the Peyer's patches, which contain high numbers of B and T cells [185]. The secretory goblet cells (GCs) produce a protective mucus, coating the epithelial lining. The mucus layer provides lubrication and acts as the first line of defense against physical and chemical injury.

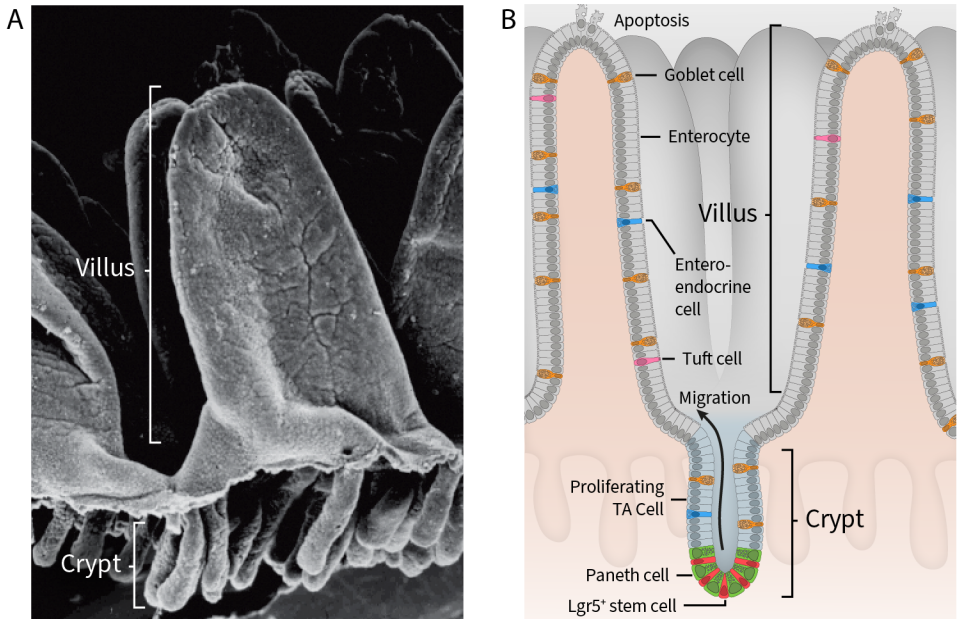

Figure 1.6: Architecture of the Intestinal Epithelium. A: Scanning electron micrograph showing the structural organization of the intestinal epithelium into crypt and villus units. **B:** Illustration depicting the organization of the intestinal epithelium. At the crypt base, ISCs are located between PCs, and constantly devide to give rise highly proliferative TA progenitor cells. The latter divide 2–3 times, differentiate into mature cells and migrate upwards along the crypt–villus axis. Cell that reach tip of the villus undergo apoptosis and are shed into the intestinal lumen. Figure A modified, with permission, from [181, 183].

Among the cells of the secretory lineage, GCs are the most frequent and recently, sentinel goblet cells, a specialized subtype that is found at the entrance of colonic crypts has been described. They sample their environment by endocytosis and secrete the mucus protein mucin 2 (Muc2) upon detection of a bacterial threat. They also stimulate mucus secretion from neighboring GCs to form a mucus barrier that effectively shields the crypt from pathogens and the digestive process above [179, 186]. In contrast to the GCs, tuft cells (TCs) are the rarest cell type and constitute less than 0.4% of the intestinal epithelium [187]. Recent evidence has shown that TCs mediate a type 2 immune response, which is induced by infection with large metazoan parasites such as helminths (parasitic worms), or the exposure to allergens. The immune response causes a number of physiological reactions, including goblet cell hyperplasia and smooth muscle hypercontractility [188, 189]. Upon helminth infection, TCs secrete interleukin (IL) 25 and activate tissue-resident group 2 innate lymphoid cells (ILC2) and recruited type 2 helper T cells (T_H2), which in turn release IL-4 and IL-13 and trigger a feed-forward loop that increases differentiation of GCs and TCs from ISCs, as well as TC proliferation [188, 189]. EECs also belong to the secretory lineage. They are scattered throughout the gastrointestinal tract and represent less than 1% of the intestinal epithelial cells [190]. EECs constantly monitor the composition of the luminal content and release a multitude of hormones in response to the nutrients they are sensing. As described above, the released hormones exert diverse functions, ranging from food intake (appetite and satiety), to intestinal motility and the release of digestive enzymes to the secretion of insulin from pancreatic β-cells [160, 161]. Like most other cell types of the intestinal epithelium, EECs migrate along crypt–villus axis, however at significantly lower

speed. It has recently been shown that EECs exit the crypt ∼60–80 hours after their specification, suggesting a significantly longer life span of these cells, and a mechanism that allows EECs to escape the migration front [191]. Moreover, EECs display hormonal plasticity and change their hormone production as they move along the BMP signalling gradient, established down the crypt–villus axis [191, 192]. The fourth secretory cell type are the Paneth cells (PCs). In contrast to all other mature cell types of the intestinal epithelium, PCs are not found on the villus surface, as they do not migrate along the crypt–villus axis. Instead, they move downwards to the base of the crypt in a specific migration process that depends on ephrinB2, ephrinB3 and phoshoinositide 3-kinase (PI3K) signaling [193, 194]. PCs have a life span of up to 8 weeks and secrete several antimicrobial peptides such as lysozyme, α-defensins and phospholipase A2, which are essential for the protection of ISCs, as well as for the mucosal immunity along the entire intestine [195–197]. In addition to their protective function, PCs constitute the primary niche for ISCs.

At the crypt base, the $Lgr5^+$ ISCs, also known as crypt base columnar (CBC) cells, are encompassed by PCs that stimulate Wnt, Notch and EGF signaling pathways, which are essential for ISC maintenance [198, 199]. The direct contact with their niche cells is required for stem cell maintenance and thus, every ISC is in contact with at least one PC. However, not only PCs provide niche signals for ISCs, but also the surrounding extra cellular matrix (ECM) and the stromal cells embedded within the ECM provide short range signals, maintaining their stemness [200]. As such, several mesenchymal cell populations have been identified that supply Wnt ligands, R-spondins, BMPs and BMP inhibitors [201–204]. ISCs that loose contact to their niche begin to differentiate into one of the intestinal lineages. The number of available PCs contacts therefore limits the size of the niche, which can accommodate on average 15 ISCs [179]. The limited niche size also promotes competition of ISCs for niche contacts where dividing ISCs push each other out of the stem cell niche, leading to a stochastic clonality of individual crypts [205, 206]. The dynamic of this competition ensures that even subtle differences in the fitness of ISCs have a strong impact on their chances to remain in the stem cell zone [179]. Thus, ISC competition provides a potent mechanism, inhibiting the formation of intestinal cancers, as tumor founding cells have to acquire the ability to proliferate independently of the niche signals, before they loose niche contact [207, 208].

In addition to the CBC stem cells, long-term lineage tracing studies have identified a candidate stem cell population of long-lived label-retaining cells that were predominantly found at the +4 position, directly above the PC zone [209–211]. The expression of several marker genes (*Bmi1* [212], *Tert* [213, 214], *Hopx* [215] and *Lrig1* [216]) have been reported to be specific for +4 cells and genetic lineage tracing experiments have shown that cells, expressing these markers are capable of regenerating the intestinal epithelium after injury [212–216]. Thus, the +4 stem cell is considered a quiescent and radiation resistant reserve stem cell that is able to reconstitute the pool of cycling ISCs (CBCs) as circumstances require [217]. However, in recent years, the concept of dedicated reserve stem cells has been questioned, as the intestinal epithelium is increasingly recognized as a highly plastic tissue, rendering intestinal regeneration as a complex and dynamic process that might involve any type of progenitor or even mature cells [179, 218–221].

A possible mechanism underlying the high plasticity of the intestinal epithelium has been revealed by studies comparing the epigenetic landscape of ISCs to progenitors and differentiated cells. On the level of histone modifications and DNA methylation, the chromatin configuration is highly similar during differentiation and in a generally permissive and open state, with lineage-specific genes being primed for transcription already in the ISC population [222–226]. Consequently, such an enhancer configuration allows a high degree of plasticity and the specification of the lineages is largely dependent on the presence or absence of lineage determining TFs at key regulatory elements.

1.10 Regulatory Pathways in the Intestinal Epithelium

The complex, dynamic processes, shaping the intestinal epithelium are governed by several key signaling pathways, including Wnt, EGF, Notch and BMP signaling (Figure 1.7). The Wnt signaling pathway is the main driver of proliferation and ISC maintenance in the crypt. Consequently, ISCs are completely lost when Wnt signaling is inactivated by the loss of Tcf7l2 (alias Tcf4), the main effector of canonical Wnt signaling [227, 228]. On the contrary, overactivation of the pathway is associated with the development of intestinal tumors [229–231]. Owing to their poor solubility, Wnt ligands act as short range signal, and most of the signaling activity in the crypt is mediated by direct cell–cell contacts. In the crypt, PCs provide Wnt3 that bind to the Frizzled receptors on the surface of neighboring ISCs. Differentiating ISCs that leave the stem cell zone then transport the surface-bound Wnt ligands away from the crypt bottom and establish a Wnt gradient, as the amount of Wnts on the cell surface is halved with every cell division [232, 233]. This mechanism couples proliferation in the crypt to its main regulator and establishes a dynamic feedback loop. If proliferation of ISCs is high, Wnt ligands are quickly diluted as many cells divide and move up the crypt, leading to a depletion of the signal at the crypt base and thus a reduction of ISC proliferation. On the other hand, if the proliferation of ISCs is low, Wnt ligands are not transported upwards and accumulate at the bottom of the crypt and promote ISC division or the dedifferentiation of progenitor cells. Thus, the gradient established through surface-bound Wnt ligands is an effective mechanism to suppress tumors and to promote regeneration upon loss of ISCs. In addition to the Wnt ligands produced by PCs, the cells from the underlying mesenchyme are an essential supplementary source of Wnt signals [201–204]. R-Spondins, a family of secreted, readily soluble proteins are another important component of the Wnt pathway [234–236]. In the crypt, they bind to the Lgr4 and Lgr5 receptors, expressed by ISCs, and potentiate Wnt signaling activity by sequestering Rnf43/Znrf3 mediated degradation of Frizzled receptors [237, 238]. It has been shown that R-Spondins are required for ISC maintenance, as the knockout of Lrg4/5 in the intestinal epithelium causes deprivation of ISCs, which resembles the loss of Wnt signaling [234].

Epidermal growth factor (EGF) signaling is another important pathway in the intestine. ISCs express the EGF receptor Erbb1, which is activated via its ligands EGF and transforming growth factor-α (TGFα), produced by neighboring PCs [198]. EGF activity is linked to ISC proliferation, and overactivation of the pathway in certain oncogenic mutants has been shown to increase ISC proliferation and to give mutated ISCs an advantage in the competition with other ISCs in the

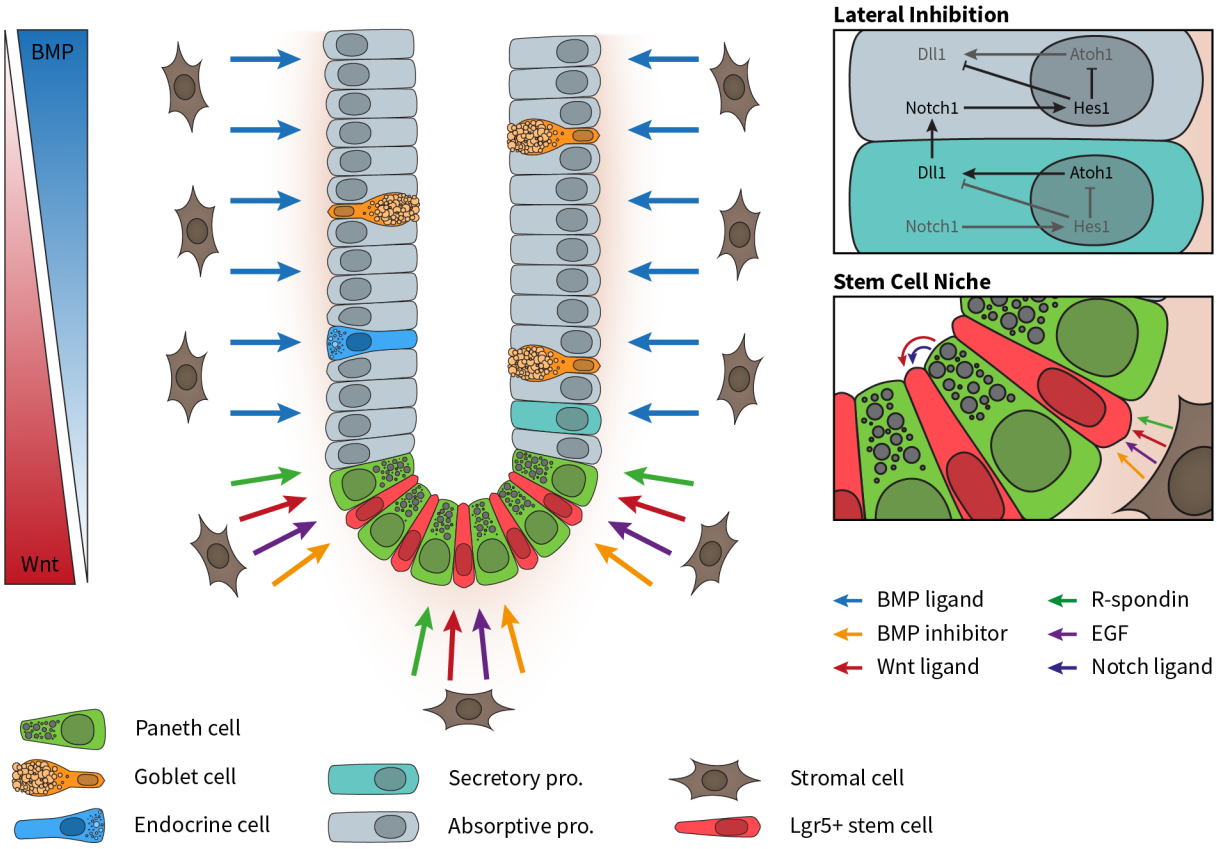

Figure 1.7: Main Signaling Pathways in the Crypt. Wnt and BMP lingands establish two opposing gradients along the crypt–villus axis. At the crypt bottom, PCs and stromal cells surrounding the crypt secrete Wnt ligands, which are essential for ISC maintenance. Wnt signals are of very short range, as Wnt ligands have a poor solubility, which results the formation of Wnt signaling gradient that quickly falls off with increasing distance to the crypt bottom. On the other hand, BMP ligands are secreted from the intercrypt and intervillus mesenchyme. BMP signals counteract Wnt signals and promote differentiation. To prevent BMP activity in the stem cell zone, BMP inhibitors are secreted from mesenchymal cells underlying the crypt, and form a gradient of increasing BMP activity along the crypt–villus axis. The mesenchyme surrounding the base of the crypt also provides R-spondins, which potentiate Wnt signaling, and EGF ligands, both of which are important for ISC maintenance. Thus, together with the PCs, which provide Notch and Wnt signals, the stromal cells around the crypt bottom form the ISC niche. In addition, Notch is an important signaling pathway in the intestinal epithelium, since it promotes ISC stemness and is involved in secretory cell differentiation. Via lateral inhibition, the progenitors of the secretory lineages, which are characterized by high levels of the Notch ligand Dll1, send signals to neighboring cells to inhibit the secretory cell fate.

crypt [208]. In order to prevent neoplastic growth, ISCs tightly control the pathway's activity by the expression of the negative regulator Lrig1. Mice, deficient of *Lrig1*, have enlarged intestines as a result of excessive crypt expansion [239]. On the other hand, blockage of EGF signaling in intestinal organoid cultures stops organoid growth and inhibits ISC proliferation, however, without disturbing ISC identity [240]. Thus, EGF signaling is important for the regulation of ISC proliferation, but unlike Wnt or Notch signaling it is not required for ISC maintenance.

Notch signaling regulates several aspects of intestinal tissue homeostasis. Direct cell–cell contacts are required for pathway activation, as both, the Notch receptor and the ligands (e.g. Dll1 and Dll4) are membrane bound proteins. Activation of the pathway by a neighboring cell

induces the expression of Hes1, the main effector of the pathway. Hes1 in turn blocks the expression of Atoh1, a transcription factor essential for the formation of secretory cells in the intestinal epithelium. The inhibition of Atoh1 also downregulates the expression of Notch ligands [241]. Thus, a signal-sending cell prevents all neighboring signal-receiving cells from sending a signal themselves, a mechanism known as lateral inhibition. In the intestine, Notch signaling activity blocks the differentiation of secretory cells and therefore determines the ratio of absorptive to secretory progenitors, and helps to maintain the stemness of ISCs at the crypt base. Hence, active Notch signaling is observed in ISCs and TA cells just above the stem cell zone, activated by PCs (Dll1 and Dll4) and secretory progenitors (Dll1) respectively [198, 218, 242]. The importance of Notch signaling in the specification of the absorptive and secretory lineage is also highlighted by loss-of-function experiments. If Notch activity is blocked, all progenitor cells adopt a secretory fate [243, 244]. Likewise, overexpression of Atoh1, forces differentiation of secretory cells, while loss of Atoh1 results in the complete absence of secretory cells [245, 246]. Moreover, the specific activation of Notch signaling in Neurog3$^+$ endocrine progenitors drives these cells to differentiate into GCs and enterocytes [247]. This indicates that in addition to the role in the initial absorptive versus secretory lineage decision, Notch signaling is also involved in the specification of the goblet and endocrine lineage.

BMP signaling, the fourth major signaling pathway in the intestinal epithelium, opposes the function of Wnt, Notch and EGF by counteracting proliferation and promoting differentiation. BMPs belong to the TGFβ superfamily of ligands and bind to the complex of type I and type II BMP receptors. Upon receptor activation, Smad1, 5 or 8 are phosphorylated, bind to the common Smad4 and translocate to the nucleus, where they engage in gene regulation [248]. BMP2 and BMP4 are the main ligands in the intestine and both are secreted from intercrypt and intervillus mesenchymal cells [249, 250]. In conjunction with BMP antagonists such as Noggin, Gremlin 1/2 or Chordin-like 1, which are secreted by mesenchymal cells beneath the crypts, a gradient of BMP activity is formed along the crypt–villus axis [204, 251, 252]. If Bmpr1a, the main BMP receptor in the intestine is conditionally deleted, or BMP inhibitors are ectopically overexpressed, the ISC and TA compartments expand and benign polyps are forming [250, 251, 253]. Thus, active BMP signaling has an inhibitory effect on proliferation and supports cell differentiation. In order to maintain the proliferative environment of the ISC zone, BMP antagonists are required to balance BMP signaling activity. As a result, decreasing Wnt and increasing BMP activity enable the gradual progression of differentiation along the crypt–villus axis.

1.11 Specification of the Intestinal Cell Types

All cell types of the intestinal epithelium are derived from the ISCs at the crypt bottom (Figure 1.8). Once they leave the stem cell zone, they start to differentiate and, guided by Notch signaling activity, undergo the first fate choice between the absorptive and secretory lineage. The absorptive fate is defined by active Notch signaling, which inhibits the expression of the secretory TF Atoh1, and is the prevalent outcome of this decision, since it is determined by lateral inhibition [241]. In mice, deletion of *Atoh1* mimics the outcome of constitutively active

Notch signaling and all progenitor cells adopt an absorptive enterocyte fate [246, 254, 255]. Thus, ISCs that maintain contact to Dll$^+$ cells after they leave the high-Wnt setting of the stem cell zone differentiate into TA enterocyte progenitors. After 2–3 rounds of cell division, enterocyte progenitors leave the TA zone and become fully differentiated enterocytes as they experience higher BMP signaling activity [179]. Unlike the other intestinal cell types, specialized absorptive M cells, found on the surface of the Peyer's patches, are not necessarily specified within the crypt. Instead, they are generated, when enterocytes or their progenitors receive RANKL signals, sent by the subepithelial stromal cells, which cover the Peyer's patches. RANKL-mediated activation of RANK then initiates the expression of the TF Spib, which is essential for the formation of M cells [256–258].

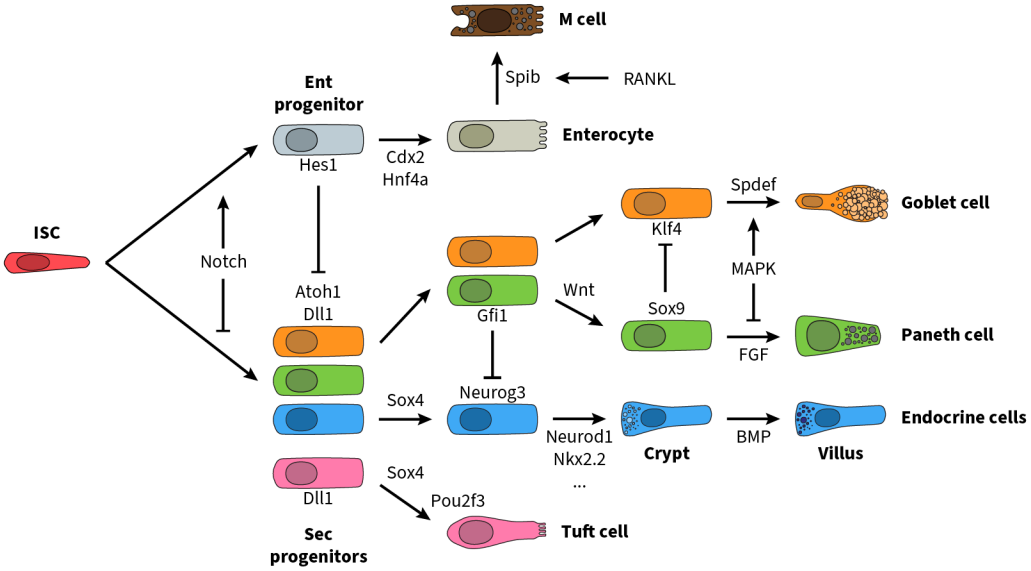

Figure 1.8: Lineage Decisions in the Intestinal Epithelium. Schematic representation of the differentiation paths from ISCs to the mature cell types of the intestinal epithelium and the main signals and TFs involved.

Tuft cells, which mediate the mucosal immune response, are considered to belong to the secretory lineage. Their specification depends on the inhibition of Notch signaling but, in contrast to the other secretory cell types, they seem to be specified independently of the secretory TF Atoh1 [187, 254, 259]. Instead, Sox4 has been shown to be an important regulator of TC differentiation. Recently, a population of *Sox4*$^+$/*Atoh1*$^-$ cells has been identified as TC progenitors and the conditional deletion of *Sox4* leads to decreased numbers of TCs [259]. Another TF, essential for TC differentiation is Pou2f3, which is specifically expressed in TC. It is required for TC differentiation and deletion of *Pou2f3* completely abolishes TC formation [189]. TC numbers are also regulated in response to parasite infection. The immune reaction triggered by TCs induces the release of IL-4 and IL-13 from ILC2 and T_H2 cells, which activate *Sox4* expression and increase TC numbers by inducing differentiation from ISCs, as well as TC proliferation [188, 189, 259].

Goblet cells are the default differentiation path for cells which have left the high-Wnt environment of the ISC zone and turned off Notch signaling. Upon genetic or pharmacological

inhibition of Notch activity, or the forced expression of *Atoh1*, all proliferative progenitors in the crypt convert to postmitotic goblet cells [243–245]. Next to Notch-mediated lateral inhibition, Shp2-mediated mitogen-activated protein kinase (MAPK) signaling is important for decision between the GC and PC fate [260]. Conditional ablation of *Shp2* in the intestinal epithelium results in reduced MAPK signaling activity and lower GC numbers, while promoting PC development. On the other hand, activation of Mek1 can rescue the *Shp2* deletion phenotype and promotes GC generation at the expense of PCs [260]. In addition, the differentiation of GCs depends on the TF Spdef, which is demonstrated by the accumulation of immature secretory progenitors in *Spdef*$^{-/-}$ mice [261, 262]. GCs are also targeted by the TC-mediated type 2 mucosal immune response. The release of IL-4 and IL-13 upon parasite infection also causes GC hyperplasia, however, via an unknown mechanism [188, 189].

The differentiation of Paneth cells, depends on the expression of the Wnt target gene *Sox9* and thus, the presence of high Wnt signals [199, 263–265]. In the case of PC deletion, ISCs lose their Notch signals and start to differentiate. In the high-Wnt environment at the crypt base, the PC lineage is the favored differentiation path. When the first PCs are formed, they provide Notch ligands and other niche factors and maintain stemness in neighboring cells [241]. Next to Wnt signaling, fibroblast growth factor (FGF) and mitogen-activated protein kinase (MAPK) signaling are important for the formation of PCs. The deletion of the FGF receptor gene *Fgfr3* results in a strong reduction in PC numbers. Conversely, the deletion of *Shp2* reduces MAPK signaling and increases the numbers of PCs, at the expense of GCs [260]. Consistently, the activation of Mek1 reduces the numbers of PCs [260].

Enteroendocrine cells belong to the secretory linage and their differentiation is also dependent on the inhibition of Notch signaling and the concomitant expression of *Atoh1* [246, 254, 255]. In addition, EEC formation depends on blocked Wnt and MAPK signaling, to prevent PC and GC differentiation, respectively [240]. Moreover, Sox4 has recently been identified as important factor for EEC differentiation, as reduced numbers of EECs were detected upon the conditional deletion of *Sox4* in the intestinal epithelium [259]. The main driver of EEC differentiation, Neurog3, is activated downstream of Atoh1 and Sox4 and is transiently expressed during a narrow window in EEC differentiation. While the knockout of *Neurog3* leads to a complete loss of all EECs [266], the forced expression of *Neurog3* in the intestine introduces a strong bias, favoring the differentiation of EECs on the expense of GCs, but does not significantly alter the overall intestinal morphology [267]. Following Neurog3 expression, several endocrine TFs, including Neurod1, Arx, Isl1, Pdx1, Nkx6.1 and Nkx2.2 are switched on and establish the endocrine identity [268–270]. In addition, a recent study identified a large panel of putative EEC regulators and showed that the TFs Rfx3, Tox3, Runx1t1, Myt1 and Zcchc12 are important for EEC differentiation [191]. Interestingly many of the genes expressed during EEC differentiation are also important for the development of pancreatic endocrine cells. As described above, EECs are a heterogeneous group of up to 20 different subtypes, each of which defined by a specific combination of TFs. However, the exact mechanisms, which define the different subtypes are not fully understood, and only a few regulators are known. For example, Pax4 and Arx are involved in the specification of the EC, D- and S-cell fates versus other EEC fates [191, 268]. As observed in the pancreatic endocrine cells, also in EEC progenitors, Pax4 and Arx are mutually

repressive, and subtypes are specified as the balance between the two TFs tips in one direction. When *Pax4* expression dominates, cells are directed towards an EC, D- or S-cell fate and, vice versa, when the balance tips towards *Arx* expression, cells adopt a non-EC/D/S fate [191, 268, 271]. Another TF that is involved in the differentiation of certain EEC subtypes is Foxa2. The combined knockout of *Foxa1* and *Foxa2* in the intestinal epithelium results in the complete loss of Glp-1/2 producing L-cells and reduced numbers of Sst expressing D-cells, as well as Pyy producing L-cells. In addition, GCs were defective and also reduced in numbers [272]. Recently, BMP signaling has been shown to regulate hormonal plasticity in EECs [192]. Individual EECs change their hormone expression profile, and express certain hormones preferentially in either crypts (e.g. Glp1) or villi (e.g. Sct). These changes in hormone production are mediated by the BMP signaling gradient along the crypt–villus axis. As EECs migrate towards the villus tips, increased signaling via BMP4 induces the switch in hormone expression of individual EECs [192].

Based on the cellular function and the principles of their differentiation, intestinal cell types are separated into the absorptive (enterocytes and M-cells) and secretory (EECs, GCs, PCs, TCs) lineages. The traditional model of cell differentiation proposes the existence of an absorptive progenitor, i.e. the TA cells, and a multipotent secretory progenitor that can generate all four secretory cell types. This secretory progenitor is specified via Notch-mediated lateral inhibition and is characterized by the expression of *Atoh1* and high levels of *Dll1* [179, 218, 273, 274]. However, the potency of secretory progenitors is controversial and the exact mechanisms that drive the differentiation of the individual cell types are not yet fully understood [191, 218, 259–261, 264, 273, 275–278]. In recent years, different models have been proposed, including multipotent [218, 273, 278] and bipotent [259, 276, 277] progenitors, or the direct specification from ISCs [179, 221]. Work form our group shows that Wnt/planar cell polarity (PCP) signaling primes $Lgr5^+$ ISCs to directly differentiate into either PCs or EECs and suggests that Wnt/PCP signaling precedes Notch-mediated lateral inhibition [279]. Thus, in our current model, all intestinal cell types are directly specified from ISCs. This does not contradict the mechanisms of cell type specification and differentiation described above but suggests that under normal conditions, a primed ISC is determined to differentiate into a specific cell type, even though it experiences common signaling cues (i.e. Notch and Wnt signaling) and goes through molecularly similar stages (i.e. $Dll1^+$/$Atoh1^+$).

1.12 The FoxA Pioneer Transcription Factors

Foxa2 is a member of the forkhead (Fox) family of TFs, which are evolutionary conserved and found from yeast to humans [280]. The *forkhead* (*fkhd*) gene was originally identified in *D. melanogaster* since mutations cause defects in the head fold involution, leading to a characteristic forked head appearance in adult flies [281]. In human and mouse, 44 different Fox genes have been recognized and are categorized into the subclasses A to S based on sequence similarity [282–284]. All Fox proteins share a highly conserved winged-helix or forkhead DNA-binding domain that is composed of three α-helices arranged in a helix-turn-helix core at the N-terminus, and two antiparallel β-strands with two less conserved winged loops at the C-terminus [285, 286]. DNA recognition is mostly mediated by the third helix, which is inserted into the major groove

of the DNA double helix. Owing the high conservation of the DNA-binding domain, most Fox TFs bind the canonical forkhead consensus sequence [280].

The three members of the FoxA subclass, Foxa1, Foxa2 and Foxa3, have been identified in the rat liver and were therefore initially termed hepatocyte nuclear factors 3 (HNF3) α, β and γ [287]. Of all the mammalian Fox genes, *Foxa2* shares the highest degree of homology with the originally discovered *Drosophila fkhd* gene [280]. The Foxa2 protein is comprised of two transactivation domains at the N- and C-terminus, respectively, and the conserved winged-helix DNA binding domain, which is flanked by two nuclear localization signals on both sides [288]. The three members are 95% identical within the forkhead domain, but the identity is less in the rest of the protein [288]. Interestingly, the structure of the DNA-binding domain of the FoxA TFs has been shown to resemble the structure of linker histones H1 and H5. This is an important feature of these TFs, as it allows them to access compact chromatin and induce local chromatin decompaction, which enables the binding of other TFs [289–292]. Moreover, the C-terminal region is able to interact with the core histones H3 and H4 [289].

The expression patterns of Foxa1, 2 and 3 are distinct in different tissues during development and adulthood but are often overlapping [288]. At E6.5, *Foxa2* is the first to be expressed in the primitive streak and node of the developing embryo. Shortly after, at E7.5 expression is detected in the mesoderm and definitive endoderm, and continues to be expressed in endoderm-derived organs, such as pancreas, liver, lung and intestine, during adulthood. Studies of conditional deletions of *Foxa2* have shown that Foxa2 alone is not essential for normal liver development but is required for the correct formation of the lung [293, 294]. Moreover, conditional *Foxa2*$^{-/-}$ mice show defects in endocrine pancreas differentiation and the maturation of α- and β-cells, resulting in impaired glucose homeostasis. However, the comparatively mild phenotypes of individual FoxA deletions suggested that FoxA factors can compensate for each other [280]. Indeed, the simultaneous deletion of *Foxa1* and *Foxa2* results in complete liver and near-complete pancreatic agenesis, as well as defects in the differentiation of EECs and GCs in the intestine, demonstrating that these factors can compensate for each other [60, 272, 295]. Moreover, it was observed that *Foxa3* is upregulated upon late-gestational deletion of *Foxa1* and *Foxa2*, and that in absence of Foxa1 and 2, Foxa3 can bind to Foxa1/2-specific target sites.

To study the function of Foxa2 during organ development and homeostasis a fluorescent Foxa2-Venus fusion (FVF) reporter mouse line has been generated in our lab [296]. The FVF knock-in reporter is expressed under the endogenous *Foxa2* promoter and the protein fusion does not interfere with the function of Foxa2, since homozygous mice are viable, fertile and do not display any obvious defects. Careful characterization of the spatial and temporal reporter activity showed that the FVF fluorescence faithfully reflected endogenous gene expression patterns in several tissues, including the gastrointestinal tract. Together, the FVF reporter line allows direct observation of Foxa2 expression using static and live-cell imaging or FACS analysis [296].

1.13 Aim of the Thesis

This thesis is principally divided into two separate parts that aim to analyze the function of the TF PDX1 in early human pancreas development and the role of the pioneer TF Foxa2 in the onset of enterendocrine differentiation in the small intestine, respectively.

Pdx1 is a master regulator of pancreas development. It is expressed throughout pancreatic organogenesis and becomes restricted to β-cells and δ-cells in the mature organ. Heterozygous mutations in the *PDX1* gene are associated with MODY and T2DM, and the homozygous loss causes pancreatic agenesis. The role of Pdx1 in pancreatic development and β-cell function is well described in model organisms such as the mouse. Owing to the restricted access to primary human material, much less is known about the function of PDX1 during human pancreas development. In the first part of this thesis, we wanted to understand how PDX1 regulates early pancreas development in humans and if PDX1 exerts different functions during development and in adulthood. Moreover, we were interested in whether T2DM-associated SNPs are enriched in active regulatory regions of pancreatic progenitors, to examine if pancreas development might affect the susceptibility to T2DM. Furthermore, we wanted to know how missense mutations in the *PDX1* coding region impact pancreas and β-cell development, and if they predispose to T2DM. This will show causality between PDX1 function and T2DM, and identify target genes that are sensitive to impaired PDX1 function, which could aid in the improvement of diagnosis and therapy of diabetic patients.

Intestinal hormones, released by enteroendocrine cells in response to food intake, are important regulators of energy and glucose homeostasis and play a role in the development of T2DM. Enteroendocrine cells are derived from intestinal stem cells but the exact mechanisms and factors driving their differentiation are not fully understood. One transcription factor, which is implicated in the differentiation of a subset of enteroendocrine cells, is Foxa2. In the second part of this thesis, we were interested in the role of Foxa2 during enteroendocrine differentiation. We asked whether the Foxa2 levels influence fate decisions and if Foxa2 acts as pioneering factor to activate regulatory elements involved in enteroendocrine cell formation. This will improve our understanding of intestinal endocrinogenesis, which is important for the development of novel incretin-based therapies for the treatment of metabolic diseases.

2 Results

2.1 The Role of PDX1 in Early Pancreatic Progenitors

Diabetes mellitus is a heterogeneous group of metabolic disorders, characterized by high blood glucose levels, caused by defects in insulin secretion, insulin action or both [1, 2]. The major forms of diabetes include type 1 diabetes mellitus, type 2 diabetes (T2DM), gestational diabetes and maturity onset diabetes of the young (MODY). MODY is a group of monogenic, autosomal-dominant inherited forms of diabetes that result from mutations in key genes, regulating β-cell development and function. One of the genes that can cause MODY when mutated at certain sites is *PDX1*, which encodes for the TF pancreatic and duodenal homeobox 1 (PDX1), a master regulator of pancreas and β-cell development. In mice with homozygous loss of Pdx1 the pancreas fails to develop and a heterozygous loss causes diabetes during adulthood due to β-cell apoptosis [131–133]. Although the function of Pdx1 is well studied in pre-clinical models, the precise function of its human ortholog during human pancreas development is still poorly understood. Moreover, it is unclear how PDX1 regulates its downstream transcriptional program and to which extent mutations of the gene impact its function.

To further the understanding of PDX1 function in pancreas development and how mutations of PDX1 affect pancreas and β-cell development, we made use of donor-specific iPSCs. We generated three novel iPSC lines from one healthy female donor (XM001) [297] and two female donors with impaired insulin secretion who carried heterozygous $PDX1^{C18R/+}$ and $PDX1^{P33T/+}$ missense mutations, respectively [298, 299]. They were identified by screening a large cohort of human subjects with a high risk to develop T2DM [300]. Furthermore, we generated three isogenic cell lines carrying homozygous $PDX1^{C18R/C18R}$ and $PDX1^{P33T/P33T}$ mutations, as well as a heterozygous $PDX1^{+/-}$ loss-of-function mutation from the XM001 control iPSC line [301]. The iPSC lines were generated from primary skin fibroblasts by reprogramming them with three episomal plasmids, encoding *OCT4, SOX2, NANOG, LIN28, KLF4* and *L-MYC* in a three stage protocol (Figure 2.1 A). All iPSC lines had a normal karyotype and were episomal vector integration-free. To study PDX1 function and the impact of PDX1 mutations, iPSCs were differentiated into β-like cells (Figure 2.1 B) and characterized by means of expression of specific markers and their ability to secrete insulin in response to a glucose stimulus. In brief, we found that early stages of pancreas development were normal in $PDX1^{P33T/+}$ and $PDX1^{C18R/+}$ cells, while later stages of development were significantly impaired and fewer β-like cells were generated, which were also less responsive to glucose. Cells with homozygous $PDX1^{P33T/P33T}$ and $PDX1^{C18R/C18R}$ or heterozygous $PDX1^{+/-}$ mutations differentiated into PP1 cells with normal efficiency (∼90%) but expression of PDX1 was reduced in $PDX1^{P33T/P33T}$ and $PDX1^{+/-}$ PPs. At the PP2 stage, $PDX1^{P33T/P33T}$ and $PDX1^{+/-}$ had fewer $PDX1^{+}$/NKX6.1^{+} cells compared to $PDX1^{C18R/C18R}$ and XM001 control cells. Differentiation into β-

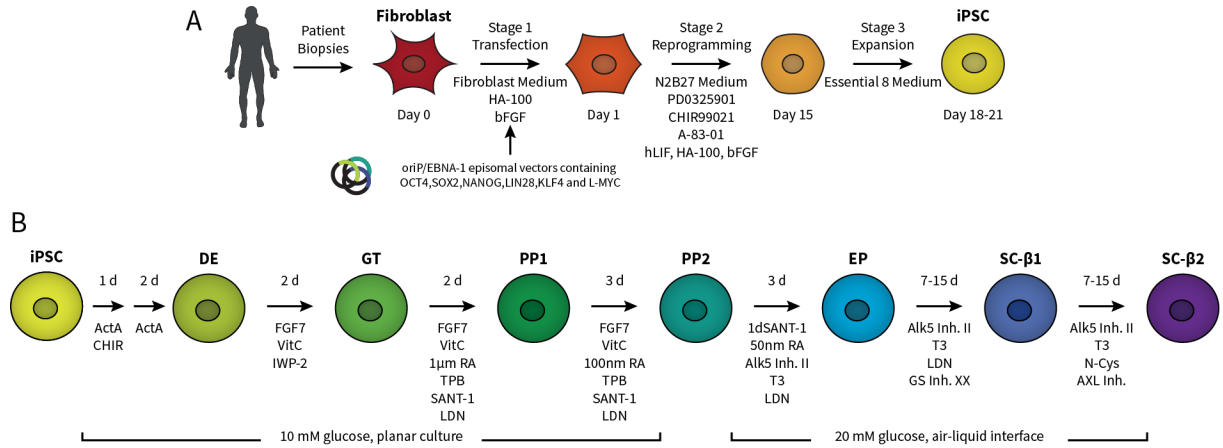

Figure 2.1: Generation and Differentiation of IPS Cell Lines. A: Schematic representation of the reprogramming steps used to generate iPSC lines. **B:** Schematic representation of the protocol used for iPSC differentiation. Figure B modified from [301] under a Creative Commons license.

like cells showed that all isogenic mutant cell lines generate fewer C-peptide-positive cells, fewer monohormonal C-peptide-positive cells and fewer cells expressing both, C-peptide and NKX6.1. Functional analyses of mutant β-like cells by glucose stimulated insulin secretion assays showed impaired glucose responsiveness, and gene expression analyses revealed deregulation of important β-cell genes [301]. Moreover, iPSCs derived PPs were characterized by transcriptome profiling and the genome-wide identification of PDX1 binding sites. This work has been published [297–299, 301] and the data on genomics of these publications are presented in this section.

2.1.1 Expression Profiling of XM001 iPS & PP Cells

The XM001 cell line generated in our lab has the potential to efficiently differentiate into pancreatic progenitors (PPs), as shown by immunofluorescent imaging and flow cytometry analysis [297]. In order to profile the gene expression changes, characteristic for the differentiation process from iPSCs to PPs, we performed Affymetrix microarray analysis on XM001 cells at the pluripotency and PP stages. Differential expression analysis identified 2658 differentially expressed genes (fold change$\geq$2 and FDR$\leq$5%), including several important pancreatic genes such as *HNF1B*, *SOX9*, *ONECUT1*, *HHEX*, *FOXA2* and *RFX6* that were, as expected, upregulated in PPs compared to iPSCs. On the other hand, pluripotency-associated genes, including *NANOG* and *SOX2* were specifically expressed in iPSCs and downregulated during PP differentiation (Figure 2.2 A). In addition, we performed ontology and pathway analysis to better understand the biological implications of the expression differences that we observed. Genes in GO terms, associated with epithelial differentiation, (early) pancreas and β-cell development or MODY were upregulated in XM001 PPs. Genes, associated with pluripotency, embryonic stem cells, cell cycle and proliferation were downregulated in differentiated PPs (Figure 2.2 B). Moreover, gene set enrichment analysis (GSEA) showed a clear enrichment of genes from the GO terms endocrine pancreas development and pancreas development in PP cells compared to iPSCs (Figure 2.2 C). To verify the microarray results, we tested the expression of several important genes involved in pancreas development (*HHEX*, *HNF1B*, *MEIS1*, *ONECUT1*, *PDX1*, *PTF1A*, *RFX6*,

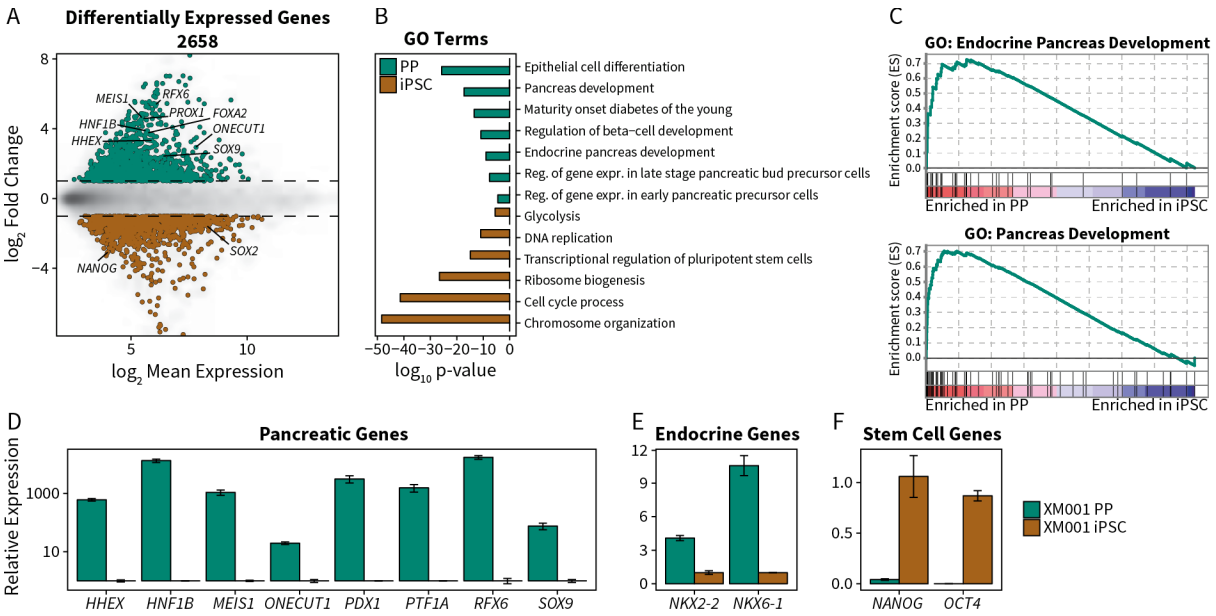

Figure 2.2: XM001 PPs Express Pancreas Specific Genes. A: MA plot showing the mean $\log_2$ expression against the $\log_2$ fold change of the differential expression analysis, comparing microarray data from XM001 PPs and iPSCs. Genes with significantly different expression ($\log_2$ fold change$\geq$1 and adjusted p-value$\leq$0.05) are shown in green (higher expression in PPs) or brown (higher expression in iPSCs). **B:** Bar graph of p-values from selected GO terms and KEGG and Reactome pathways from differentially expressed genes, shown in A. **C:** Gene set enrichment analysis of pancreas related GO terms. **D-F:** qPCR validation of the expression of pancreas (D), endocrine (E) and stem cell (F) related genes in XM001 iPSCs and PPs (n = 3). Errorbars, standard deviation (SD). Figure modified from [297] under a Creative Commons license.

and *SOX9*), endocrine formation (*NKX2-2*, *NKX6-1*) and stem cell maintenance (*NANOG*, *OCT4*) by quantitative PCR (qPCR) and validated the results from the microarray screen (Figure 2.2 D-F). Thus, after 10 days of *in vitro* differentiation, XM001 iPSCs were efficiently differentiated into PP cells, characterized by the expression of PDX1 and other lineage-specific genes and the downregulation of pluripotency genes.

2.1.2 The Enhancer Landscape of XM001 PP Cells

The correct specification of tissue and cell type-specific enhancers is an integral part of the differentiation of individual cell types [302]. The histone modification H3K27ac is associated with active regulatory elements [303–307] and is widely used to profile the active enhancer landscape. We therefore analyzed H3K27ac in XM001 cells after 10 days of differentiation using chromatin immunoprecipitation followed by next generation sequencing (ChIP-seq) . H3K27ac was enriched around TSS and the levels of H3K27ac at the promoters correlated with the expression levels of corresponding genes (Spearman's ρ= 0.57, Figure 2.3 A-B). Recently, Wang et al. [302] thoroughly characterized the progression of chromatin activity during several stages of pancreatic endoderm differentiation (embryonic stem cells (ESCs), definitive endoderm (DE), gut tube (GT), posterior foregut (FG; corresponds to PP stage of this study) and late pancreatic progenitors/pancreatic endoderm (PE)) using H3K27ac and showed that the chromatin state

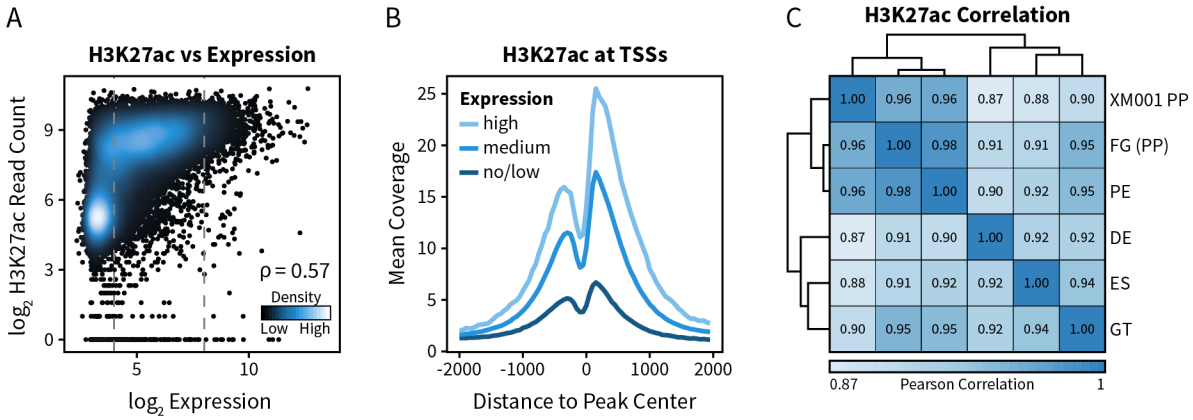

Figure 2.3: H3K27ac in XM001 PPs. A: Scatter plot of H3K27ac $\log_2$ read counts from 1 kb regions around TSSs against the $\log_2$ microarray expression values of the corresponding genes. Spearman's rank correlation shows a positive correlation between the amount of H3K27ac at promoters and the expression of the gene (ρ= 0.57). Vertical dashed lines indicate the expression thresholds for genes with no or weak, medium and high expression, as used in B. **B:** H3K27ac coverage profile of 2 kb regions around TSSs shows an enrichment of H3K27ac at TSSs that is increasing with increasing gene expression. **C:** Hierarchical clustering of Pearson correlation coefficients from H3K27ac data from XM001 and published datasets [302] from embryonic stem cells (ES), definitive endoderm (DE), gut tube (GT), posterior foregut (FG) and pancreatic endoderm (PE) cells. Figure C is modified from [297] under a Creative Commons license.

at lineage-specific enhancers is crucial for developmental competence. We next compared our H3K27ac profile to those from Wang et al. [302] to test if XM001 PPs have a similar pattern of H3K27ac. Unsupervised hierarchical clustering of H3K27ac profiles showed that XM001 PPs cluster together with FG and PE cells derived from hESCs, indicating that our iPSC-derived PPs represent an early stage of pancreas development, shortly after PDX1 induction (Figure 2.3 C). Thus, despite the different origin of cells and the different protocols used to generate PP-/PE-stage cells, the landscape of active enhancers as defined by H3K27ac is highly similar.

2.1.3 Characterization of PDX1 Binding in XM001 PP Cells

The transcription factor PDX1 plays a crucial role in the specification and differentiation of the pancreatic lineages. In mouse, expression of Pdx1 becomes evident at E8.5–9 and is necessary for pancreas development, as homozygous Pdx1 knock out mice suffer from pancreatic agenesis [56, 57]. During later stages of development, Pdx1 becomes restricted to β-cells and δ-cells [41, 54, 58, 125, 126]. Thus, we were interested in transcriptional targets of PDX1 to gain insight into the development of the human pancreas.

In order to profile genome-wide PDX1 binding sites in human PPs, we performed ChIP-seq analysis in XM001 iPSCs after 10 days of differentiation, a time point at which PDX1 is robustly expressed in ~85% of the differentiated cells [297]. In total, we identified 8088 PDX1-bound sites and 5664 associated putative target genes with PDX1 binding sites within 20 kb of their TSSs or within their gene bodies. These genes include *PDX1* itself and other important pancreatic genes such as *RFX6*, *MEIS1* and the MODY-gene *HNF1B* (Figure 2.4 A). PDX1-bound regions were predominantly found in intergenic (44.2%) and intronic (40.4%)

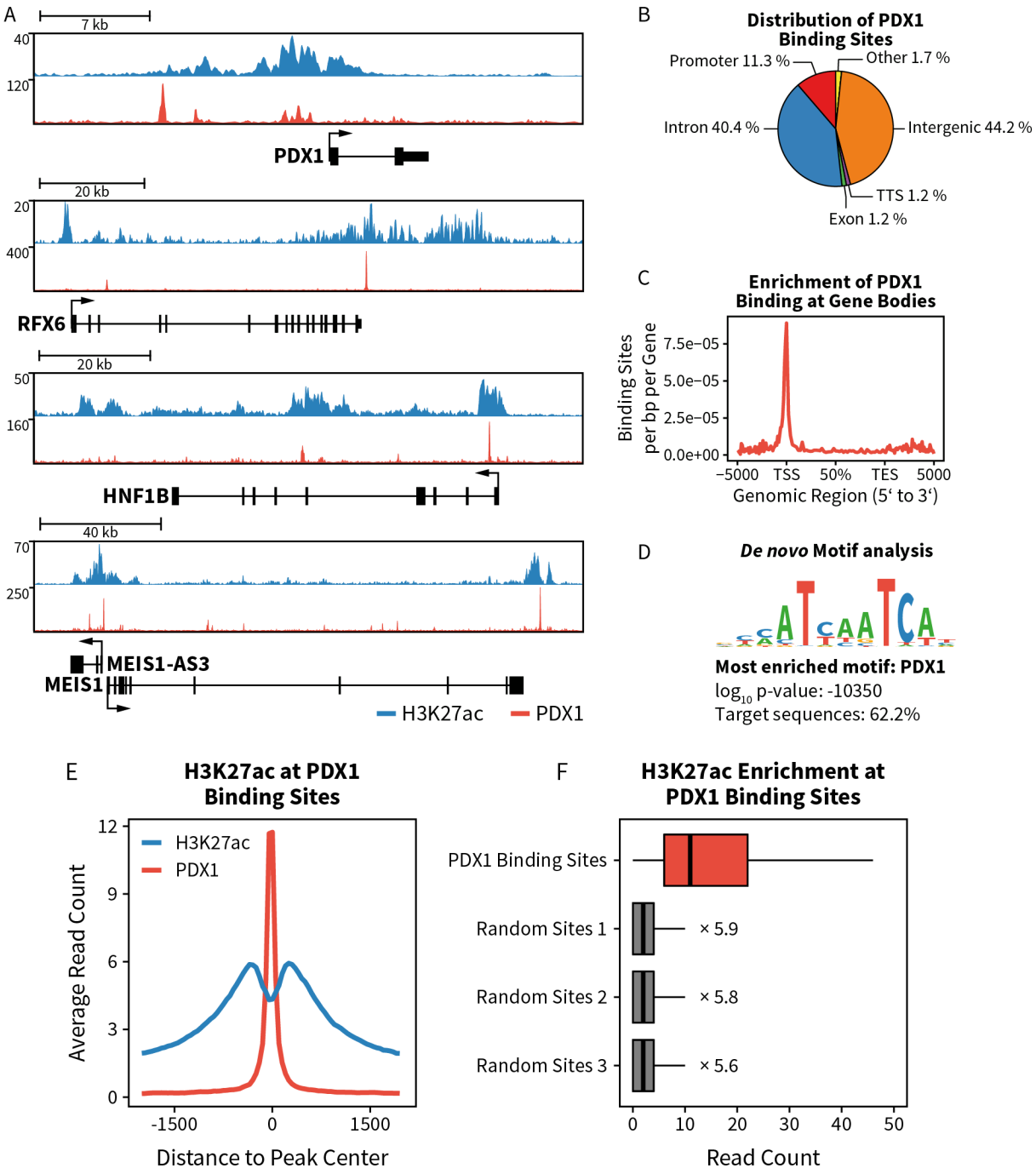

Figure 2.4: Characterization of PDX1 Binding in XM001 PP Cells. A: ChIP-seq data tracks showing the enrichment of H3K27ac (blue) and PDX1 (red) at the loci of selected important pancreatic genes. **B:** Distribution of PDX1 binding sites across genomic features shows preferential binding to intergenic and intronic regions. **C:** Meta-gene profile plot showing enrichment of PDX1 binding sites around TSSs displayed as binding sites per bp per gene over the genomic regions of all RefSeq genes. **D:** Most enriched motif discovered by *de novo* motif analysis resembles the known PDX1 consensus sequence and is identified in 62.2% of all PDX1-bound sequences. **E:** ChIP-seq coverage profiles of H3K27ac (blue) and PDX1 (red) at PDX1 binding sites shows enrichment of H3K27ac at PDX1-bound sites. **F:** Read count of H3K27ac at the 8088 PDX1 binding sites in XM001 PPs and at three sets of 8088 random sites of the same length distribution. H3K27ac is enriched > 5.6× at PDX1-bound sites compared to the random sites. For all comparisons, p-value < 2.2×10^{-16}. Figure modified from [297] under a Creative Commons license.

regions but are enriched at promoter regions that harbor 11.3% of the binding sites (Figure 2.4 B-C). To identify sequence motifs, enriched in PDX1-bound sites, we performed *de novo* motif analysis and found that the most enriched motif is present in 62.2% of the binding sites and resembles the known binding motif of PDX1, confirming the specificity of the PDX1 ChIP-seq approach (Figure 2.4 D). Moreover, we found that the histone modification H3K27ac is significantly enriched at PDX1 binding sites. Compared to a set of random sites of the same length distribution, H3K27ac was enriched > 5.6× at PDX1-bound sites (Figure 2.4 E-F).

2.1.4 Functional Characterization of PDX1-Bound Genes in XM001 PPs

In the next step, we integrated the PDX1 binding profile with the expression data from XM001 PPs and examined the function of PDX1-bound genes. From the 5664 PDX1-bound genes, 4523 were covered by the microarray and 755 of the PDX1-bound genes were differentially expressed between iPSC and PP stage XM001 cells. This accounts for 28.5% of all differentially expressed genes between these two developmental stages (Figure 2.5 A). Among the PDX1-bound genes that are specifically expressed in PP cells, we found several important pancreatic genes, including *RFX6*, *FOXA2*, *HNF1B* and *MEIS1*. Of note, we also found *RFX3* and *DLL1* among the upregulated putative PDX1 target genes which are known to be involved in endocrine differentiation in the mouse [67, 308, 309], suggesting a potential role of these genes also during human pancreas development (Figure 2.5 B). GO term and pathway analysis further confirmed that PDX1-bound genes, expressed in XM001 PP cells were associated with pancreas, endocrine and β-cell development and revealed that they are also enriched in Notch and Hippo signaling pathway genes (Figure 2.5 C). These results suggest that PDX1 induces a pancreatic program and primes for endocrine differentiation.

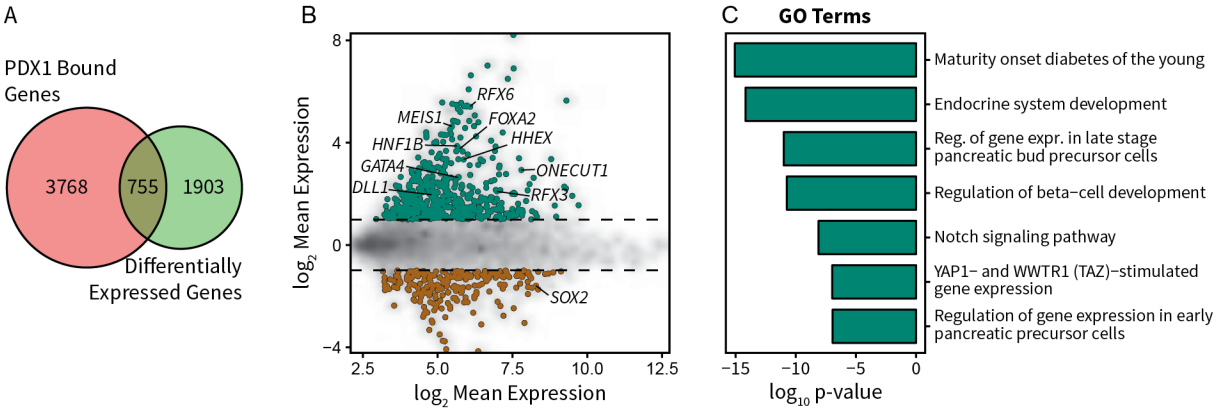

Figure 2.5: Functional Characterization of PDX1-Bound Genes in XM001 PPs. A: Venn diagram showing the overlap between PDX1-bound genes and the differentially expressed genes identified by microarray analysis. **B:** MA plot of the mean $\log_2$ expression against the $\log_2$ fold change of PDX1-bound genes. Differentially expressed genes are plotted in green (higher expression in PPs) or brown (higher expression in iPSCs). **C:** Bar chart of $\log_{10}$ p-values from enriched GO terms and KEGG and Reactome pathways of PDX1-bound genes upregulated in PPs. Figure modified from [297] under a Creative Commons license.

2.1.5 Definition of High Confidence PDX1 Binding Sites

To our knowledge, this PDX1 binding profile is the first generated from iPSC derived PPs, however, several PDX1 ChIP-seq data sets from hESC derived PP or PE cells have been published [302, 310, 311]. Since all these data sets were generated in different labs, using different cell lines and protocols for ChIP-seq and differentiation, we were interested in the differences between them and compared the data sets from Wang et al. [302] and Teo et al. [311] to ours. The overlap of the PDX1 binding sites showed some obvious differences and only 1606 sites were found in all three data sets. Very likely, this results from the differences between hESC and iPSC, genetic background variations and differences in the *in vitro* differentiation stage or protocol. Similarly, available PDX1 ChIP-seq data from adult human islets, obtained from different individuals with a unique genomic background and generated in different labs, showed variability in the detected PDX1 binding sites. However, the overlap of different data sets is useful to define a set high confidence PDX1 binding sites of adult islets and PP stage cells that is not influenced by any of the aforementioned factors. We took advantage of the published data sets of PDX1 ChIP-seq from PPs and adult human islets to derive two sets of high confidence PDX1 binding sites, comprised of 4979 and 6967 PDX1 sites, consistently bound across multiple data sets from PPs and adult islets, respectively (Figure 2.6).

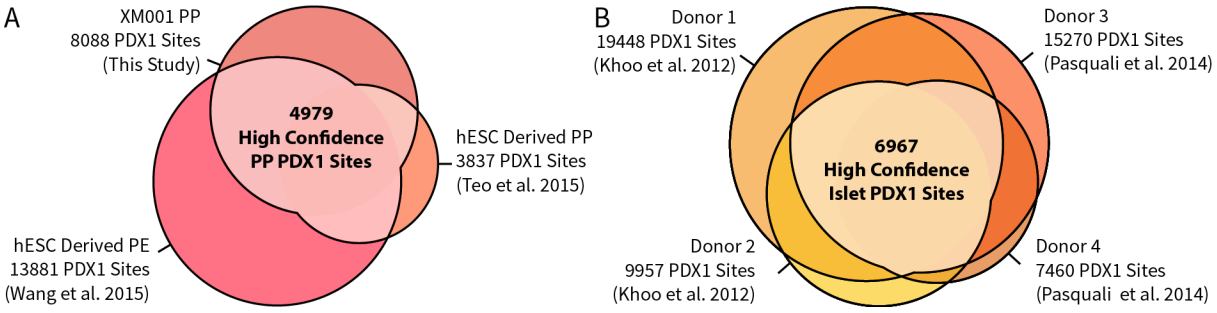

Figure 2.6: Definition of High Confidence PDX1 Binding Sites. A: Venn diagram showing the overlap between PDX1 binding sites in XM001 PP cells, hES-derived PP cells from Dunn lab [311] and hES-derived PE cells from Sander lab [302]. **B:** Venn diagram showing the overlap between PDX1 binding sites in adult human islets from four different donors [312, 313]. Sites bound in two out of three PP data sets or three out of four islet data sets were considered as high confidence PDX1 binding sites.

2.1.6 Function of PDX1-Bound Genes in Development & Adulthood

During pancreas development in mouse, Pdx1 follows a distinct tempo-spatial expression pattern. In early stages of development, Pdx1 is expressed in all PPs and becomes gradually restricted to mature β- and δ-cells in the fully developed organ. Based on that, we hypothesized that Pdx1 might control distinct molecular programs in a stage-specific manner.

To test if PDX1 serves different functions during development and adulthood, we compared PDX1 binding sites from PPs and adult human islets. We took advantage of the published data sets of PDX1 ChIP-seq from adult human islets to derive another set of high confidence PDX1 binding sites, comprised of 6967 PDX1 sites, consistently bound across multiple data sets, which we compared to the high confidence binding sites from PP cells. Of the 4978 high confidence binding sites in PP cells, 1018 (20.4%) were overlapping with sites from the islet binding data. This indicates that only a minority of binding sites are conserved between PPs and adult islets and that most of the PDX1 binding sites are changing during the differentiation and maturation processes (Figure 2.7 A). Correspondingly, we found genes important for pancreas development, such as *GATA4*, *HES1*, *HHEX*, *HNF1A*, *NEUROG3*, *ONECUT1*, *TBX1* and *YAP1* near PP-specific high confidence binding sites. In contrast, genes linked to β-cell maturation and identity as well as mature β-cell functions like insulin secretion and glucose sensing, including *GLPR1*, *KCNJ11*, *SLC2A2*, *NKX6-1* and *MAFA* were found near islet-specific binding sites. Genes in the vicinity of shared binding sites are implicated in endocrine development and include *NKX2-2*, *PAX4*, *RBPJ* and *TGFB1* (Figure 2.7 A-B). In addition, we observed genes with a combination of both, stage-specific and stable PDX1 binding sites. Among these were *HNF1B*, *RFX3* and *RFX6*, *MNX1*, *FOXA2*, the *MEG3* locus and other important pancreatic genes (Figure 2.8). Taken together, our results suggest a dynamic pattern of PDX1 binding that changes during the process of pancreas and endocrine cell development to meet the functional requirements of the different stages during development and homeostasis.

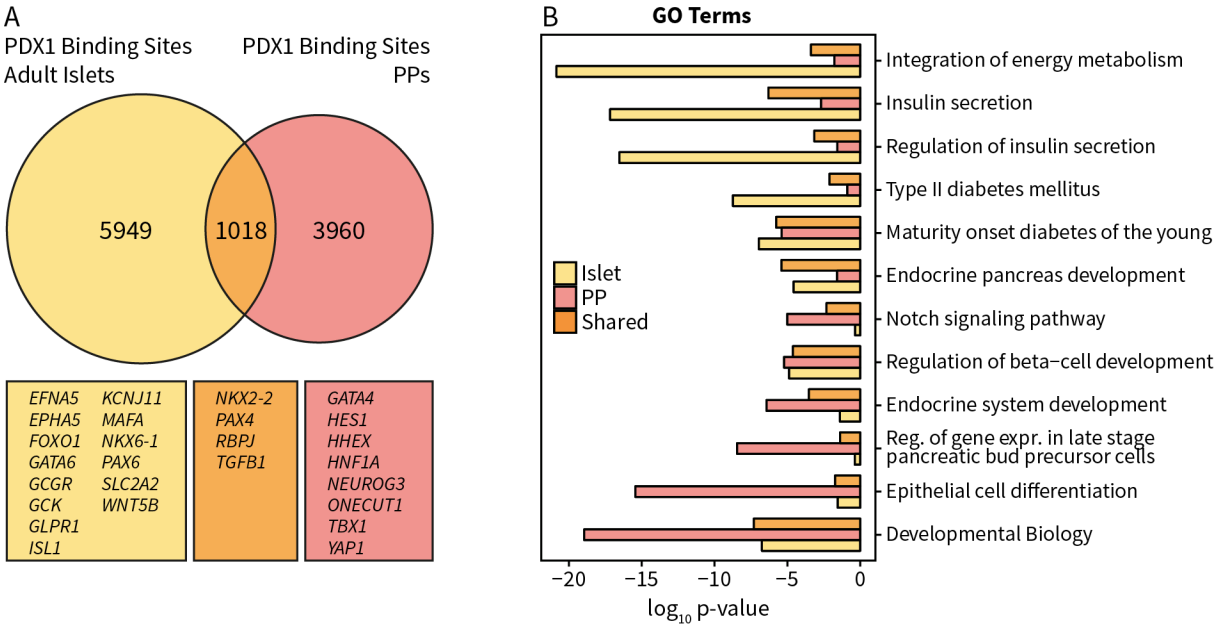

Figure 2.7: Different Functions of PDX1-Bound Genes in Development and Adulthood. A: Venn diagram showing the overlap of high confidence PDX1 binding sites from adult human islets and PPs with selected annotated genes. **B:** Bar chart of $\log_{10}$ p-values from enriched gene ontology and KEGG/Reactome pathways from genes bound in adult islets and/or PPs. Figure modified from [297] under a Creative Commons license.

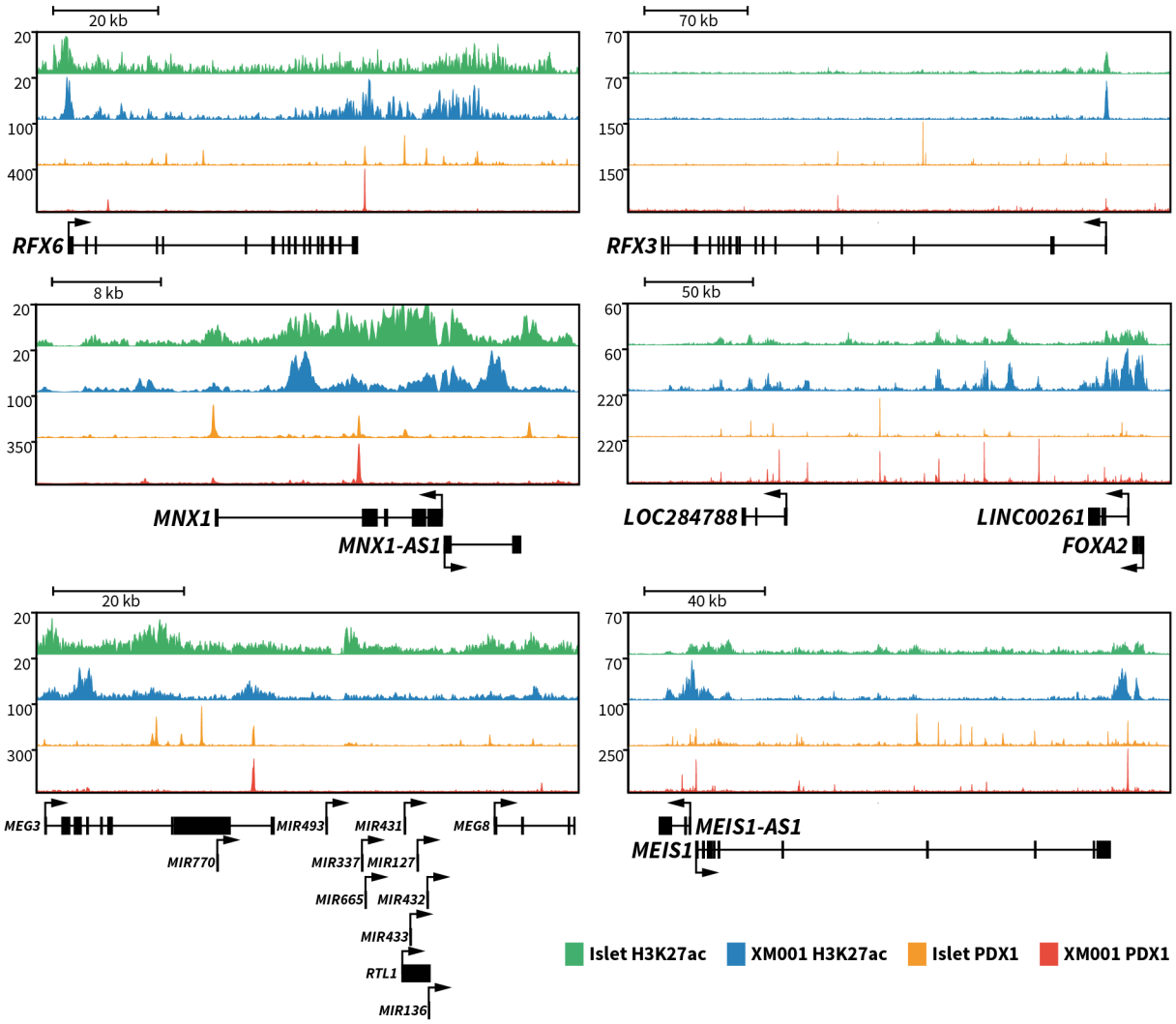

Figure 2.8: Differential Binding of PDX1 in XM001 PPs and Adult Islets. ChIP-seq data tracks showing H3K27ac and PDX1 from XM001 PP and adult islet cells [302,312,313] at selected loci with stage-specific PDX1 binding. Figure modified from [297] under a Creative Commons license.

2.1.7 Analysis of T2DM SNPs in XM001 PPs & Adult Islets

Single nucleotide polymorphisms (SNPs) are the most common form of genetic variation and approximately four to five million are found in each person's genome. Most SNPs have no impact on health or development but some may increase the risk of developing a particular disease or may even be causal for a given disease. Genome-wide association studies (GWAS), aiming to identify disease associated SNPs, identified more than 300 genes that are linked to T2DM [314]. While severe mutations in MODY genes lead to monogenetic forms of diabetes [13–16], most SNPs associated to T2DM occur in non-coding regions and increase the susceptibility to T2DM. In XM001 PPs 104 (32%) of the 325 known T2DM-linked genes are bound by PDX1 and it has been shown that islet enhancers are enriched in T2DM-associated SNPs [313, 315]. Therefore,

we asked whether active chromatin regions of XM001 PPs, defined by enrichment of H3K27ac, are also enriched in T2DM-associated SNPs and if different variants are enriched in PPs and adult islets.

We used the DIAGRAMM 1000G GWAS meta-analysis data [316] and identified SNPs with a high association significance (1788 SNPs, $p \leq 5 \times 10^{-8}$) located within active chromatin regions in XM001 PPs and adult islets [302, 313]. We found a significant enrichment of T2DM-associated SNPs in active chromatin of both PPs (437 SNPs, $p=1.0156 \times 10^{-106}$) and islets (918 SNPs, $p=8.2266 \times 10^{-263}$). Next, we analyzed whether T2DM-SNPs are localized in active chromatin regions, shared between PPs and mature islets and identified 380 SNPs in regions of stable H3K27ac enrichment. In addition, we found 57 SNPs in regions, specifically active in PPs (Figure 2.9 A). Mapping of the SNPs to genes identified 857 SNPs in the vicinity of known genes (20 kb up-/ downstream of or within a gene body) that included mostly known T2DM-associated genes, e.g. *TCF7L2*, *IGF2BP2*, *HNF1B*, and *FTO* (Figure 2.9 B). For most of the identified genes, SNPs lie in regions, active in both, PPs and adult islets, while genes with SNPs in cell type-specific active regions were less frequent. SNPs in islet-specific regions were found near the genes *FTO*, *HNF1A*, *KCNJ11*, *PPP2R2C*, and *SYN2* and SNPs in PP-specific regions near *SYN2*, *IGF2BP2*, *KIF11*, and *PPARG*. Notably, we found only a single T2DM-related locus, the *PPARG* locus, in which a T2DM-associated SNP falls exclusively into a PP-specific active region and that does not have any T2DM-associated SNP in a commonly active or islet-specific active site. Together, the enrichment of T2DM SNPs in active regulatory regions of PP cells raises the possibility that the susceptibility to T2DM is, at least in parts, already established during pancreatic development, possibly through miss-regulation of key genes and developmental programs.

Since PDX1 is a key player in pancreas development and β-cell function, we tested if T2DM-associated SNPs lie within PDX1 binding sites. Interestingly, we found only six T2DM-SNPs that fall into PDX1-bound sites (3 SNPs in PP PDX1 sites and 3 SNPs in islet PDX1 sites) (Figure 2.9 B). The SNP rs8044769, rs78681698 and rs150111048 were in PDX1-bound regions in islets and were near *FTO*, *THADA*, and *IGF2BP2*. In PPs, 3 SNPs near *HNF1B* and *TCF7L2* were found in PDX1-bound regions. One of them is the SNP rs11263763 that was located at a PP-specific PDX1 binding site in the first intron of the *HNF1B* gene (Figure 2.9 C). Moreover, at the *HNF1B* locus, there were two PP-specific enhancer regions, in the 4^{th} intron and downstream of *HNF1B*, suggesting that *HNF1B* expression is regulated through different mechanisms during development and adulthood. The other two SNPs, rs12762233 and rs7087006, were located in PP-specific PDX1 binding sites at the *TCF7L2* locus and might impact the regulation of *TCF7L2* expression during development. Intriguingly, SNPs near the *TCF7L2* locus have the highest association to T2DM and mutations of *HNF1B* can cause MODY5. Thus, mutations in these cis-regulatory regions of these genes may contribute to an increased susceptibility to T2DM. This further supports the notion that the risk for T2DM might be established during early pancreas development.

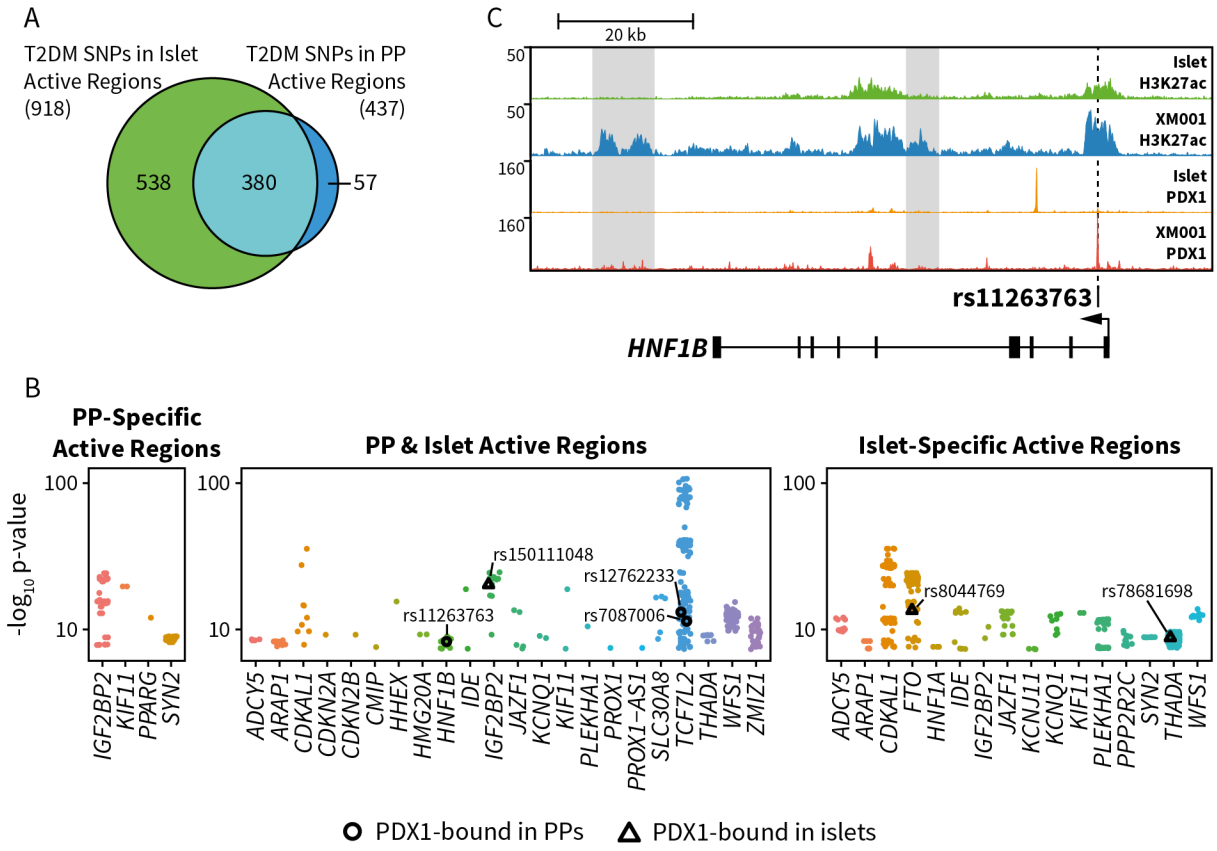

Figure 2.9: T2DM SNPs in XM001 PPs and Adult Islets. A: Venn diagram showing the overlap of T2DM SNPs found in active regions of adult islets and XM001 PPs. **B:** Disease-association p-values of T2DM SNPs near known T2DM-associated genes found in active regions of adult islets [312, 313] and XM001 PPs. The p-values of SNPs bound by PDX1 in PPs and islets are shown as black circles and triangles, respectively. **C:** ChIP-seq data tracks showing H3K27ac and PDX1 from PP and islet [302, 312, 313] cells at the *HNF1B* locus. The SNP rs11263763 is located in a PP specific PDX1 binding site in the first intron of *HNF1B*. The PP specific enhancer regions are shaded in gray. Figure modified from [297] under a Creative Commons license.

2.1.8 Transcriptional Profiling of Isogenic Mutant PPs

Mutations in key pancreatic genes can cause maturity onset diabetes of the young (MODY) which refers to hereditary, monogenic forms of diabetes mellitus. There are 11 recognized forms of MODY and two forms of monogenetic neonatal diabetes, each caused by mutations in different genes, such as *HNF4A*, *GCK* and *HNF1A* in MODY1 to MODY3, respectively [317]. MODY4 is characterized by a heterozygous P63fsdelC frameshift mutation in the *PDX1* gene that results in a truncated protein, lacking the DNA-binding homeodomain. Homozygosity of this loss-of-function mutation leads to pancreatic agenesis [16]. In addition to this MODY mutation, several other, less severe missense mutations, including the C18R and P33T mutations of the transactivation domain of PDX1 have been associated with an increased risk of T2DM or gestational diabetes [136, 139]. To study the effect of the PDX1 C18R and P33T mutation on pancreatic and endocrine development, we generated iPSC lines from heterozygous carriers of these mutations [298, 299] and differentiated them into β-like cells *in vitro* [301]. Both cell

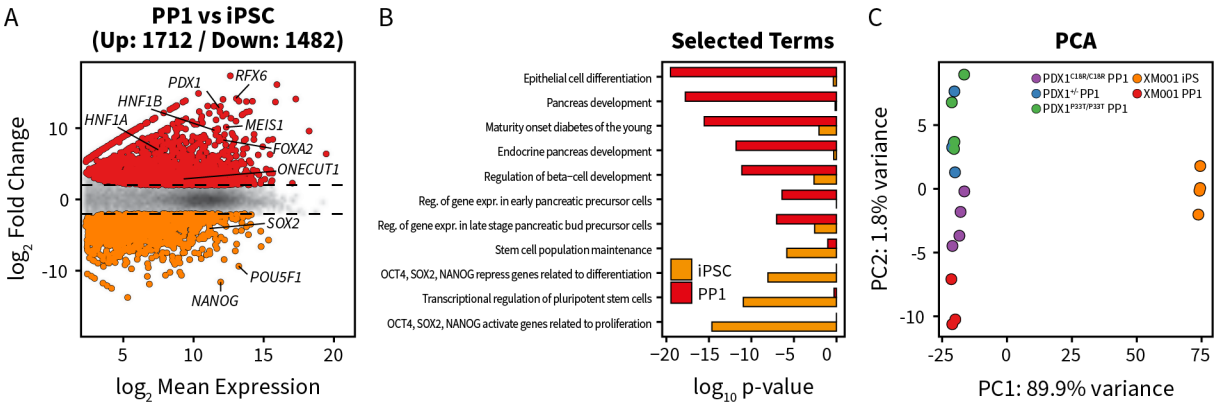

Figure 2.10: Differential Gene Expression in iPSCs and Isogenic Mutant PPs. A: MA plot showing the mean $\log_2$ expression against the $\log_2$ fold change of RNA-seq expression data from control XM001 iPSCs and the union of PPs from control and mutant cell lines. Genes with significantly different expression (fold change $\geq$4, FDR $\leq$5%) are shown in red (higher expression in PPs) or orange (higher expression in iPSCs). **B:** Bar graph of p-values, obtained from GO term and pathway enrichment analysis in differentially expressed genes, showing the enrichment of selected terms in PPs and iPSCs. **C:** Scatter plot showing the PCA of individual RNA-seq samples. The major sources of variance are the differentiation stage (PC1) and mutations in the *PDX1* gene (PC2). Figure modified from [301] under a Creative Commons license.

lines exhibited impaired differentiation capability and reduced glucose-stimulated insulin secretion [301]. To control for genomic-background mediated effects, we introduced homozygous C18R and P33T mutations as well as a heterozygous PDX1 loss-of-function mutation ($PDX1^{+/-}$) in XM001 iPSCs. In line with previous results, the isogenic cell lines displayed defects in β-like cell differentiation and glucose responsiveness [301]. To study these mutations in early pancreatic development, we differentiated isogenic mutant and control cell lines into PP1 stage cells and found that PPs were generated with similar efficiencies (∼90%) in all cell lines, but that PDX1 expression levels are substantially lower in $PDX1^{P33T/P33T}$and $PDX1^{+/-}$PP1 cells. Also, the differentiation to the PP2 stage of $PDX1^{P33T/P33T}$and $PDX1^{+/-}$cells was impaired, as evidenced by a significantly lower fraction of ∼20% PDX1/NKX6.1 double-positive cells compared to the ∼40% detected in $PDX1^{C18R/C18R}$and XM001 control PP2s [301].

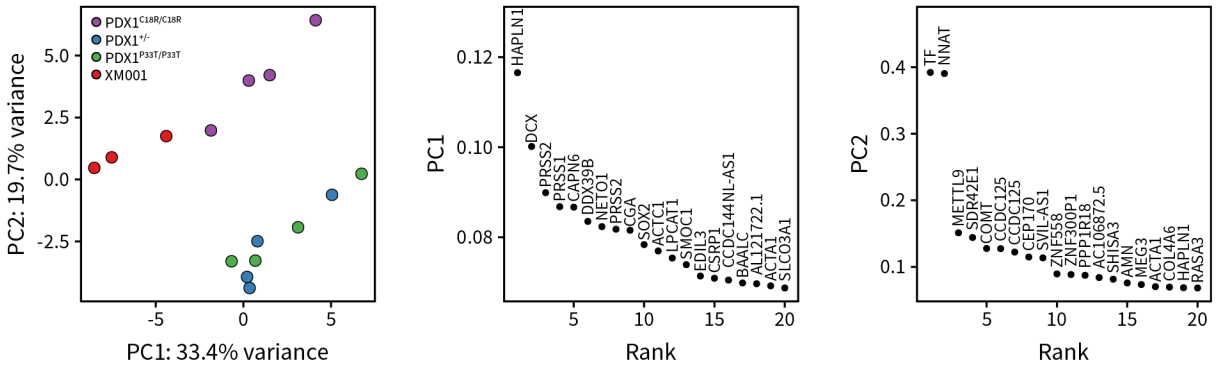

Figure 2.11: PCA of RNA-seq from Control, $PDX1^{+/-}$, $PDX1^{P33T/P33T}$and $PDX1^{C18R/C18R}$Mutants. A: PCA of individual RNA-seq samples from PPs. Here, the major source of variance reflects the differences between control and *PDX1* mutations. **B:** Genes ranked by their contribution to PC1. **C:** Genes ranked by their contribution to PC2. Figure modified from [301] under a Creative Commons license.

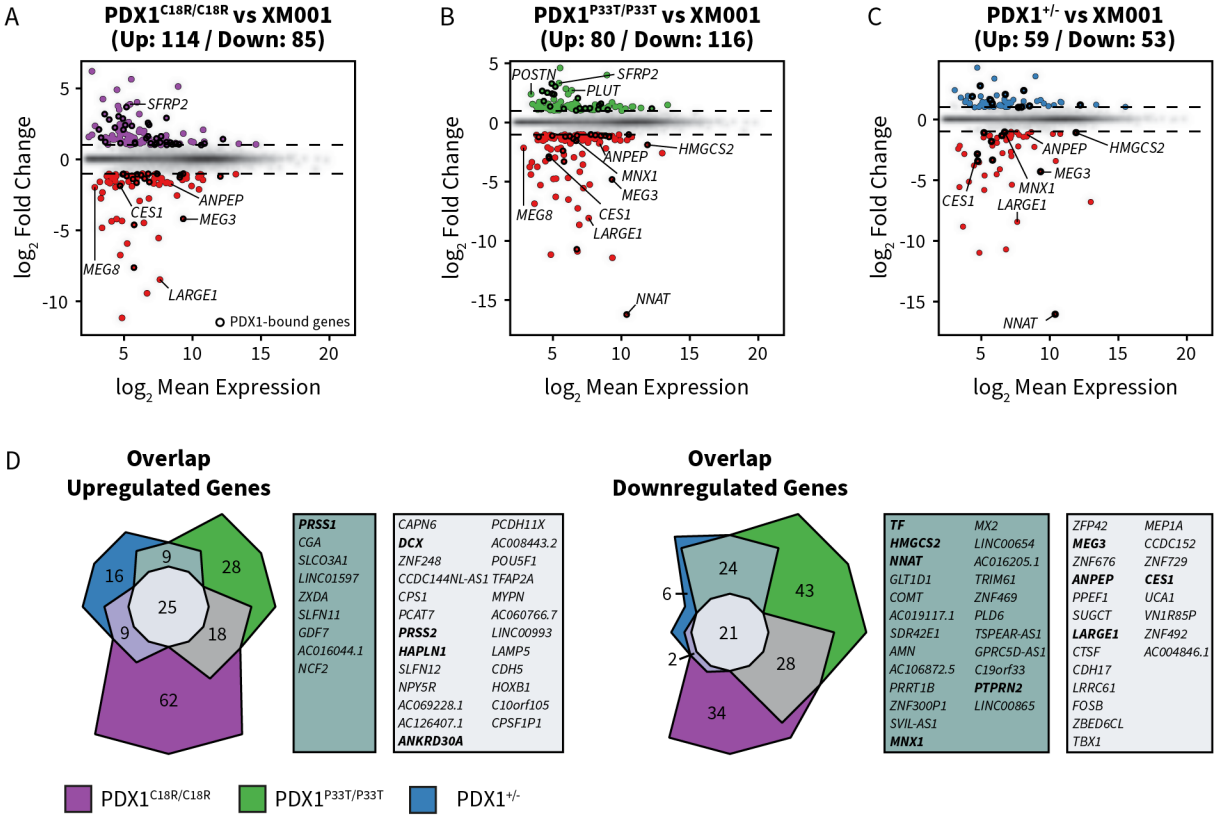

Figure 2.12: Differential Gene Expression in *PDX1*$^{+/-}$, *PDX1*$^{P33T/P33T}$ and *PDX1*$^{C18R/C18R}$ Mutant Lines. **A-C:** MA plots showing the mean $\log_2$ expression against the $\log_2$ fold change of the RNA-Seq data obtained from *PDX1*$^{+/-}$, *PDX1*$^{P33T/P33T}$ and *PDX1*$^{C18R/C18R}$ compared to XM001. Genes with significantly different expression ($\log_2$ fold change ≥ 1 and adjusted p-value ≤ 0.1) are drawn in color. Red depicts increased expression in XM001 control PPs, purple, green and blue depict increased expression in *PDX1*$^{C18R/C18R}$ (**A**), *PDX1*$^{P33T/P33T}$ (**B**) and *PDX1*$^{+/-}$ (**C**), respectively. Black circles mark genes with PDX1 binding sites in XM001 PPs. **D:** Venn diagrams showing the overlap of up- and downregulated genes from *PDX1*$^{C18R/C18R}$, *PDX1*$^{P33T/P33T}$ and *PDX1*$^{+/-}$ PPs compared to XM001 PPs. Some relevant pancreatic and disease-associated genes are highlighted in bold. Figure modified from [301] under a Creative Commons license.

To identify the immediate early PDX1 target genes, which react to haploinsufficiency and the homozygous point mutations and which could explain the impaired PP2 differentiation, we performed RNA-seq at the PP1 stage. As a quality control step, we included wild type XM001 iPSCs in our analysis to monitor PP differentiation. We compared the XM001 control iPSCs to the average of all, mutant and control, PPs and detected the upregulation of a panel of pancreatic genes (including *PDX1*, *HNF1B*, *SOX9*, *ONECUT1*, *HHEX*, *FOXA2* and *RFX6*), indicative of proper PP differentiation (Figure 2.10 A). In line with this, GO-term and pathway analysis showed upregulation of genes associated with early pancreas and endocrine pancreas development, while stem cell-related genes were downregulated (Figure 2.10 B). In addition, principal component analysis (PCA) clearly showed two distinct clusters along the 1st principal component (PC), separating iPSCs and PPs. The axis of the 2nd PC, reflected the differences between the mutant and control PPs (Figure 2.10 C).

For a more precise picture of the differences between mutant and control PPs, PCA was repeated using only PP samples. Here, the 1st PC reflected the variance between control and mutant PPs

and the genes that contribute most to these differences included *HAPLN1, DCX, PRSS1/2* and *CAPN6*. The 2nd PC was predominantly defined by the genes *TF* and *NNAT* and represented variation of $PDX1^{P33T/P33T}$and $PDX1^{+/-}$compared to $PDX1^{C18R/C18R}$and control (Figure 2.11).

Next, we performed differential expression analysis to identify genes with significantly different expression (fold change $\geq$2, FDR $\leq$10%) in PPs with PDX1 mutations compared to controls and found 199, 196 and 112 deregulated genes in $PDX1^{C18R/C18R}$, $PDX1^{P33T/P33T}$and $PDX1^{+/-}$PPs, respectively. Of these, 54, 41 and 20 genes were bound by PDX1, respectively. (Figure 2.12 A-C). Interestingly, we found several genes linked to T2DM risk or glucose resistance, including *MEG3, LARGE1, CES1* and *ANPEP* to be consistently deregulated in all three PDX1 mutants (Figure 2.12 D). Since in both, $PDX1^{P33T/P33T}$and $PDX1^{+/-}$, PDX1 protein levels were reduced, we were specifically interested in the genes, co-regulated in these mutants. Among these genes, we found *MNX1, NNAT* and *PTPRN2,* which code for a pancreatic transcription factor, a proteolipid and a protein tyrosine phosphatase, all of which are relevant for pancreatic development, T2DM and/or insulin secretion [318–321].

2.1.9 Impact of Heterozygous P33T Mutation on PDX1 Binding Patterns

Since the homozygous P33T mutation in isogenic cells severely impaired PP2 cell differentiation and the PDX1 P33T heterozygous donor was diagnosed with gestational diabetes, we characterized the consequences of the heterozygous P33T mutation at the PP stage. To this end, we differentiated the iPSCs, derived from the donor carrying the heterozygous P33T mutation and control subject, into PP cells for 10 days. We then performed ChIP-seq to profile genome-wide PDX1-binding sites in control [297] and donor-derived iPSCs to understand if the binding patterns of PDX1 are perturbed by the heterozygous P33T mutation. We identified 8970 and 8288 PDX1 binding sites in the donor-derived and control PPs, respectively, and motif analysis showed that the most enriched motif was the PDX1 consensus sequence that was found in 71.3% and 66.6% of PDX1-bound sites, respectively (Figure 2.13 A). Binding sites were similarly distributed in both samples, with an enrichment around TSSs and the major fraction of binding sites found in intergenic and intronic regions (Figure 2.13 B-C). A comparison of the PDX1 binding sited in XM001 control and $PDX1^{P33T/+}$PPs showed that a majority of 6444 sites were detected in both samples and a minor fraction of 1844 and 2526 binding sites was called only in XM001 and $PDX1^{P33T/+}$cells, respectively (Figure 2.13 D). However, analysis of the ChIP-seq signal at shared and unique sites revealed enrichment of PDX1 at all sites in both samples, at a slightly smaller magnitude in the cell line, in which the binding sites were not called (Figure 2.13 E). Thus, PDX1 binding patterns in control and mutant PPs are highly similar, which is also reflected in the high correlation between the two profiles (Spearman's ρ=0.688, Figure 2.13 F). Thus, we concluded that the heterozygous P33T mutation does not significantly impair PDX1 binding. Next, we mapped PDX1-binding sites, shared between the control and $PDX1^{P33T/+}$PPs, to genes and identified 3978 potential target genes. Among those, we found important pancreatic genes such as *PDX1, MNX1* or *HNF1B* (Figure 2.13 G-I).

We also profiled H3K27ac enrichment in control (see 2.1.2) and donor-derived PPs by ChIP-seq to see if the PDX1 mutation has an impact on chromatin activity. In line with the previous

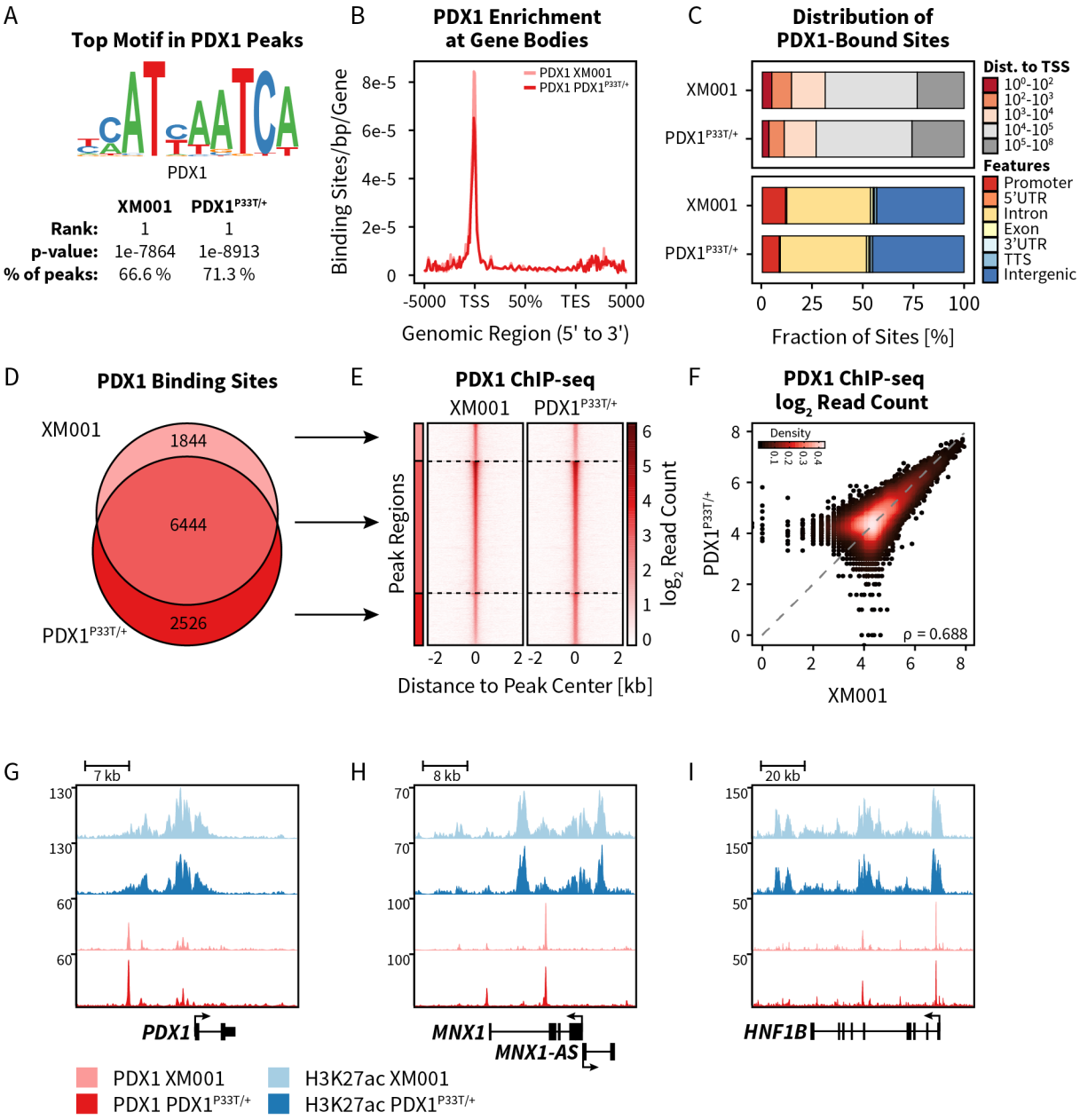

Figure 2.13: Characterization of PDX1 Binding in *PDX1*$^{P33T/+}$PPs. A: Most enriched motif detected by motif analysis resembles the known PDX1 consensus sequence and is present in 71.3% and 66.6% of all PDX1-bound sequences in *PDX1*$^{P33T/+}$and control PPs, respectively. **B:** Meta-genomic plot showing the enrichment of PDX1 at the transcriptional start sites (TSS) of its target genes displayed as binding sites per base pair (bp) per gene over the genomic regions of all RefSeq genes. **C:** Distribution of PDX1 binding sites among genomic features. PDX1 P33T binds predominantly to intergenic and intronic regions, more than 1 kb to the nearest TSS. A significant fraction (∼10%) however, binds to promoter regions less than 1 kb from a TSS. **D:** Venn diagram showing the overlap of PDX1-binding sites in XM001 PPs and *PDX1*$^{P33T/+}$PPs. **E:** Heatmap of PDX1 ChIP-seq signal at all PDX1 binding sites in XM001 and *PDX1*$^{P33T/+}$PPs, showing high resemblance of PDX1 binding in these cells. **F:** Scatter plot comparing the $\log_2$ read counts from PDX1 ChIP-seq in *PDX1*$^{P33T/+}$and XM001. **G-I:** ChIP-seq data tracks showing the enrichment of H3K27ac and PDX1 at the loci of important pancreatic genes in XM001 and PDX1 P33T PPs. Figure modified from [301] under a Creative Commons license.

results, H3K27ac was enriched around PDX1 binding sites (Figure 2.14 A). In both, XM001 control and *PDX1*$^{P33T/+}$PPs, H3K27ac was enriched at almost identical levels (∼5–6×) around PDX1 sites, compared to randomly selected sites of the same length distribution (Figure 2.14 B). Moreover, Spearman's correlation showed that H3K27ac profiles from control and mutant PPs were highly similar (ρ=0.979, Figure 2.14 C), suggesting that in *PDX1*$^{P33T/+}$cells the active chromatin landscape, in terms of H3K27ac, was not significantly altered.

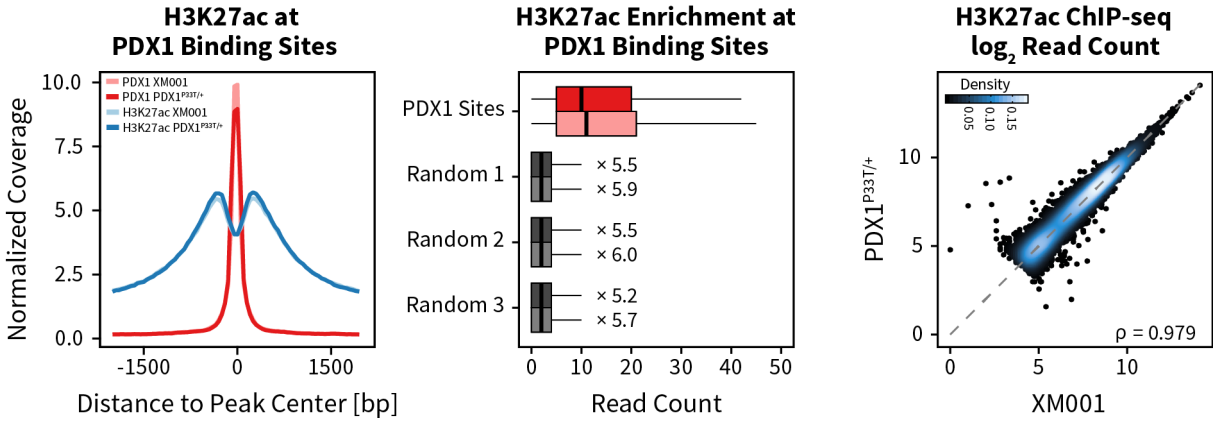

Figure 2.14: Enrichment of H3K27ac in *PDX1*$^{P33T/+}$PPs. A: Comparison of the H3K27ac signal around PDX1 binding sites in *PDX1*$^{P33T/+}$and control. **B:** Read count of H3K27ac at the 8970 and 8288 PDX1 binding sites in *PDX1*$^{P33T/+}$and XM001 PPs, respectively, and at three sets of 8970 and 8288 random sites of the same length distribution. H3K27ac is enriched > 5× at PDX1-bound sites compared to the random sites in both, *PDX1*$^{P33T/+}$and XM001 controls. P-value for all comparisons < 2.2×10^{-16}. **C:** Scatter plot comparing the $\log_2$ read counts from H3K27ac ChIP-seq in *PDX1*$^{P33T/+}$and XM001. Figure modified from [301] under a Creative Commons license.

2.1.10 Transcriptional Profiling of PPs with Heterozygous PDX1 P33T Mutation

Furthermore, we performed Affymetrix microarray analysis of control and donor-derived *PDX1*$^{P33T/+}$cells at the pluripotency and PP stages. Comparing these two stages, we identified 2370 differentially expressed genes (fold change ≥2, FDR ≤5%) including a panel of important pancreatic genes such as *HNF1B*, *SOX9*, *ONECUT1*, *HHEX*, *FOXA2* and *RFX6* that were upregulated in PPs and pluripotency-associated genes including *NANOG* and *LIN28A* that were specifically expressed in iPSCs and downregulated in PPs (Figure 2.15 A). In addition, GO term and pathway analysis revealed a clear enrichment of terms associated with pancreas and endocrine pancreas development in PPs, while pluripotency and stem cell-related terms were enriched in iPSCs (Figure 2.15 B). These results confirm that donor-derived iPSCs can be differentiated into PPs, despite their heterozygous PDX1 P33T mutation.

The next question was whether PPs with P33T mutation are different from control PPs in terms of gene expression. To answer this, we compared the mRNA profiles from *PDX1*$^{P33T/+}$PPs and control PPs and differential expression analysis detected a total of 140 deregulated genes (Figure 2.15 C). Of these genes, the majority were pancreatic genes and 34 had PDX1 binding sites in control and donor-derived cells. Three genes were either experimentally identified (*POSTN* and *NNAT*) or predicted as PDX1 target genes (*SPHKAP*), but of these, only *NNAT* was

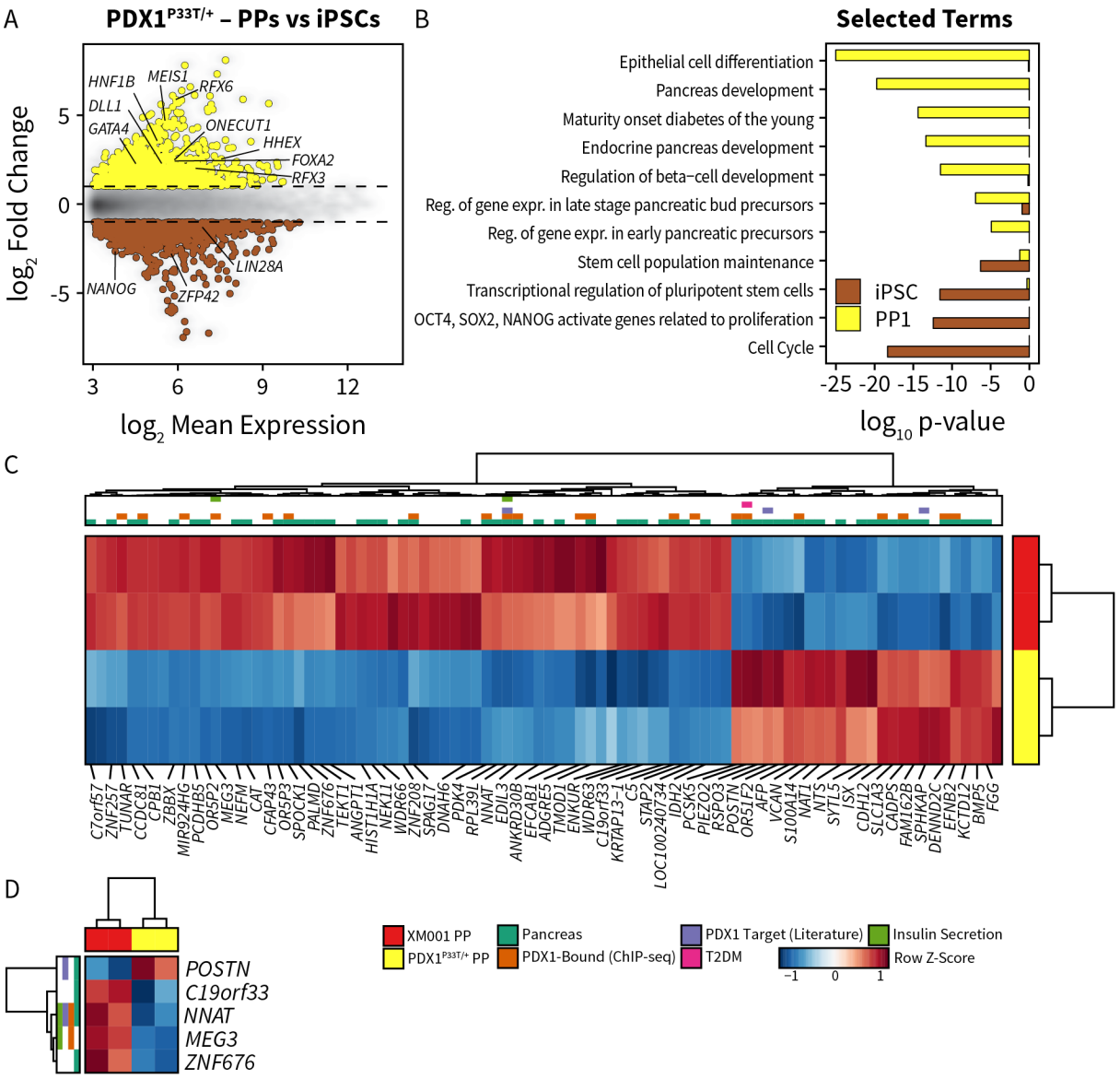

Figure 2.15: Expression Profiling of donor-derived *PDX1*$^{P33T/+}$ PPs. A: MA plot showing the mean $\log_2$ expression against the $\log_2$ fold change of the microarray data obtained from donor-derived iPSCs and PPs. Genes with significantly different expression ($\log_2$ fold change $\geq$1 and adjusted p-value $\leq$0.1) are drawn in color. Yellow depicts increased expression in PPs, whereas brown depicts increased expression in iPSCs. **B:** Bar graph of p-values from pathway enrichment analysis, showing selected GO terms and KEGG and Reactome pathways from differentially expressed genes. **C:** Heatmap of differentially expressed genes between donor-derived *PDX1*$^{P33T/+}$ and XM001 control PPs. **D:** Heatmap showing consistently deregulated genes in donors *PDX1*$^{P33T/+}$ and isogenic *PDX1*$^{P33T/P33T}$ PPs. Figure modified from [301] under a Creative Commons license.

bound by PDX1 in our data set. Periostin which is encoded by the *POSTN* gene, is involved in pancreas and β-cell regeneration [322] and the mouse ortholog of *SPHKAP* is downregulated during pregnancy [323]. Interestingly, among the deregulated genes were also the T2DM and insulin secretion-associated genes *PIEZO2* and *MEG3*.

Next, we were interested if the observed expression changes in heterozygous P33T mutants were in agreement with those observed in homozygous P33T mutants. Thus, we screened differentially expressed genes from both experiments for compliant genes. We identified five genes that were significantly deregulated in the same direction in both experiments, *ZNF676*, *MEG3*, *NNAT*, *C19orf33* and *POSTN* (Figure 2.15 D). Interestingly, from the differentially expressed genes in *PDX1*$^{P33T/+}$PPs, both genes associated with insulin secretion, the only clinically evident symptom in the donor the iPSCs were derived from, *MEG3* and *NNAT* were among the consistently deregulated genes.

2.2 Foxa2 in the Enteroendocrine Lineage Formation

Next to the pancreas, the intestine plays an important role in metabolic control. In response to food intake, enteroendocrine cells (EECs) release a number of hormones that regulate various processes, ranging from food intake and gut motility to insulin secretion [161]. Both, pancreas and intestine are derived from the endoderm during development and the formation of endocrine cells in both systems share some common mechanisms and TF networks [17, 161]. All mature cell types of the intestinal epithelium, including the EECs, but also goblet cells (GCs) and Paneth cells (PCs), as well as the absorptive enterocytes (Ent) are derived from intestinal stem cells (ISCs). They located at the bottom of the crypts and each lineage is directly recruited from the ISCs [276, 279]. However, the decision between an absorptive and a secretory fate is regulated by Notch signaling activity. Thus, the differentiation of all secretory cell types depends on a common mechanism, triggering the suppression of Notch signaling and the expression of the TF Atoh1. Downstream of Atoh1, further cell fate specification is regulated by a number of TFs such as Neurog3 or Spdef, which drive EEC and GC differentiation, respectively. The commitment to different EEC subtypes is regulated through lineage specific TFs like Neurod1, Isl-1, Pax4, Arx and others [324]. The pioneer TF Foxa2, which is a key regulator of endoderm development and the specification of all endoderm-derived organs, is known to be involved in EEC and GC differentiation, as upon loss of Foxa1/2 certain EEC subtypes and functional GCs fail to develop [272]. However, whether Foxa2 is important for ISC function and for the initiation of EEC differentiation is still unknown.

Here, a combination of bulk and single-cell transcriptomics together with ChIP-seq and ATAC-seq is employed to dissect the EECs lineage decision and describe the role of Foxa2 and its impact on the chromatin landscape during EEC differentiation.

2.2.1 Foxa2 Marks ISCs & the Secretory Lineage

The pioneer TF Foxa2 is expressed in endoderm progenitors and all endoderm-derived organs where it plays a key role in lineage priming and acquisition [288, 325]. As both Foxa2 and the ISC marker and R-spondin receptor Lgr5 are canonical Wnt/β-catenin target genes [326], we tested the idea if Foxa2 is a novel marker of ISCs. To approach this, we used, (i) the Foxa2-Venus fusion (FVF) knock-in mouse model [296] to characterize the expression patterns of Foxa2 in the intestinal epithelium and (ii) the Foxa2$^{\text{nEGFP-CreERT2/+}}$;Gt(ROSA)26$^{\text{mTmG/+}}$ lineage-tracing model to mark Foxa2 expressing cells and their progeny.

Using the FVF model, which is an accurate reporter of endogenous Foxa2 expression [296], we observed FVF^{+} cells at the base of the crypt and distributed in a scattered fashion across villi and crypts. In order to define the nature of FVF^{+} cell types, we performed immunofluorescent analysis in combination with cell type specific markers. We found that virtually all Chga^{+} EECs, Muc2^{+} GCs and Lyz1^{+} PCs were also positive for the FVF reporter. However, none of the Dclk1^{+} TCs showed a FVF staining (Figure 2.16). Thus, apart from TCs, all secretory cell types of the intestinal epithelium express Foxa2.

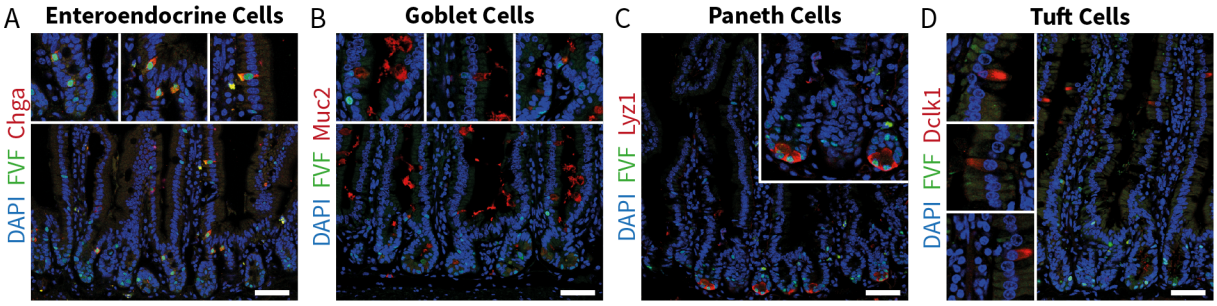

Figure 2.16: Foxa2 Expression in EECs, GCs, PCs and TCs. A-D: Confocal laser scanning microscopy (CLSM) images of DAPI (blue), FVF (green) and cell type specific markers (red) in the duodenum. Scale bars, 50 µm. Stainings were performed by Alexandra Aliluev and Anika Böttcher.

Moreover, we found FVF^+, slender, column-shaped cells between PCs at the crypt base, reminiscent of ISCs (Figure 2.17 A). To test if these FVF^+ cells are indeed ISCs, we performed a pulse-chase genetic lineage tracing experiment using a $Foxa2^{nEGFP\text{-}CreERT2/+}$;$Gt(ROSA)26^{mTmG/+}$ mouse line to mark Foxa2-expressing cells and their progeny. In this double-fluorescent reporter mouse model, one copy of the *Foxa2* gene is fused to a *2A-nEGFP-2A-CreET2* cassette, expressing a tamoxifen inducible $CreER^{T2}$ [327]. In addition, a membrane-targeted tandem dimer Tomato (mT) is ubiquitously expressed from a floxed trans gene in *ROSA26* locus. Upon Cre-mediated excision of the floxed *mT* gene, a membrane-targeted green fluorescent protein (mG) is expressed from a trans gene, immediately downstream of *mT* [328]. Hence, cells that express Foxa2 at the time of induction and all their descending cells are marked by a green fluorescent membrane (Figure 2.17 B). First, we analyzed recombination 48 hours after a single dose of orally administered tamoxifen and could detect recombination, marked by GFP staining in the membrane, in single cells and small patches of 2-3 cells that were scattered across the entire epithelial lining. Next, we assessed recombination 30 days after tamoxifen induction, which revealed ribbon-like patterns of recombined mG^+ cells, typical for ISC lineage tracings [205, 326] (Figure 2.17 C). At this time point, single recombination events, outside of ribbon structures, were restricted to long-lived Paneth cells at the crypt base. To confirm that the identified ribbons contain all lineages of the intestinal epithelium, we co-stained mG with cell type specific markers and identified $Chga^+$ EECs, $Muc2^+$ GCs and $Lyz1^+$ PCs as well as $Dclk1^+$ TCs within the Foxa2-lineage ribbons (Figure 2.17 D-G).

Collectively, we could show that in the epithelium of the small intestine, Foxa2 is expressed in ISCs, EECs, PCs and GCs. On the other hand, we did not detect expression of Foxa2 in enterocytes or TCs.

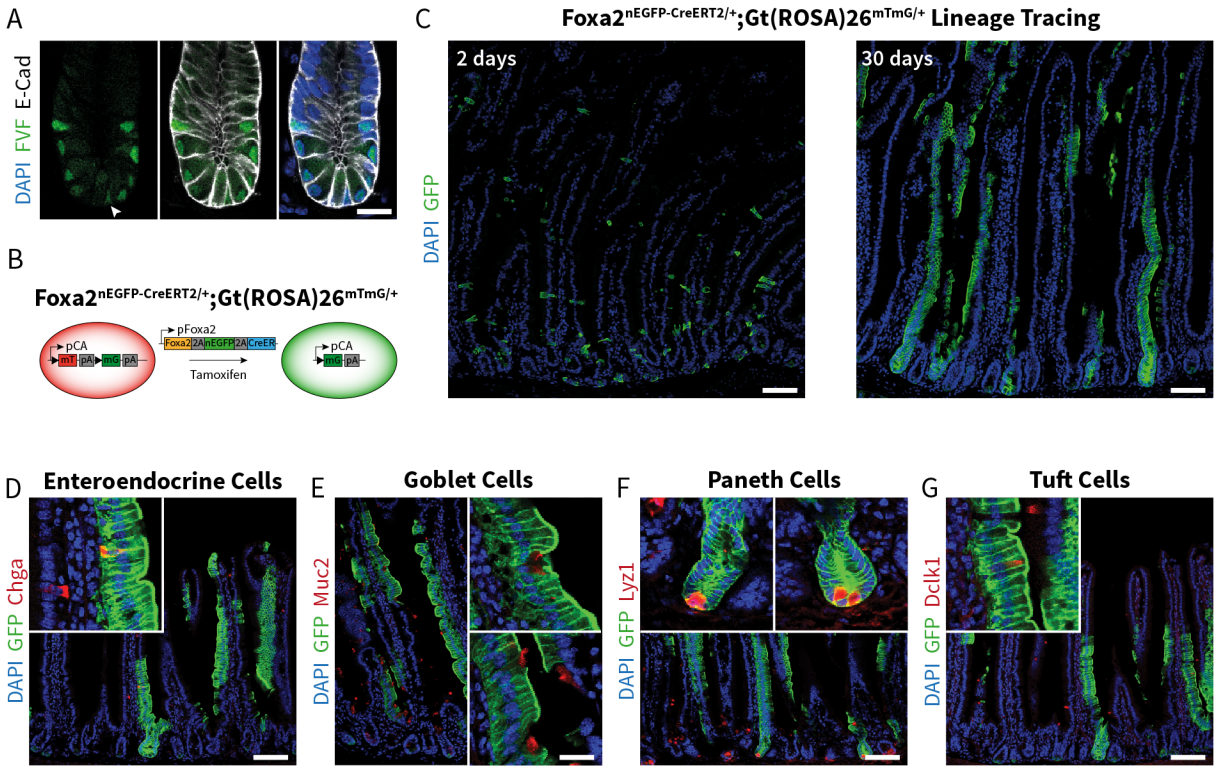

Figure 2.17: Foxa2 is Expressed in ISCs. **A:** CLSM images of DAPI (blue), FVF (green) and E-cadherin (white) in a SI crypt. Arrow head points to a putative FVF^{+} ISC. Scale bar, 25 µm. **B:** Schematic representation of the $Foxa2^{nEGFP\text{-}CreERT2/+}$;$Gt(ROSA)26^{mTmG/+}$ lineage-tracing model. Cells are red in the default state and irreversibly convert into green cells when Foxa2 is expressed and Tamoxifen is present. **C:** Representative LSM images of DAPI (blue) and recombined cells (green) in the duodenum of $Foxa2^{nEGFP\text{-}CreERT2/+}$;$Gt(ROSA)26^{mTmG/+}$ mice, 48 hours and 30 days after a single dose of Tamoxifen. Scale bars, 100 µm. **D-G:** Representative LSM images of DAPI (blue), recombined cells (green) and cell type specific markers (red). Scale bars, 100 µm. Lineage tracing experiment and stainings were performed by Alexandra Aliluev.

2.2.2 Foxa2 Expression Levels Discriminate Intestinal Lineages

Using the FVF mouse model, we noticed that Foxa2 is expressed at different levels within the crypt. Cells expressing lower levels of Foxa2 were found at the bottom of the crypt, while cells expressing higher levels of Foxa2 were positioned mainly above the crypt bottom (Figure 2.18 A). We therefore hypothesized that the levels of Foxa2 expression might distinguish between different types of $Foxa2^{+}$ cells. Flow cytometry analysis of cells, isolated from small intestinal crypts, showed two populations of $Foxa2^{+}$ cells with mean fluorescence values of 2413 ± 35 and 5581 ± 95, referred to as Foxa2 low (FVF^{low}) and Foxa2 high (FVF^{high}), respectively (Figure 2.18 B-D).

To characterize these cells, we used FACS to sort FVF^{neg} (Foxa2 negative), FVF^{low} and FVF^{high} cells and subjected each of these populations to mRNA profiling by Affymetrix microarray (Figure 2.19 A). Differential expression analysis revealed sets of 1137, 312 and 928 genes that are specifically expressed in FVF^{neg}, FVF^{low} and FVF^{high} cells, respectively (fold change ≥ 2, FDR $\leq 5\%$). Among these population-specific genes, we found several well-known marker genes, characteristic for the different cell types of the intestinal epithelium. We found marker genes

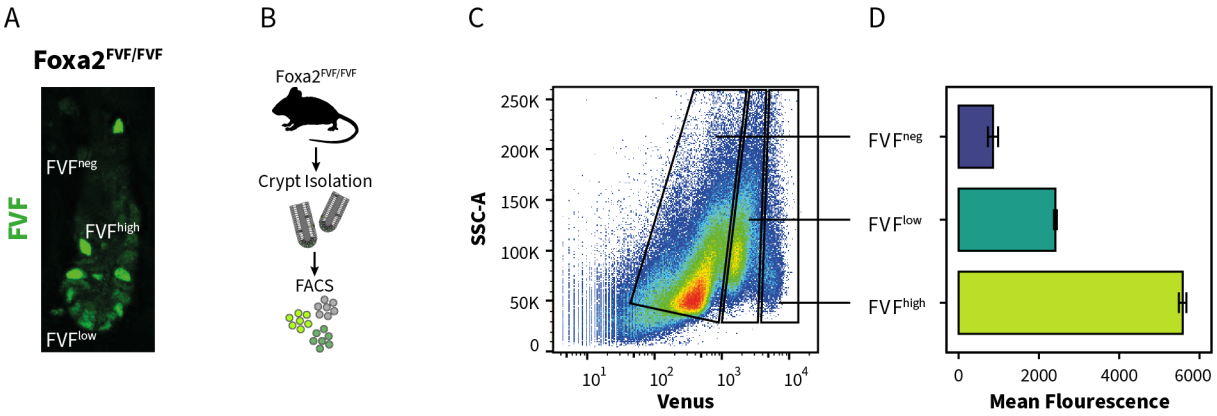

Figure 2.18: Foxa2 Expression Levels Define Distinct Cell Populations Within the Crypt. Foxa2 is expressed at different levels within the small intestinal crypt. **A:** The FVF-reporter shows cells with low (FVF^{low}) and high (FVF^{high}) expression of Foxa2. **B:** FACS scheme for the analysis of FVF^{+} cells. **C:** FACS data shows three distinct populations, separated by their expression of Foxa2 (FVF^{neg}, FVF^{low} & FVF^{high}). **D:** The mean fluorescence values of the FVF populations show a 2.3×increase in Foxa2 expression between FVF^{low} and FVF^{high} cells.

for Enterocyte (*Alpi*, *Apoa1*, *Fabp1*, *Sis*, *Lct*) and TCs (*Dclk1*, *Trpm5*) to be expressed in FVF^{neg} cells, while marker genes for ISCs (*Lgr5*, *Ascl2*, *Gkn3*) as well as PCs (*Lyz1*, *Mmp7*, *Ang4*) were expressed in FVF^{low} cells. FVF^{high} cells were enriched in genes known to be involved in EEC differentiation (*Neurod1*, *Neurog3*, *Nkx2-2*). Most GC genes such as *Agr2*, *Spdef*, *Spink4* or *Tff3* were enriched in both, FVF^{low} and FVF^{high}, compared to FVF^{neg} cells. However, *Dll1* and *Dll4*, associated with pan secretory but also specifically GC differentiation were enriched in FVF^{high} (Figure 2.19 B-D). The expression of several ISC, EEC and Ent marker genes, as well as the expression of Foxa genes was also determined by quantitative PCR (qPCR) (Figure 2.19 E-H). These results show that with the exception of GC genes that are expressed in FVF^{low} and FVF^{high} cells, other cell type-specific markers are enriched in a single FVF population, suggesting that the different cell types can be distinguished by their expression levels of Foxa2.

To gain insight into the function of genes associated with the different FVF populations, GO-term and pathway enrichment analysis was performed. In line with the expression of enterocyte marker genes, FVF^{neg} cells showed enrichment of terms related to metabolic and catabolic processes, digestion, absorption and the brush border membrane of enterocytes. Genes enriched in the FVF^{high} population, were linked to neuroendocrine differentiation and function, exocytosis and negative regulation of cell proliferation. In contrast, there was no specific term found for genes enriched in FVF^{low} cells. The most significant term was related to secretory vesicles but also showed enrichment in FVF^{high} cells, although to a lesser extent. Other terms enriched in FVF^{low} were related to fat digestion and absorption or signal release, but these show considerably higher enrichment in FVF^{neg} and FVF^{high}, respectively. (Figure 2.20). Thus, pathway analysis supports the finding that the different cell types in the intestinal epithelium can be discriminated by their Foxa2 expression levels.

Based on the expression patterns of intestinal marker genes and enriched pathways, we concluded that FVF^{neg} cells consist of Enterocytes and TCs, FVF^{low} cells of ISCs and PCs and FVF^{high}

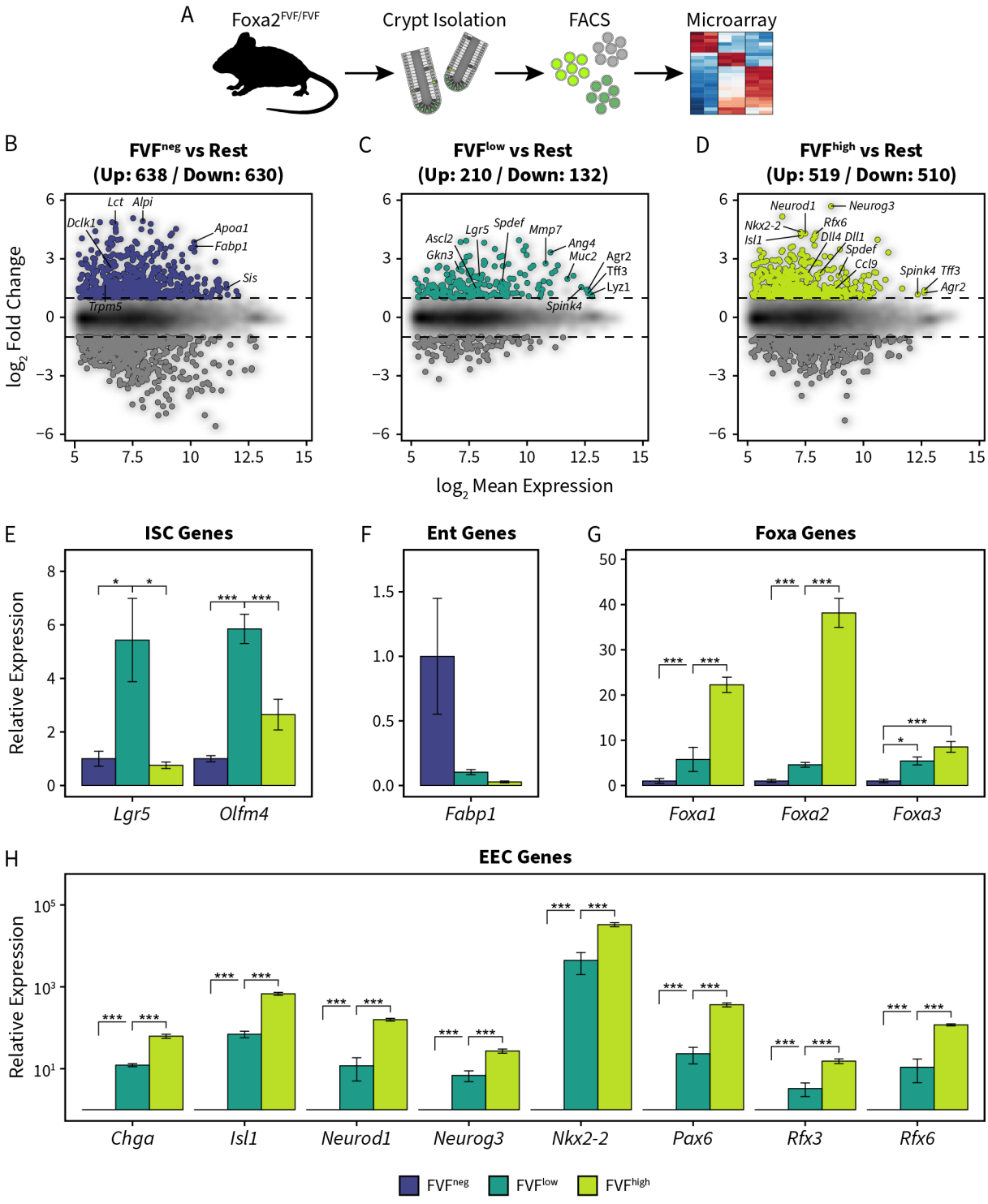

Figure 2.19: Population-Specific Gene Expression in the Intestinal Crypt. A: Experimental outline for mRNA profiling of FVF populations by microarray. **B-D:** MA-plots showing the mean $\log_2$ expression against the $\log_2$ fold change of each FVF population compared to the average of the other populations (rest). Significantly enriched genes in FVF^{neg} (B), FVF^{low} (C) and FVF^{high} (D) cells, respectively, are highlighted in color. **E-H:** qPCR analysis of the selected marker genes for ISCs (E), Enterocytes (F) and EECs (H) as well as Foxa genes (G) confirm microarray results. ANOVA, *: p<0.05, **: p<0.01, ***: p<0.001. Errorbars, standard error of the mean (SEM).

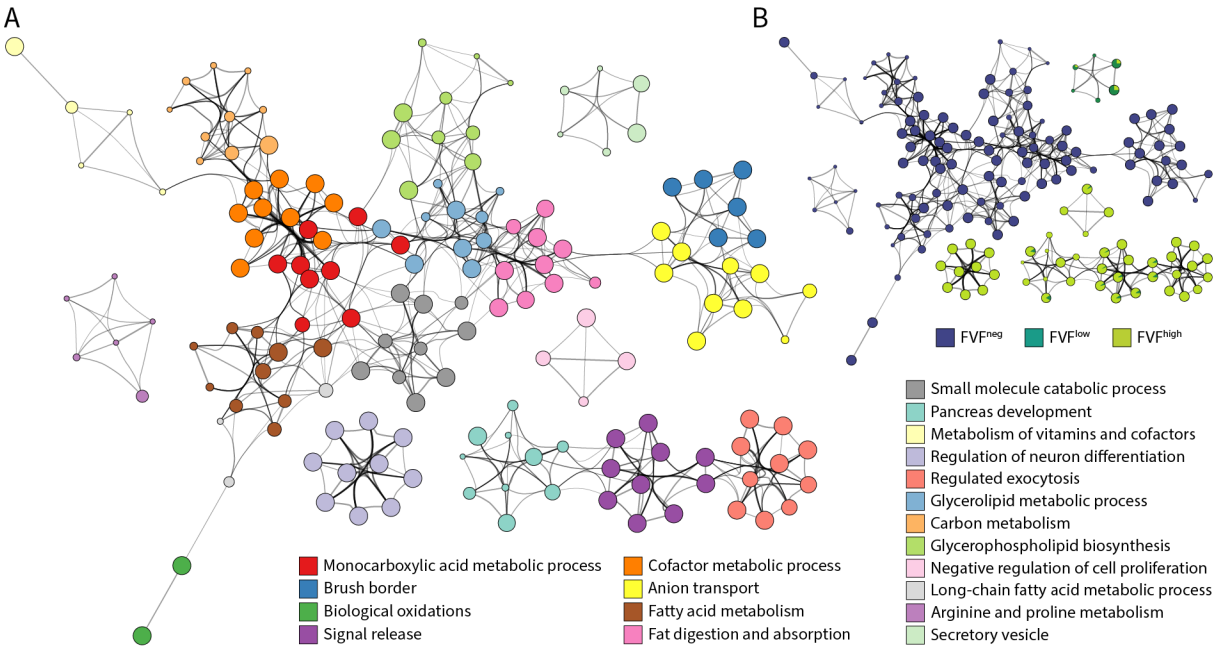

Figure 2.20: Gene Function Enrichment Analysis of Population Specific Genes. Pathway and process enrichment analysis of genes with population specific expression. **A:** Network representation of enriched GO-terms and pathways. Each node represents an enriched term and its size is proportional to the total number of hits (from all FVF populations) falling in that specific term. Terms are grouped into clusters based on their membership similarities. Nodes are colored by their cluster affiliation and clusters are named by the most significant term within the cluster. Terms with a similarity greater than 0.3 are connected by edges. **B:** Nodes are represented by pie-charts, color coded by FVF populations, where the portions of the pie represent the proportion of genes originating from a specific FVF populations.

cells mainly of EECs and their progenitors. The expression of GC genes, in both FVF^{low} and FVF^{high}, indicates that GC express intermediate Foxa2 levels and are therefore present in both populations (Figure 2.21). The expression of *Dll1* and *Dll4* in FVF^{high} cells, however, suggests that GC progenitors might be found in the FVF^{high} population. Thus, the FVF marker is a useful tool to enrich for low abundance cell types such as EECs or ISCs, by sorting FVF^{+} cell and to dissect the EEC lineage formation by analyzing FVF^{low} and FVF^{high} cells.

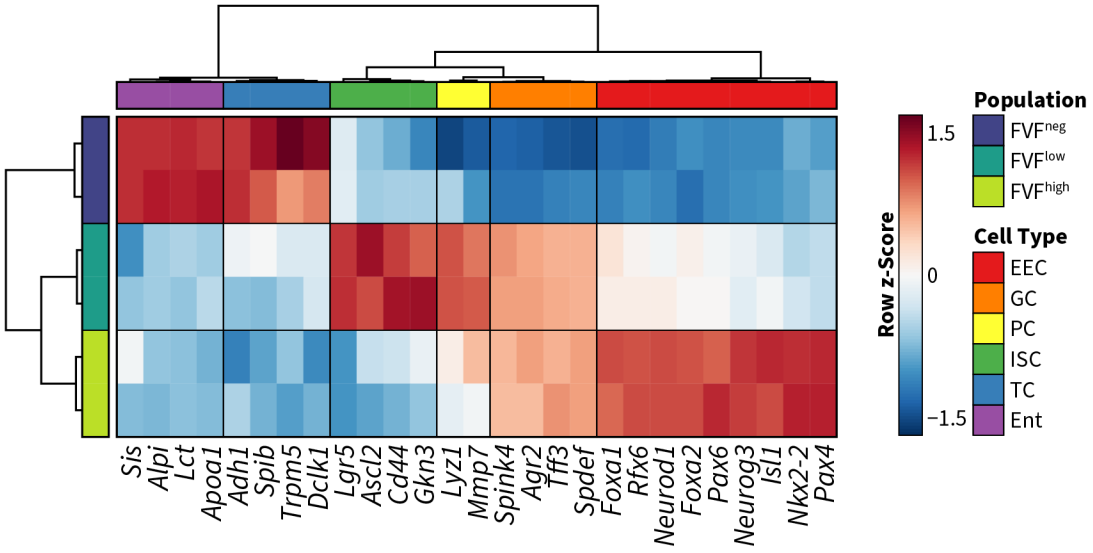

Figure 2.21: Marker Gene Expression Patterns in FVF Populations. Heatmap showing the expression of several marker genes for the different intestinal lineages.

2.2.3 Active Histone Modifications are Similar in FVF^{low} & FVF^{high} cells

The gradual 2.3-fold increase of the FVF reporter allowed us to monitor EEC recruitment from ISCs and to gain further insight into the lineage decision process. To examine the chromatin dynamics underlying the enteroendocrine lineage formation, we profiled the active promoter and enhancer landscape by performing ChIP-seq of the histone modifications H3K4me3 and H3K27ac on FACS purified FVF^{low} and FVF^{high} cells (Figure 2.22 A). H3K4me3 is associated with active promoters, while H3K27ac marks both, active enhancers and promoters [304–307, 329, 330]. In concordance, levels of H3K4me3 and H3K27ac at transcription start sites (TSS) correlated with the expression levels of the respective genes. Spearman's ρ was higher for the correlation between H3K27ac enrichment and expression levels than for H3K4me3 ($\rho_{H3K27ac} = 0.56$ and $\rho_{H3K4me3} = 0.45$), which suggests that H3K27ac is a more accurate predictor of genes expression (Figure 2.22 B-C). We then compared H3K4me3 and H3K27ac between FVF^{low} and FVF^{high} cells on a genome-wide scale and found a striking similarity between the two cell populations, showing that changes are few and on a small scale (Figure 2.22 D-F).

However, we found a positive correlation between the change in H3K27ac levels at transcription start sites and the change in expression of the corresponding gene (Figure 2.22 G). Thus, even though there were mostly minor differences between the FVF^{low} and FVF^{high} population, the observed changes correlate well with changes in gene expression. Among, the loci with the most striking differences in H3K4me3 and H3K27ac were those of endocrine lineage determinants such as *Neurog3* and *Neurod1* which were also among the most upregulated genes in FVF^{high} cells or the PC marker *Ang4* which is specifically expressed in FVF^{low} cells (Figure 2.22 H-J).

The observed similarities on the level of active histone modifications were in line with previous findings, showing a high similarity of H3K4me2 and H3K27ac patterns in ISCs, secretory (Sec Pro) and enterocyte (Ent Pro) progenitor cells [222, 224]. In a heterogeneous cell population with

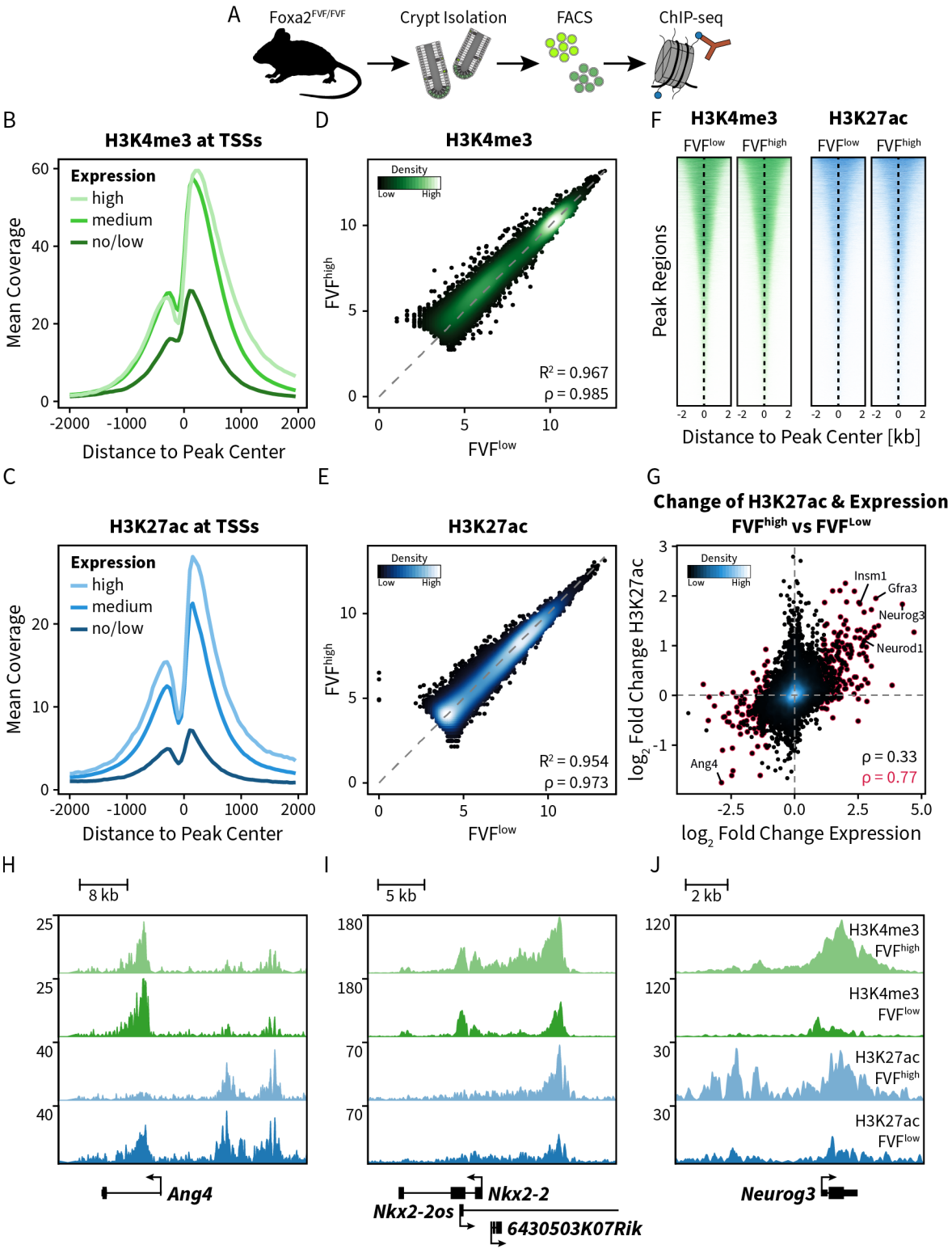

Figure 2.22: Active Histone Modifications are Highly Similar in FVF^{low} and FVF^{high} cells. Caption on next page.

Figure 2.22: Active Histone Modifications are Highly Similar in FVF^{low} and FVF^{high} cells (continued). A: Experimental outline for profiling histone modifications in FVF^{low} and FVF^{high} populations by ChIP-seq. **B-C:** H3K4me3 (B) and H3K27ac (C) coverage profile of the averaged signal from FVF^{low} and FVF^{high} in a 2 kb regions around TSSs shows an enrichment of H3K4me3 and H3K27ac at TSSs that is increasing with increasing gene expression (mean expression from FVF^{low} and FVF^{high}). Compared to H3K4me3, the correlation of H3K27ac with the expression is higher ($\rho_{H3K27ac}$ = 0.56 and $\rho_{H3K4me3}$ = 0.45). **D-E:** Scatter plots comparing the $\log_2$ read counts from H3K4me3 (D) and H3K27ac (E) peaks in FVF^{low} and FVF^{high} cells. **F:** Enrichment heatmap showing the high concordance between H3K4me3 and H3K27ac profiles in FVF^{low} and FVF^{high}. **G:** Scatter plot comparing $\log_2$ fold changes between FVF^{high} and FVF^{low} in gene expression (all genes covered by the microarray) against H3K27ac at the respective gene promoter. Points marked in red correspond to genes with significantly different expression. Spearman's correlation of significant genes only is written in red. **H-J:** ChIP-seq data tracks showing H3K4me3 and H3K27ac from FVF^{low} and FVF^{high} cells at the *Ang4* (H), *Nkx2-2* (I) and *Neurog3* (J) locus as examples for loci that exhibit pronounced changes in H3K4me3 and H3K27ac.

such a homogeneous configuration of active histone modifications, cell fate decisions are made based on the presence or absence of lineage determining TFs that have to be tightly regulated. This generally permissive chromatin landscape is likely the basis of the high plasticity observed in the intestinal crypt [222].

2.2.4 Selective Chromatin Accessibility Reflects Lineage Recruitment

Due to the high similarity of the chromatin configuration on the level of the profiled histone modifications, replicate ChIP-seq experiments are necessary for accurate quantification of such small differences. Owing to the small numbers of cells available from the FVF^{high} population, it was not feasible to perform the required replicate experiments. To overcome this limitation, we assayed chromatin accessibility by ATAC-seq as approximation of chromatin activity in purified FVF^{low} and FVF^{high} cells (Figure 2.23 A), which requires about two orders of magnitude fewer cells. ATAC-seq utilizes the Tn5 transposase, loaded with adapters for high-throughput sequencing, for the simultaneous fragmentation of genomic DNA and the ligation of sequencing adapters. The Tn5 transposase preferentially integrates its adapters into regions of high chromatin accessibility and yields amplifiable fragments that can be directly used for sequencing, making the method very efficient in terms of material requirements. Initial analysis of ATAC-seq data revealed the typical periodic length distribution of the transposed fragments, which results from the protective effect of nucleosomes, inhibiting transposition of nucleosome-bound DNA sequences and the pitch of the DNA helix. The first peak of small fragments below $\sim$150 bp corresponds to nucleosome-free regions of active chromatin. The larger peaks with sequentially decreasing height and correspond to nucleosome-bound regions. Consecutive peaks result from mono-nucleosomes, di-nucleosomes and tri-nucleosomes [331] (Figure 2.23 B). Similarly, when decomposed into the nucleosome-free and nucleosome-bound fractions, the ATAC-seq signal shows a characteristic pattern around active TSSs, caused by phasing of nucleosomes at actively transcribed genes [332]. The nucleosome-free fragments accumulate in a sharp peak centered at the TSSs, while the nucleosome-bound fragments are depleted from the TSSs and accumulate in periodic peaks around phased upstream and downstream nucleosomes [331] (Figure 2.23 C).

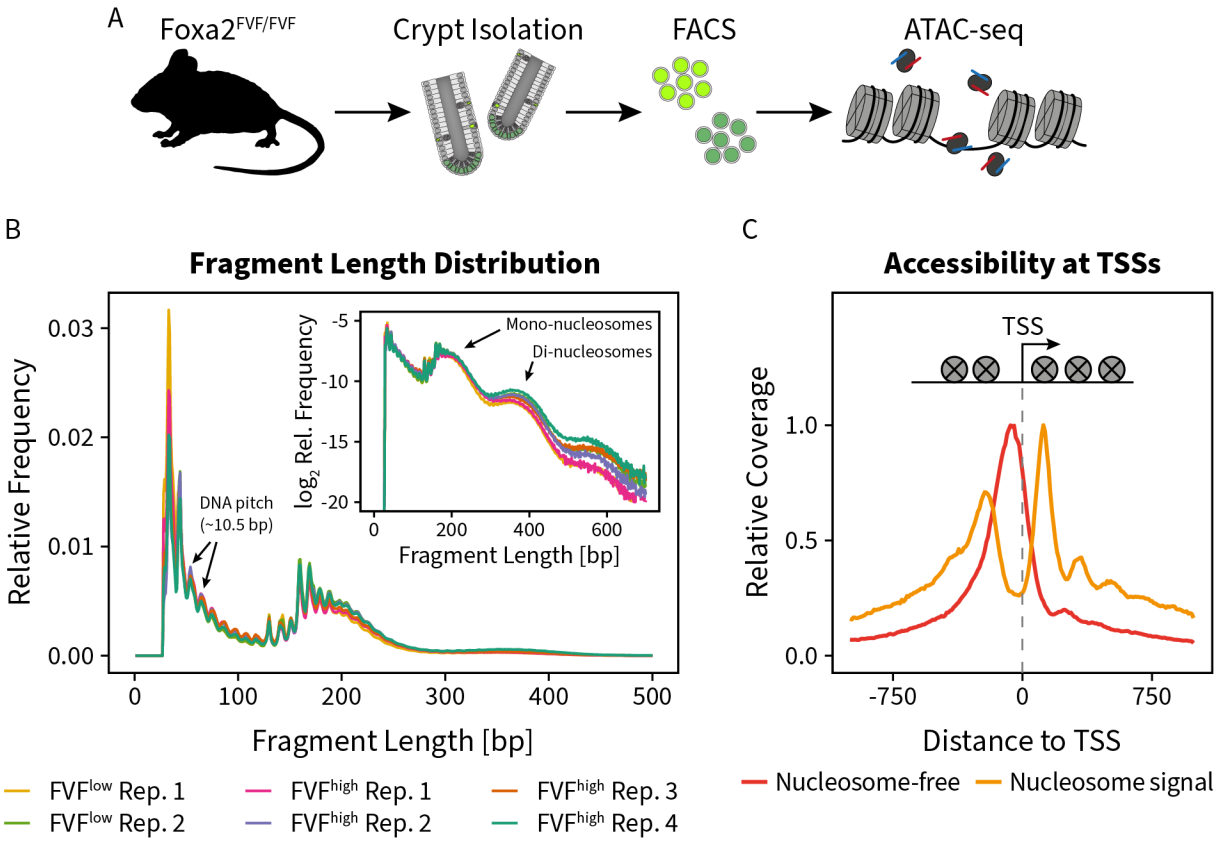

Figure 2.23: Primary ATAC-seq Analysis. A: Experimental outline for accessibility profiling of FVF populations by ATAC-seq. **B:** Fragment length distribution of ATAC-seq profiles from different biological replicates. Distribution patterns show the typical, consecutive peaks of nucleosome-free and nucleosome-bound DNA, as well as a good consistency between biological replicates. **C:** Relative ATAC-seq coverage around active TSSs shows accumulation of nucleosome-free fragments at the TSSs, flanked by a periodic signal from nucleosome-bound fragments that is the result of phased nucleosomes around active TSSs.

To identify differences in chromatin opening between FVF^{low} and FVF^{high} cells, we analyzed ATAC-seq data using DESeq2 [333]. Differential enrichment analysis identified 1344 regions with significant differences (s-value < 0.1) in accessibility between FVF^{high} and FVF^{low}. The majority of differential sites (1103) were selectively open in FVF^{high} cells and genes in the vicinity include EEC and goblet cell differentiation genes such as *Neurog3*, *Neurod1* or *Spdef*. On the other hand, only 241 regions showed a higher accessibility in FVF^{low} cells. However, consistent with the finding that the FVF^{low} population contains ISCs, genes near FVF^{low} specific open sites included the ISC marker genes *Olfm4* and *Ascl2* (Figure 2.24 A). Interestingly, when compared to the bulk of open chromatin regions, selectively open regions were almost exclusively found in intergenic and intronic regions that are more than 1 kb away from the closest TSS. In contrast, among all accessible sites approximately one out of four is found within 1 kb of a TSS (Figure 2.24 B). In fact, only 39 differentially accessible sites were found in a promoter region (1 kb upstream and 100 bp downstream of a TSS), suggesting that regulation of gene expression is achieved through distal regulatory elements rather than through their promoters (Figure 2.24 A). To investigate the function of genes near the sites with changed accessibility, we performed gene function

enrichment analysis. Genes associated with increased chromatin opening in FVF^{high} cells have a function in the regulation of secretion and integration of energy metabolism as well as in pathways involved in endocrine and neuron differentiation. Neuronal differentiation genes often have a function also in endocrine differentiation and include some of the most important endocrine decision maker genes such as *Neurog3* and *Nkx2-2* [161]. Moreover, genes from the Mapk signaling pathway were also associated with increased chromatin accessibility. This is in line with the finding that Mapk signaling is involved in the differentiation of enteroendocrine cells *in vitro* [240] (Figure 2.25).

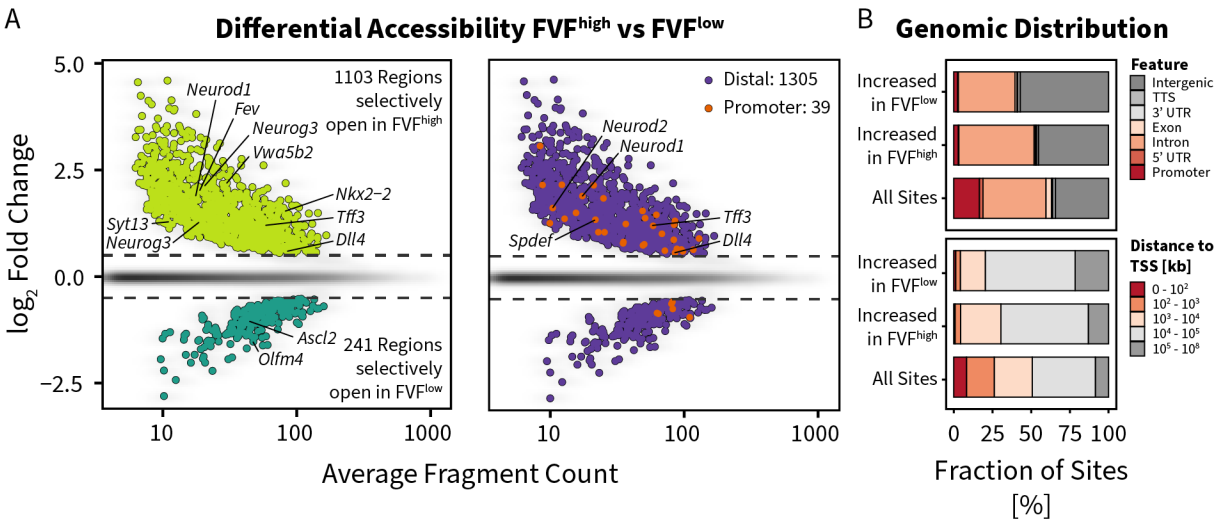

Figure 2.24: Differential Chromatin Accessibility in FVF^{low} and FVF^{high} Cells. A: MA plot showing differentially accessible sites in FVF^{low} and FVF^{high} cells. Sites with significant differences are shown in color. In the left panel, light and dark green depict sites with higher accessibility in FVF^{high} and FVF^{low}, respectively. Selected relevant genes with differentially accessible sites in their vicinity (20 kb up/downstream of the TSS or within gene body) are indicated. On the right, sites are colored according to the location in the genome. Significantly different sites, overlapping promoters (-1 kb to 100bp of a TSS), are marked in orange, all other significant sites are considered as distal and shown in purple. **B:** Distribution of accessible sites in terms of genomic features and distance to the nearest TSS.

Cooperative binding of sequence-specific TFs to regulatory elements in the genome is characterized by markedly increased DNA accessibility [334]. Using the consensus sequence information, it is possible to infer the TF-associated accessibility in ATAC-seq data. One commonly used technique is the analysis of TF footprints which emerge from TF binding events that protect the DNA strand from being cleaved [335]. However, not every TF-chromatin interaction is stable enough to cause footprints in accessibility data and samples have to be sequenced at great depth in order to accurately detect footprints [335]. To avoid these limitations, we employed a recently developed approach, chromVAR, that measures the gain or loss of chromatin accessibility within peaks sharing a given motif or annotation and allows the *de novo* discovery of motifs associated with changes in chromatin accessibility [331, 336]. We used a combination of *de novo* motif discovery and analysis of known motifs from TFs expressed in our microarray dataset to identify TFs associated with variation in accessibility. Using the *de novo* approach, we found eight motifs that were significantly associated with alterations of chromatin opening in FVF^{high} compared to FVF^{low}. Three motifs, a basic helix-loop-helix (bHLH) similar to the Neurod1 motif, a forkhead

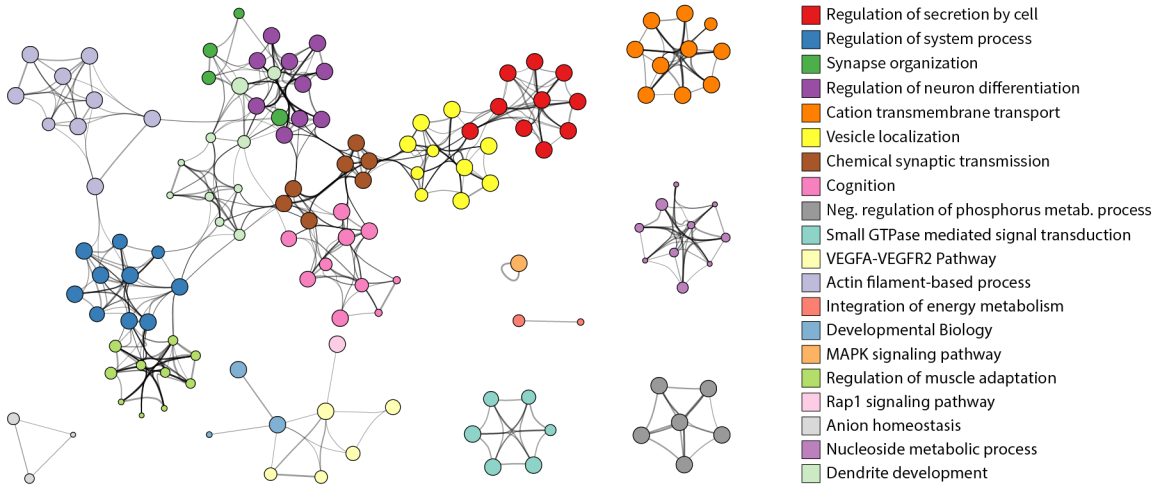

Figure 2.25: Gene Function Near Differential Accessibility. Network representation of enriched GO-terms and pathways of genes near FVF^{high} specific open chromatin regions. Each node represents an enriched term and its size is proportional to the total number of hits falling in that specific term. Terms are grouped into clusters based on their membership similarities. Nodes are colored by their cluster affiliation and clusters are named by the most significant term within the cluster. Terms with a similarity greater than 0.3 are connected by edges.

(FKHD) motif similar to the Foxa1/2 consensus and a regulatory factor X-type (RFX) motif were associated with increased chromatin accessibility in FVF^{high} cells and five motifs were linked to chromatin opening in FVF^{low} cells. These included a nuclear receptor (NR) motif reminiscent of the Hnf4 or Tcf7l2 motif, a GATA like motif, a basic leucine zipper (bZIP) motif, a interferon consensus sequence (IRF) and a zink finger (ZF) motif (Figure 2.26 A). Similar results were obtained by the analysis of known TF motifs that also link bHLH, forkhead and RFX motifs but also some homeobox motifs to accessibility in the FVF^{high} population. Nuclear receptor, bZIP, GATA and Sox motifs were found in sites with increased chromatin opening in FVF^{low} cells (Figure 2.26 B). Thus, TFs like Hnf4a and Gata4 known to be important for intestinal patterning (Gata4 [337–340]) and the differentiation of enterocytes (Hnf4a [341]) were linked to chromatin opening in the FVF^{low} population, containing ISCs and early enterocyte progenitors, while well-known endocrine TFs such as Neurog3, Neurod1, Rfx6 and Foxa1/2 or Atoh1, which were important for all intestinal secretory cells and especially Goblet cells, are strongly associated with chromatin opening in FVF^{high} cells.

Together, the ATAC-seq data substantiate the finding that FVF^{high} cells are differentiating towards the goblet and endocrine lineage and suggests that Forkhead and bHLH TFs, such as Foxa2, Neurod1, Neurog3 and Atoh1, contribute to this process. In this case, Foxa2 might act as a pioneering factor for transcription factors that are important for goblet and enteroendocrine lineage differentiation in FVF^{high} cells, as reflected by the motif analysis in sites of selectively open chromatin. In contrast, motifs, associated with TFs that are not explicitly involved in secretory cell fate, such as Hnf4a, correlate with loss of chromatin accessibility in FVF^{high} cells.

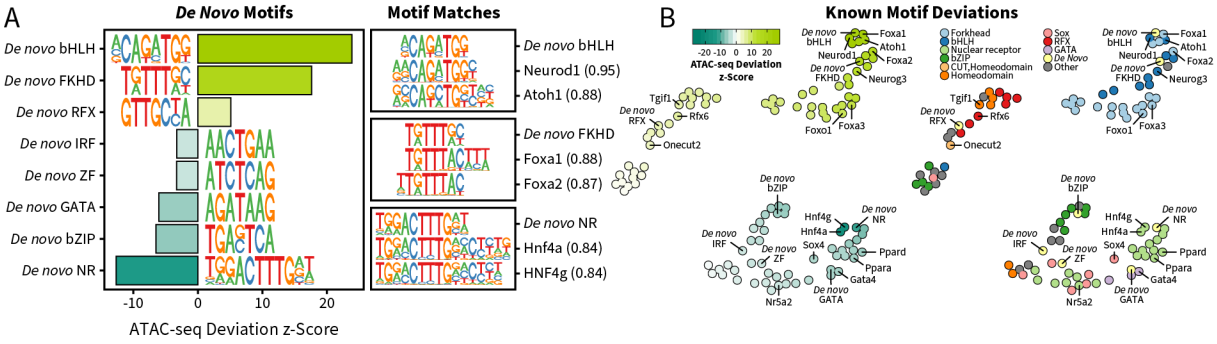

Figure 2.26: Analysis of Motifs Associated with Variation in Chromatin Accessibility. A: *De novo* discovery of motifs associated with variations in chromatin opening in FVF^{low} (dark green) and FVF^{high} (light green) cells. Bar plot on the left shows ATAC-seq deviation scores of each discovered motif as well as the motif itself. Motif matches on the right show the 3 motifs with most extreme deviation scores with their two best matching, non-redundant known motifs. **B:** UMAP representation of known TF motifs and their associations to chromatin accessibility changes. On the left, motifs are colored and distributed according to their ATAC-seq deviation score. Color code as in A. On the right, motifs are colored according to the TF family affiliation.

2.2.5 Foxa2 Binds Important Intestinal Genes

The pioneer TF Foxa2 is known to be involved in the differentiation of EECs and goblet cells [272]. Our ATAC-seq data link Forkhead TFs with EEC differentiation in FVF^{high} cells and thus, to high levels of Foxa2 expression. Therefore, we set off to characterize genome wide Foxa2 binding patterns in order to investigate whether the rising expression levels of Foxa2 are associated with binding and activation of cis-regulatory regions of EEC lineage determining factors.

To address these questions, we performed ChIP-seq of Foxa2 using an antibody specific to the Venus fluorescent tag in both, FVF^{low} and FVF^{high} cells (Figure 2.27 A). We identified 10984 Foxa2 binding sites in FVF^{low} cells and 18838 binding sites in FVF^{high} cells. The distribution of the Foxa2 binding sites in both population was similar and showed a preference for intronic ($\sim$45%) and intergenic ($\sim$41%) regions (Figure 2.27 B). However, with $\sim$10% of the binding sites occurring in promoter regions, Foxa2 binding sites were enriched around TSSs (Figure 2.27 C-D).

Moreover, we found the active histone modifications H3K4me3 and H3K27ac enriched around Foxa2 bound sites in FVF^{low} and FVF^{high}. Similarly, chromatin accessibility as measured by ATAC-seq was increased at Foxa2 bound sites (Figure 2.27 E). We compared the enrichment of H3K4me3, H3K27ac and ATAC-seq in a 1 kb region around all 20377 Foxa2-bound sites in FVF^{low} and FVF^{high} to a set of 20377 randomly chosen 1 kb regions to quantify the increase of active chromatin features around Foxa2 sites. This analysis revealed that on average H3K4me3, H3K27ac and ATAC-seq were 12.5×, 9.8× and 11.6× more abundant at Foxa2 binding sites compared to the randomly chosen background (Figure 2.27 E).

To test whether the known Foxa2 consensus motif was enriched in our Foxa2-bound sites, we performed motif analysis in the identified peaks from FVF^{low} and FVF^{high} cells. As expected, we identified the known Foxa2 consensus sequence as primary motif in peak regions from both, FVF^{low} and FVF^{high} cells. The Foxa2 motif was found in 68.8% and 69.1% of the Foxa2 peaks, confirming the high quality and specificity of the ChIP-seq data (Figure 2.27 F).

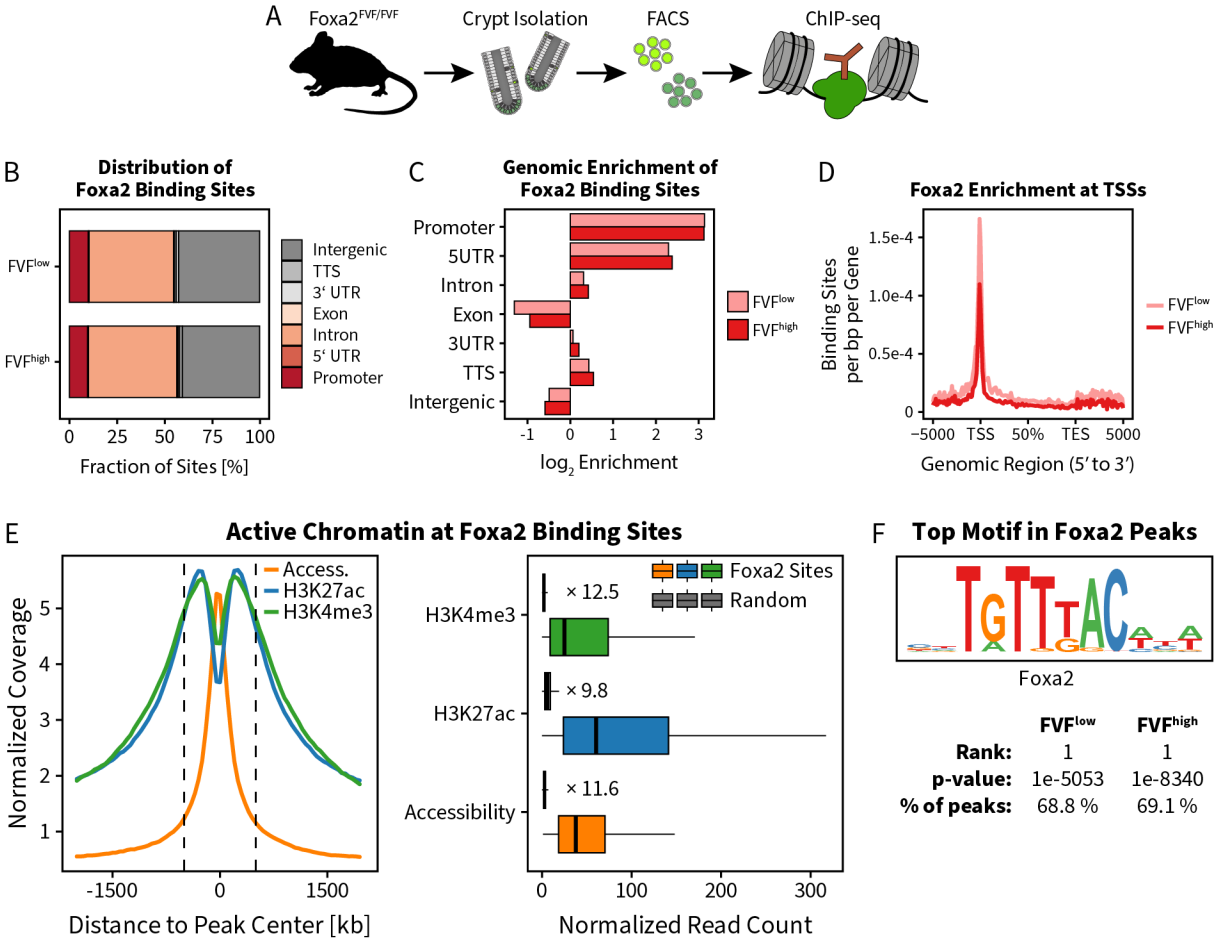

Figure 2.27: Distribution of Foxa2 Binding Sites in FVF^{low} and FVF^{high} Cells. A: Experimental outline for profiling Foxa2 binding in FVF^{low} and FVF^{high} cells by ChIP-seq. **B:** Distribution of Foxa2 binding sites across genomic features in FVF^{low} and FVF^{high} cells. **C:** Log_2 enrichment of Foxa2-bound sites, showing that Foxa2 is enriched mainly at promoters and 5' UTRs but depleted from intergenic regions, even though ∼40% of the binding sites occur in these regions. **D:** Meta-gene profile plot highlighting the enrichment of Foxa2 binding sites around TSSs displayed as binding sites per bp per gene over the genomic regions of all RefSeq genes. **E:** Left: Normalized coverage of H3K4me3, H3K27ac and ATAC-seq around all 20377 Foxa2 binding sites, averaged over FVF^{low} and FVF^{high}. Dashed lines mark the region used for quantification of the active chromatin marks. Right: Average normalized read counts of H3K4me3, H3K27ac and ATAC-seq in FVF^{low} and FVF^{high} at a 1 kb region centered on the 20377 total Foxa2 binding sites in FVF^{low} and FVF^{high} and at a set of 20377 randomly chosen 1 kb sites. H3K4me3, H3K27ac and ATAC-seq are enriched ∼10× at Foxa2-bound sites compared to random sites. P-value for all comparisons < 2.2×10^{-16}. **F:** Most enriched motif detected by motif analysis is the known Foxa2 consensus sequence and is identified in ∼69% of the Foxa2-bound sequences in FVF^{low} and FVF^{high}.

Next, binding sites were attributed to potential target genes by assigning each Foxa2 peak that lies within a 20 kb radius around the TSS or within the body of a given gene, to the corresponding gene. Using this technique, the identified binding sites were mapped to 8666 and 11335 potential target genes in FVF^{low} and FVF^{high} cells, respectively. Foxa2 binding was found near many genes, important for the function and identity of the intestinal epithelium, such as *Cdx2*, *Hnf4a* or *Gata4* (Figure 2.28 A-C). Moreover, the ensemble of Foxa2 bound genes contains 70% of the differentially expressed genes between FVF^{high} and FVF^{low} and 63% of the ISC signature

genes [342], including the ISC marker Lgr5 (Figure 2.28 D-E). This indicates that Foxa2 is an important intestinal TF and also suggests role for Foxa2 in the regulation of ISC function.

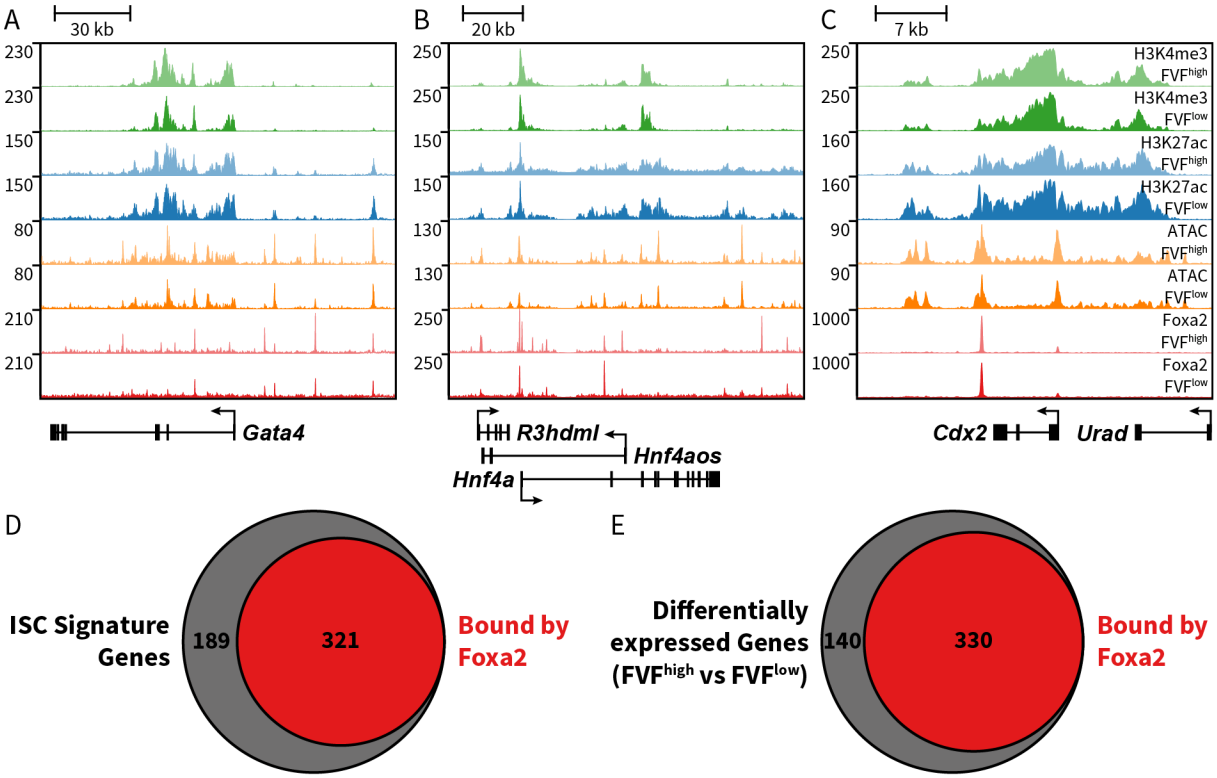

Figure 2.28: Foxa2 Binds Important Intestinal Genes. A-C: ChIP-seq data tracks showing H3K4me3, H3K27ac, ATAC-seq and Foxa2 from FVF^{low} and FVF^{high} cells at the *Gata4* (A), *Hnf4a* (B) and *Cdx2* (C) locus as examples for important intestinal genes, bound by Foxa2. **D:** Venn diagram showing the overlap between ISC signature genes [342] and Foxa2-bound genes (FVF^{low} and FVF^{high}). **E:** Venn diagram showing the overlap between differentially expressed genes (FVF^{high} vs FVF^{low}) and Foxa2-bound genes (FVF^{low} and FVF^{high}).

2.2.6 High Foxa2 Levels Correlate with Binding of EEC Specific Target Genes

Our data show that EEC and GC progenitors are associated with increased Foxa2 expression and loss of Foxa1/2 leads to defects in the differentiation of incretin secreting EECs and goblet cells [272]. We therefore wanted to know if Foxa2 binds EEC and GC differentiation genes and if binding is restricted to FVF^{high} cells. The comparison of Foxa2 binding profiles from FVF^{low} and FVF^{high} cells revealed, that 86% of sites bound in FVF^{low} cells were also bound in FVF^{high} cells, with only a small number of binding sites that were exclusively bound in FVF^{low} cells. On the other hand, 49% of Foxa2 binding sites in FVF^{high} cells, were only bound in this population, suggesting that there is a large panel of target sites that require higher levels of Foxa2 to be bound (Figure 2.29).

Next, we examined the functions of genes in the vicinity of Foxa2 binding sites using gene ontology and pathway analysis. In general, Foxa2-bound genes were involved in epigenetic gene

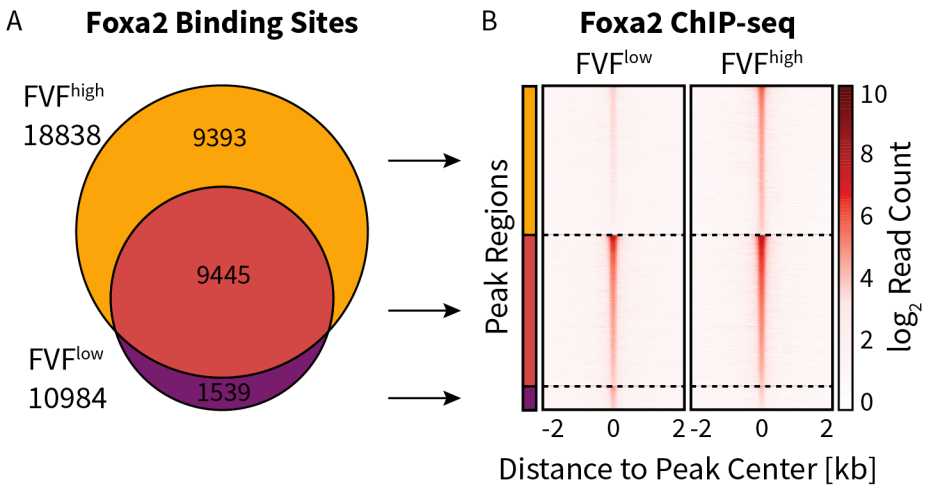

Figure 2.29: Differential Foxa2 Binding in FVF^{low} and FVF^{high} Cells. A: Venn diagram showing the overlap between Foxa2 binding sites in FVF^{low} and FVF^{high}. **B:** Heatmap of Foxa2 ChIP-seq signal in a 4 kb region around the peak center. Peak regions were sorted by signal strength and grouped by their presence or absence in FVF^{low} and FVF^{high}.

regulation, Wnt signaling and epithelial cell differentiation. Genes, bound by Foxa2 in both, FVF^{low} and FVF^{high}, populations, were linked to metabolic pathways and tissue homeostasis, while genes, specifically bound in FVF^{high} cells, were associated with chromatin organization, endocrine differentiation and the downregulation of EGFR (Figure 2.30 C). Inhibition of EGF signaling has been shown to regulate ISC quiescence and endocrine lineage allocation *in vitro* [240]. Among the genes exclusively bound in FVF^{high} cells, we found master regulators of enteroendocrine (*Neurog3*, *Neurod1*, *Rfx6* and *Insm1*) and GC (*Spdef*) differentiation (Figure 2.30 A-B). This suggests that in order to commit to the endocrine lineage, cells upregulate Foxa2 thereby facilitating the activation of endocrine fate determining genes.

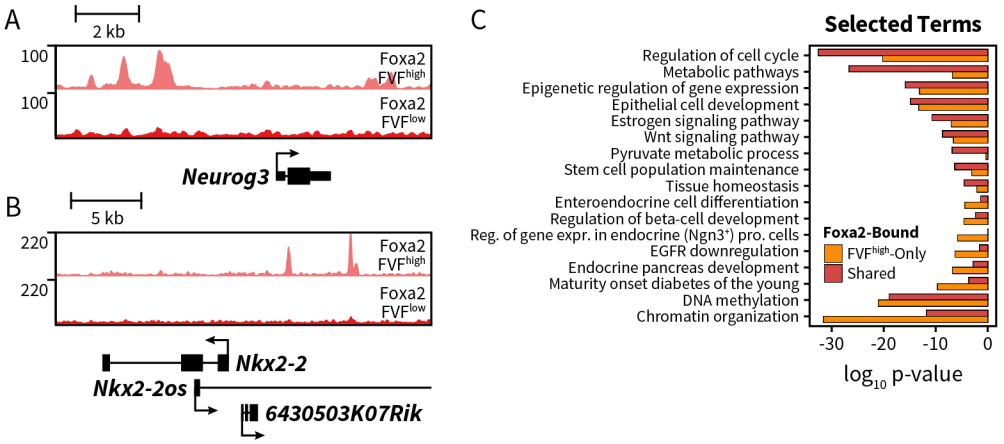

Figure 2.30: Functional Implications of Differential Foxa2 Binding. A-B: ChIP-seq data track showing Foxa2 from FVF^{low} and FVF^{high} at the *Neurog3* (A) and *Nkx2-2* (B) loci. **C:** Selected GO terms and pathways enriched in genes associated exclusively with FVF^{high} specific and shared Foxa2 binding sites.

2.2.7 Foxa2 Binding Promotes Chromatin Opening

Foxa2 is a pioneer transcription factor and is thus able to induce remodeling and opening of chromatin in the vicinity of its binding sites [290, 343]. The increased accessibility in turn enables binding of co-factors that are required to regulate gene expression [344]. Consistent with its pioneering activity, the vast majority of Foxa2 binding sites (83%) show chromatin accessibility and only a minor fraction of the binding sites were found in closed chromatin. Vice versa, a fraction of 10% and 15% of all identified accessible regions, were bound by Foxa2 in FVF^{low} and FVF^{high}, respectively, suggesting that Foxa2 is an important transcription factor in these cells. Strikingly, 70% of sites, showing specific opening in FVF^{high} cells were bound by Foxa2, while only 28% of sites with specific opening in FVF^{low} cells were bound (Figure 2.31).

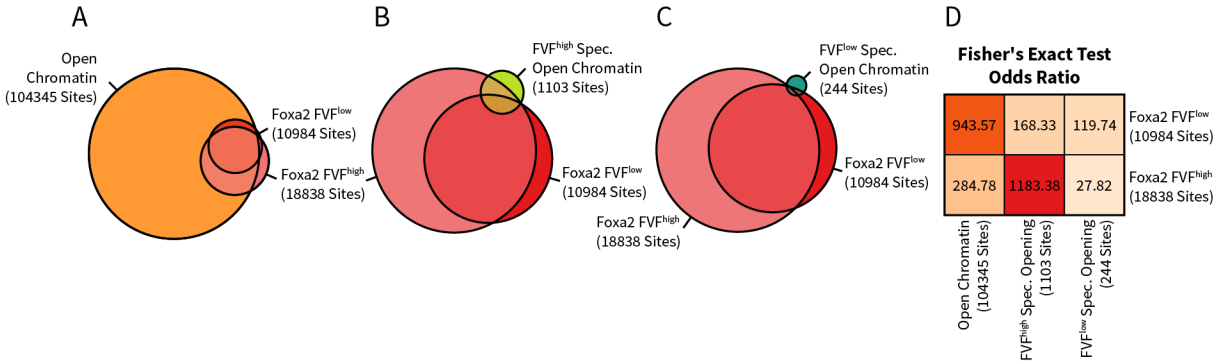

Figure 2.31: Foxa2 is Enriched at Open Chromatin Sites. A-C: Venn diagrams showing the overlap of all open chromatin regions (A), FVF^{high}-specific open chromatin (B) and FVF^{low}-specific open chromatin regions (C) and the Foxa2 binding sites from FVF^{low} and FVF^{high} cells. **D:** Heatmap showing odds rations calculated by the Fisher's exact test for each comparison. All p-values were $< 2.2 \times 10^{-16}$.

We were interested in a possible connection between Foxa2 binding and chromatin opening and compared chromatin accessibility at Foxa2 binding sites in FVF^{low} and FVF^{high} cells. At shared binding sites, bound in both populations, the mean accessibility, as measured by the normalized tag count per kb from ATAC-seq data at a given set of sites, was similar but slightly higher in FVF^{high} cells. At FVF^{high}-specific Foxa2 binding sites, the differences in chromatin opening were stronger, showing markedly higher ATAC-seq signals in FVF^{high} cells. The opposite trend was observed at FVF^{low}-specific binding sites that showed a slightly increased accessibility in FVF^{low} cells (Figure 2.32 A). This result is in accordance with the findings from the chromVAR motif analysis that showed a strong association between openness of chromatin in FVF^{high} cells and the FKHD consensus motif. The motif analysis also identified NR and some homeodomain motifs to be associated with specific chromatin opening in FVF^{low}. To test if these motif-based analyses can be recapitulated with actual binding data, we analyzed publicly available ChIP-seq data sets from intestinal Atoh1 [222], Hnf4a and Cdx2 [345] and measured the ATAC-seq signal at the identified binding sites. We applied the chromVAR method to calculate ATAC-seq deviations associated with ChIP-seq peak regions, which showed a strong association of shared and FVF^{high}-specific Foxa2 and Atoh1 binding sites with increased openness in FVF^{high} cells. Interestingly, this analysis also indicated that, especially at FVF^{high}-specific Foxa2 binding sites, chromatin in FVF^{low} cells is closed. On the other hand, at FVF^{low}-specific Foxa2, Cdx2 and Hnf4a

peaks, chromatin was more accessible in FVF^{low} cells but the differences were less pronounced and these sites were not completely closed in FVF^{high} cells (Figure 2.32 B). We next counted the ATAC-seq fragments at Atoh1, Cdx2 and Hnf4 sites. At Atoh1 sites, counts were increased in FVF^{high} cells and remarkably, at sites shared by Atoh1 and Foxa2, the increase was the strongest (Figure 2.32 C). At the Cdx2 and Hnf4a sites, we did not see marked differences between FVF^{low} and FVF^{high} cells. This discrepancy is very likely due to the different normalization strategies used for fragment counts and by chromVAR. Nevertheless, both methods clearly show that Foxa2 binding is linked to chromatin opening.

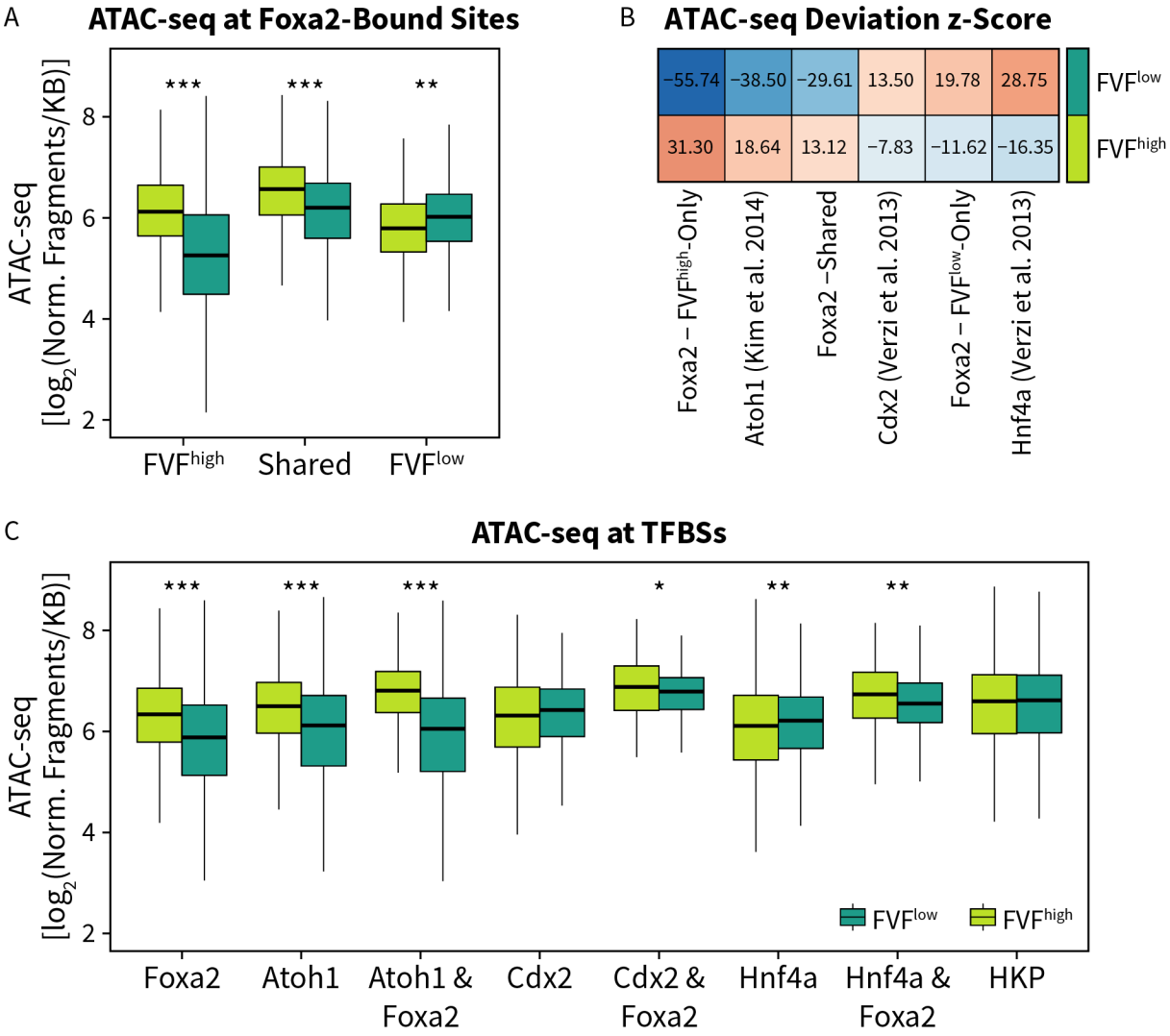

Figure 2.32: Chromatin Accessibility is Increased at Foxa2 Binding Sites. A: Normalized ATAC-seq fragments per kb from FVF^{low} and FVF^{high} cells at Foxa2-bound sites that are FVF^{high}-specific, shared and FVF^{low}-specific, respectively. **B:** Heatmap of ATAC-seq deviation z-score calculated for Foxa2, Atoh1, Cdx2 and Hnf4a binding sites. **C:** Normalized ATAC-seq fragments per kb at Foxa2, Atoh1, Cdx2 and Hnf4a binding sites. Sites co-occupied by Foxa2 and Atoh1, Cdx2 or Hnf4a are plotted separately. Wilcoxon rank sum test with continuity correction, *: $p<5\times10^{-2}$, **: $p<10^{-7}$, ***: $p<10^{-100}$.

To see whether changes in Foxa2 occupancy are associated with changes in chromatin accessibility, we compared the Foxa2 ChIP-seq signal at sites with differential accessibility in FVF^{low} and FVF^{high} cells. We found that at Foxa2-bound sites with higher accessibility in FVF^{high} cells the enrichment of Foxa2 was also higher, indicating a relationship between Foxa2 binding strength and chromatin opening. This was also observed in an opposite manner at sites more accessible in FVF^{low} cells (Figure 2.33 A-B). Moreover, a direct comparison of the fold changes in Foxa2 ChIP-seq and ATAC-seq signal revealed that changes in Foxa2 binding were positively correlated (ρ=0.63) with changes in chromatin accessibility (Figure 2.33 C). Interestingly, we also saw that Foxa2 binding to closed chromatin was associated with weak enrichment of Foxa2 (Figure 2.33 B).

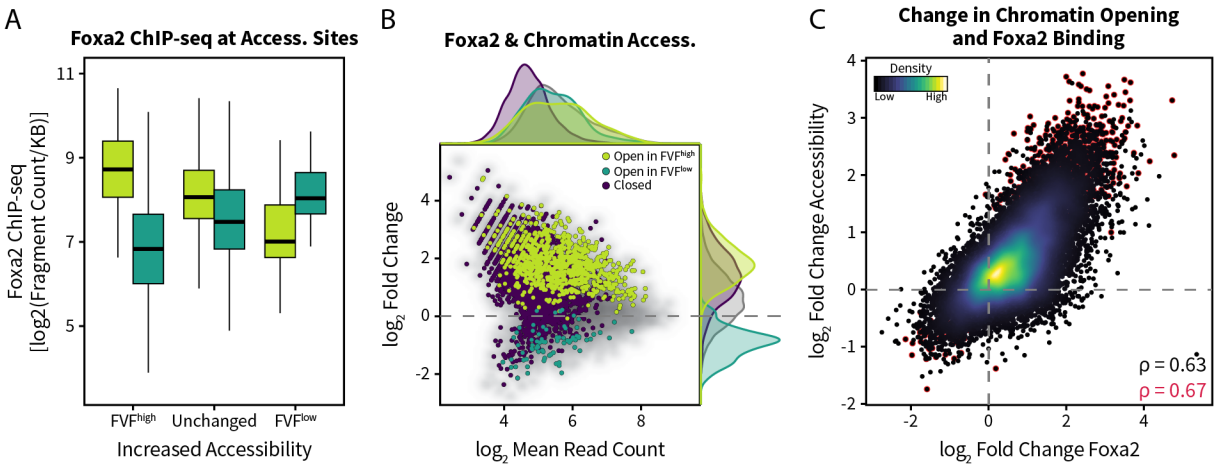

Figure 2.33: Foxa2 Binding Strength Correlates with Chromatin Opening. A: Foxa2 ChIP-seq fragments per kb at open chromatin sites that were specifically open in FVF^{high} cells, open in both populations (unchanged) and specifically open in FVF^{low} cells, respectively. Wilcoxon rank sum test with continuity correction, p-value < 10^{-9} for all comparisons. **B:** MA plot showing the $\log_2$ fold change over the $\log_2$ mean read count of the Foxa2 ChIP-seq in FVF^{low} and FVF^{high} cells. Marked in color are binding sites that overlap different categories of open chromatin. Densities of all (gray) and the marked sites (colored) are shown on the margins of the plot. **C:** Scatter plot of the $\log_2$ fold changes of Foxa2 against the $\log_2$ fold changes of chromatin accessibility (non-shrunken). Points corresponding to significantly different ATAC-seq peaks are marked in red. Spearman's rank correlation coefficient is shown for all sites and significant sites only (red).

Together, these results clearly show that Foxa2 binding is associated with changes in chromatin accessibility and the correlation between changing Foxa2 enrichment and chromatin opening suggests a causal link. Thus, our results show that upon higher expression levels of Foxa2, new sites are bound, which then gain accessibility and participate in the regulation of enteroendocrine differentiation. In additions, these results suggest that Foxa2 might cooperate with other TFs such as Atoh1.

2.2.8 Cooperative Binding Defines Foxa2 Function

Enhancers are often co-occupied by multiple TFs that orchestrate transcriptional regulation in concert [346]. Foxa2 has been shown to interact with several different TFs in various tissues to carry out specific tasks. In the pancreas, Foxa2 cooperates with TFs such as Pdx1 and Neurog3

or Insm1 and Neurod1 to control pancreas and endocrine development or β-cell function, respectively [87, 347, 348]. In the liver, Foxa2 works together with TFs like Hnfa4 or the other Foxa family members Foxa1 and Foxa3 [348–351]. Since Foxa2 is able to cooperate with a panel of different TFs in different tissues, we were interested in finding possible partner TFs that cooperate in a cell type-specific manner. Thus, in order to identify TFs, potentially cooperating with Foxa2, the chromVAR method was applied to the Foxa2 ChIP-seq data. Among the known motifs of the set of expressed transcription factors, we found several bHLH, FKHD, RFX and homeodomain motifs to be associated with increased Foxa2 binding in FVF^{high} cells and nuclear receptor and bZIP motifs associated with Foxa2 binding in FVF^{low} cells (Figure 2.34). These results are consistent with what we found in the ATAC-seq data, which might be due to the large fraction of Foxa2 bound sites among the sites with differential accessibility.

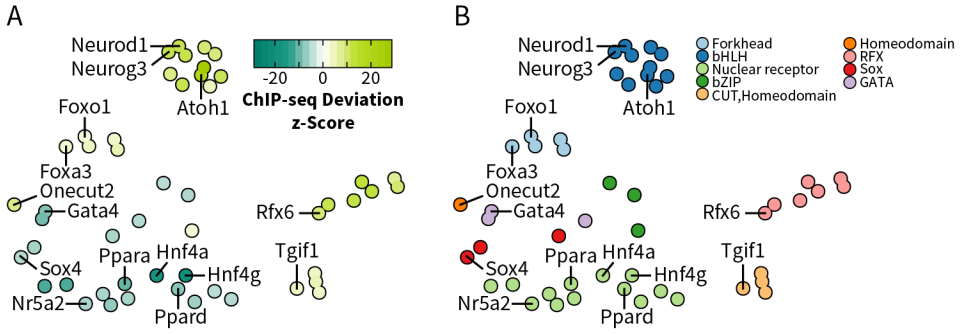

Figure 2.34: Motif Analysis of Foxa2 Binding Sites. UMAP representation of known TF motifs and their associations to changes in Foxa2 binding. Motifs are distributed according to their similarity to each other. **A:** Motifs are colored according to their ChIP-seq deviation score, representing increased enrichment in FVF^{low} (dark green) and FVF^{high} (light green) cells. **B:** Motifs are colored according to the TF family affiliation.

Since the motif analysis suggests that there is a connection between the Foxa2 binding in FVF^{low} and FVF^{high} cells and possible co-binding TFs, we again used the ChIP-seq data sets from intestinal Atoh1 [222], Hnf4a and Cdx2 [345] and identified regions that were co-occupied by Foxa2. In total, 1486, 916 and 256 Sites were found to be co-occupied by Foxa2 and Atoh1, Hnf4 or Cdx2, respectively. The density distributions of the sites showed that Foxa2 binding at sites, co-bound by Atoh1 had stronger Foxa2 binding in FVF^{high} cells, which reflects the overall density distributions of Foxa2 which has generally a higher signal in FVF^{high} cells. In contrast, sites that were also bound by Hnf4a or Cdx2 appeared to be not differentially enriched with Foxa2 in FVF^{low} and FVF^{high} cells. Sites that were bound by other TFs tended to show comparably high Foxa2 binding (Figure 2.35 A). We then calculated the relative distance between Foxa2 binding sites and each of the other three TFs, to see if there were spatial correlations in the binding patterns. For all three TFs, the observed relative distances showed that their binding sites were closer to Foxa2 sites than it would be expected by chance, suggesting that the observed overlap between Foxa2 and Atoh1, HNf4a and Cdx2 were not random observations, but rather functionally relevant (Figure 2.35 B). Also, the Fisher exact test yielded high odds rations (p-values $< 2.2 \times 10^{-16}$), indicating a non-random relationship between the binding of Foxa2 and the tested TFs (Figure 2.35 D).

To know which genes were affected by co-binding of Foxa2 and Atoh1, Hnf4a or Cdx2, binding sites were mapped to genes as described before (2.2.5) and associated genes were subjected

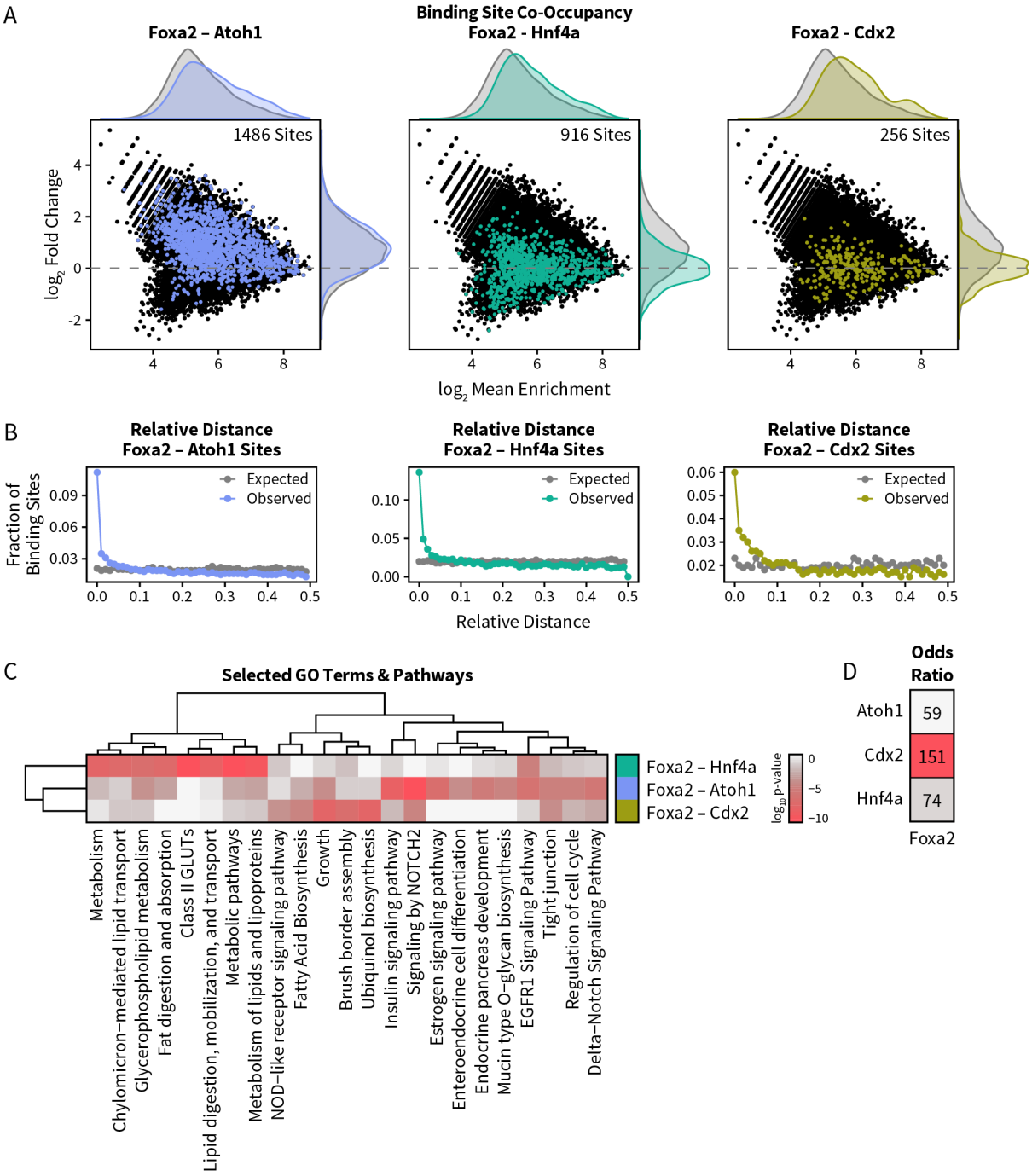

Figure 2.35: Foxa2 Cooperates with Atoh1, Cdx2 and Hnf4a. A: MA plots showing the $\log_2$ fold change over the $\log_2$ mean read count of the Foxa2 ChIP-seq in FVF^{low} and FVF^{high} cells. Marked in color are binding sites that overlap with binding sites from Atoh1 (left), Cdx2 (middle) and Hnf4a (right). Densities of all (gray) and the marked sites (colored) are shown on the margins of the plots. **B:** Relative distances between Foxa2 and Atoh1 (left), Cdx2 (middle) and Hnf4 (right), respectively. Observed distances are shown in color, expected distances in gray. **C:** Heatmap of notable GO terms and pathways enriched in the genes near cooperative binding sites. **D:** Odds rations calculated by the Fisher's exact test for each comparison. All p-values were $< 2.2 \times 10^{-16}$.

to pathway enrichment analysis. As already indicated by the different Foxa2 ChIP-seq signal at sites co-occupied with Atoh1 compared to those co-occupied with Hnf4a or Cdx2, the function of nearby genes was different between each of the groups. Genes near Foxa2–Atoh1 sites were associated with Delta–Notch, EGFR1 and estrogen signaling, cell cycle regulation, mucin biosythesis and (entero)endocrine differentiation. Foxa2–Hnf4a sites were associated with genes linked to metabolism as well as nutrient absorption and transport and Foxa2–Cdx2 sites with growth, brush border assembly and the sensing of bacteria by the innate immune system (Figure 2.35 C). Thus, depending on the co-occupying TF, Foxa2 regulates various functions in the small intestine that reach from PC function and enterocyte differentiation in conjunction with Cdx2 to the regulation of metabolic processes and nutrient uptake together with Hnf4a. However, most interesting in the context of the role of Foxa2 in secretory and especially endocrine cell development, is the finding that Foxa2 seems to cooperate with Atoh1 to regulate certain aspects of the endocrine lineage and GC differentiation.

2.2.9 scRNA-seq Shows Upregulation of Foxa2 During EEC & GC Differentiation

The advent of scRNA-seq technologies has enabled the parallel profiling of thousands of cells and has rapidly improved our understanding of the cellular coherence in the intestinal epithelium [162, 179, 192, 240, 276, 277, 279, 352–354]. Assessing the transcriptome on the single-cell level allows the precise identification of all cell types, their progenitors and even cell subtypes such as the different types of EECs present in the intestine [162, 179]. Hence, this is the state-of-the-art approach to uncover transcriptional networks governing cellular differentiation. We used scRNA-seq data generated for a different project in our lab [354] to study the correlation between Foxa2 expression and EEC and GC differentiation. This data set was particularly suited for that matter, since it was generated using an enrichment strategy that significantly increased the number of secretory cells, including EECs. More specifically, the data set was generated from FACS purified small intestinal cells of FVF mice, where 50% of the cells were sorted without considering the FVF marker (i.e. FVF^- & FVF^+ cells) and the other 50% were sorted to be FVF^+ (Figure 2.36 A) [354]. The data set comprises a total of 13366 cells and on average, 3566 genes were detected per cell. Unsupervised graph-based clustering was used to identify different cell populations that were then annotated using established marker genes. With this approach all mature intestinal cell types and most of their progenitors could be identified. Due to the absence of a specific known marker gene signature, one cluster could not be matched to any intestinal cell type (Figure 2.36 B-C). Next, diffusion pseudotime [355] was calculated to reconstruct developmental progression and sort the cells according to their pseudo-temporal order of differentiation. To discover new marker genes that effectively distinguish the individual clusters, we performed a pair-wise differential expression analysis of each cluster against every other cluster and defined genes, positively associated with a given cluster in every comparison as markers for this given population. Using these marker genes, we performed GSEA in the microarray data obtained from FACS purified FVF populations to revisit our attribution of different intestinal cell types to FVF populations. In general agreement with the previous classification, GSEA showed an association of EECs with FVF^{high} cells, ISCs and PCs with FVF^{low} cells and enterocytes and TCs with FVF^{neg} cells. Notably, enterocyte progenitors were found to

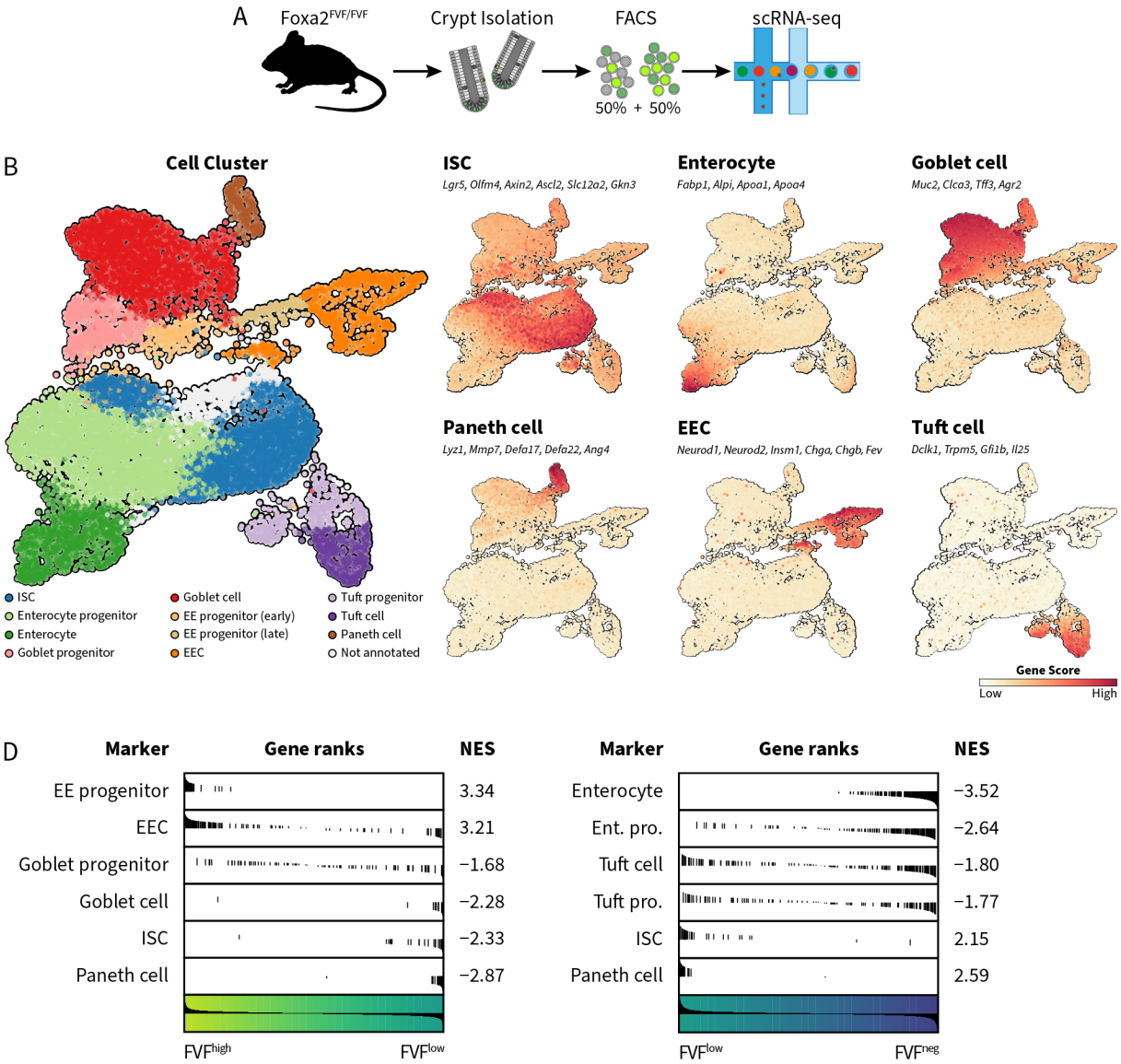

Figure 2.36: Single-Cell Transcriptomics of Intestinal Crypt Cells. A: Schematic representation of the experimental layout scRNA-seq of crypts from FVF mice with enrichment of FVF^{+} secretory cells. **B:** UMAP plot of the 13366 single cells pooled from three replicates. Different clusters, identified based on the expression of known marker genes are highlighted in different colors. *C:* Expression scores of marker gene sets, specific for each intestinal cell lineage. *D:* GSEA of identified marker genes from each cluster of the dataset. Markers for early and late EE progenitors were pooled. Data analysis performed in cooperation with Sophie Tritschler.

be in-between the FVF^{low} and FVF^{neg} population. Moreover, with the new marker genes, we revealed that GC progenitors were found in the FVF^{high} population, while mature GCs were rather found among the FVF^{low} cells (Figure 2.36 D).

Foxa2 was expressed at low levels, which results in low read counts originating from the gene and hence a lower signal-to-noise ratio and increased rates of dropout events [356]. To overcome this, we used a computational method, using a deep count autoencoder network (DCA) to

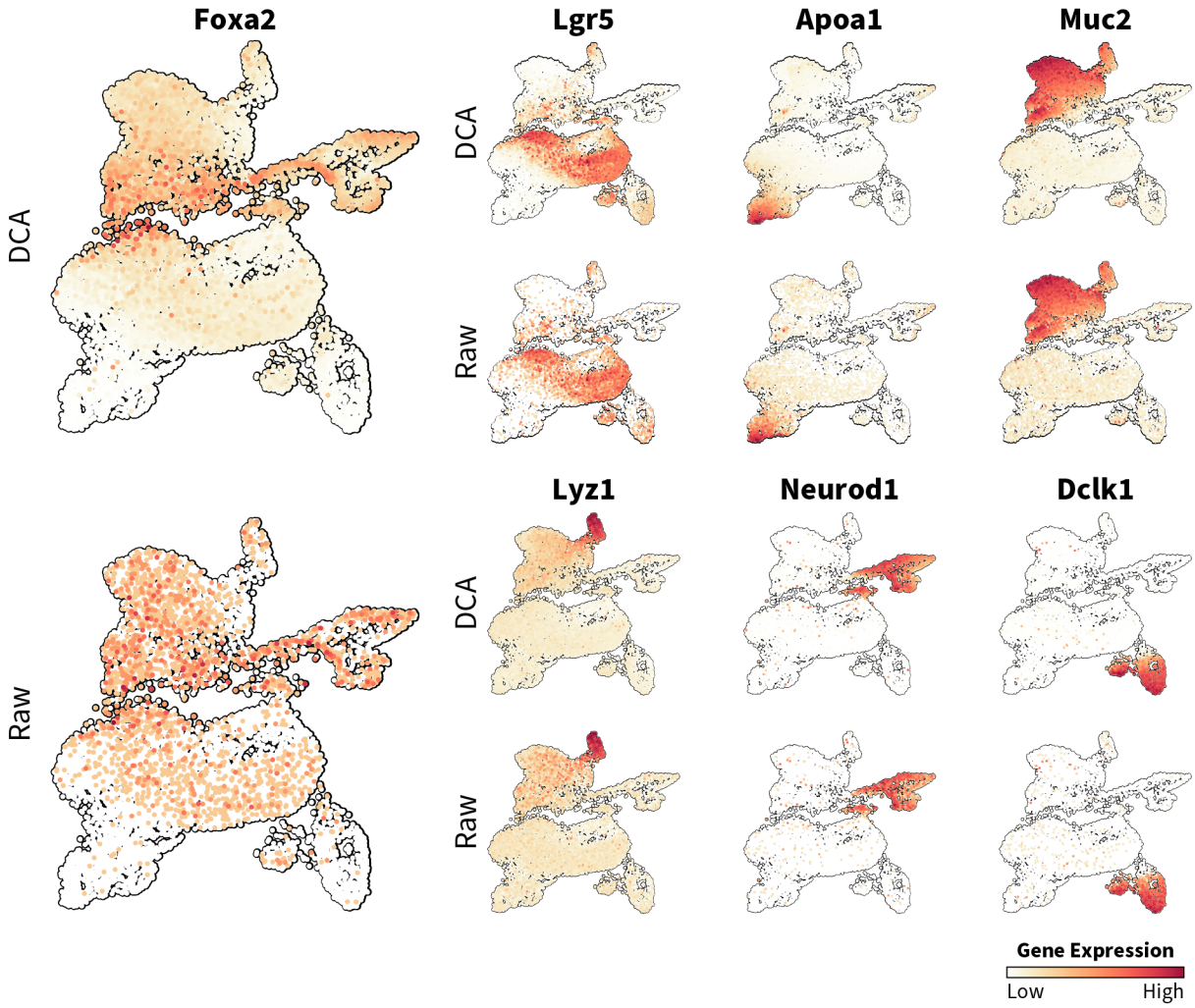

Figure 2.37: Denoising of scRNA-seq Data. Expression patterns of selected intestinal marker genes and *Foxa2*, comparing denoised (DCA) and raw expression values.

denoise the scRNA-seq data [356]. The DCA approach results in a marked improvement in the representation of lowly expressed genes, including *Foxa2*, while preserving overall expression patterns (Figure 2.37).

With the knowledge of cluster identities and the pseudo-temporal order of our data set, we were able to asses *Foxa2* expression in each cell type and along differentiation trajectories. In agreement with previous results we saw highest *Foxa2* levels in GC progenitors, EECs and EEC progenitors. Mature GCs, PCs, ISCs, enterocyte progenitors and TC progenitors had decreasingly lower *Foxa2* expression. Enterocytes and TCs did not express *Foxa2* (Figure 2.38 A). We next examined the differentiation trajectories of EECs and GCs in respect to the expression of *Foxa2*, Foxa2-bound genes and important lineage determining factors. As suggested by the expression patterns in the different cell clusters, the pseudo-temporal ordering showed a strong increase of *Foxa2* expression in early EE progenitors that peaks at the transition of from early to late EE

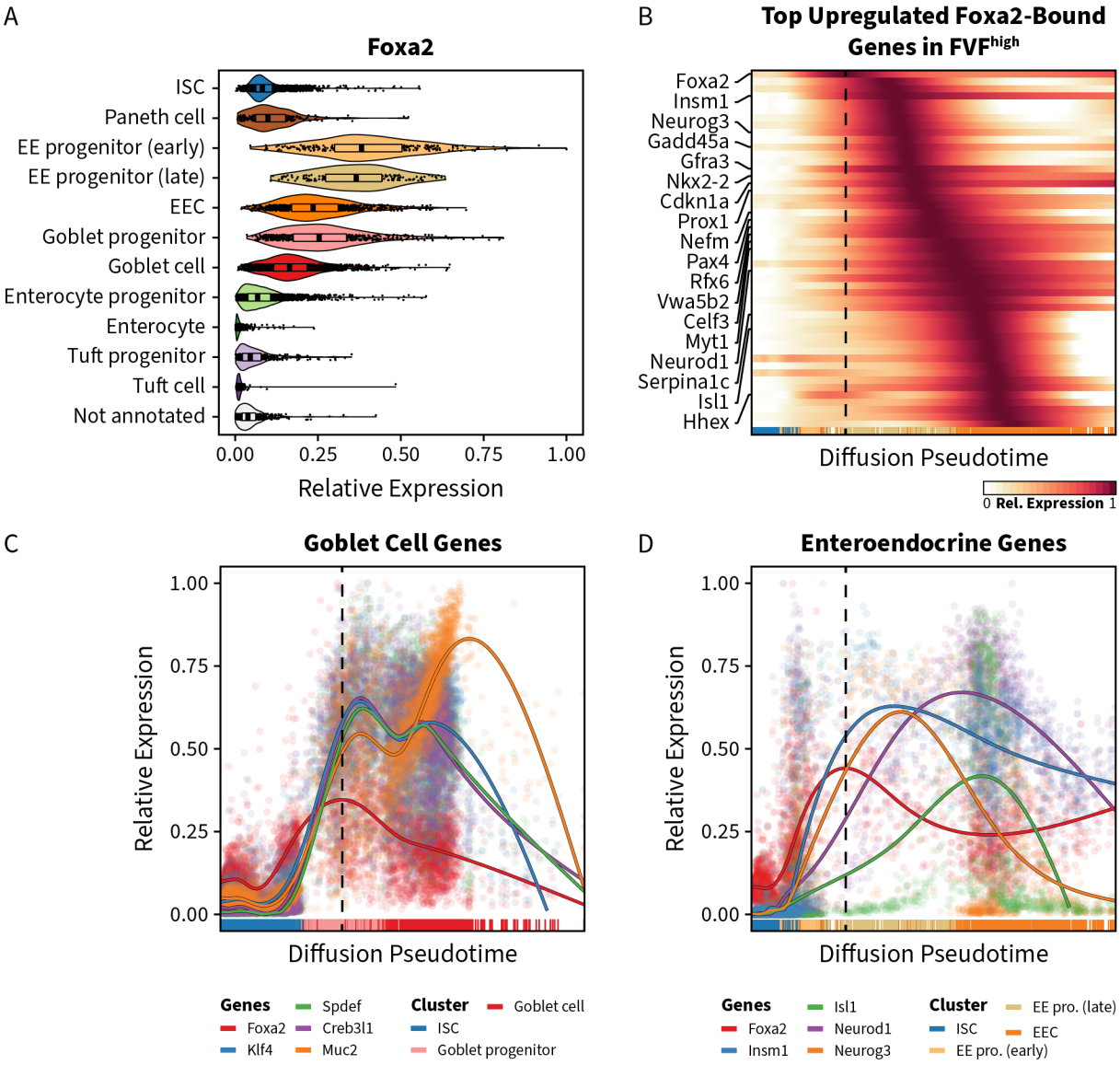

Figure 2.38: Expression Patterns During Lineage Differentiation. A: Expression levels of *Foxa2* in the different intestinal lineages, as defined by clustering of the single-cell data. **B:** Heatmap showing the expression of the most upregulated Foxa2-bound genes in FVF^{high} cells along the EEC differentiation path. Colors of the bottom row represent cluster affiliations. **C–D:** Profiles of the relative expression of Foxa2 and selected genes of the GC and EEC lineages, respectively. Colored lines represent the Loess-smoothed mean expression. Dashed lines indicate maximum *Foxa2* expression. Colors at the bottom represent cluster affiliations of the individual cells. Data analysis performed in cooperation with Sophie Tritschler.

progenitors. The top upregulated genes in FVF^{high} cells with Foxa2 binding sites, as identified by microarray analysis, include some of the most important endocrine TFs, such as *Insm1*, *Neurog3*, *Neurod1* or *Isl1*, and all of them had highest expression levels in late EE progenitors or EECs and generally after upregulation of *Foxa2* (Figure 2.38 B & D). Similarly, on the path of GC differentiation, Foxa2 expression was rising in GC progenitors and expression of Foxa2-bound GC differentiation genes, such as *Spdef*, *Klf4* or *Creb3l1* succeeded Foxa2 upregulation (Figure 2.38 C). Moreover, we identified genes, that share a similar expression pattern with

Foxa2, by gene expression correlation analysis and found 269 genes ($\rho > 0.5$) co-regulated genes. These contained many known EEC and GC lineage differentiation genes, as well as marker genes identified in our data and 68% of the co-regulated genes were bound by Foxa2, suggesting that these genes are regulated by Foxa2.

Thus, our scRNA-seq data shows that Foxa2 is upregulated during EEC and GC differentiation and places it upstream of important lineage determining genes, such as *Neurog3*. This is in line with our conclusions from microarray and genomics data and supports the initial hypothesis, that high levels of Foxa2 drive the differentiation towards the EEC and GC lineage.

3 Discussion

3.1 Modeling Pancreas Development

The overall goal of stem cell research is to gain insight into pathomechanisms and to provide novel strategies for disease treatment. Since it has been shown that islet transplantation can successfully revert insulin dependence in diabetic patients, cell-replacement therapies are a viable approach in the treatment of diabetes [150]. The shortage of islet donors and the necessity of life-long suppression of the immune system to prevent graft rejection, render hESCs and in particular patient-derived iPSCs a promising source to generate large quantities of transplantable β-cells. However, the efficient generation of fully functional and mature β-cells still remains a challenge, even though the key events of pancreatic endocrine cell formation *in vivo* are recapitulated during *in vitro* differentiation [155–157, 357]. To improve the current differentiation protocols it is therefore essential to understand the molecular mechanisms, which govern the development of pancreatic endocrine cells. In this context, it is important to characterize the specification of early human pancreatic lineages that give rise to functional β-cells. The tempo-spatial patterns of TF activity are critical for the differentiation process, and the knowledge of the expression patterns of key pancreatic TFs and their target genes is important to improve the recapitulation of the differentiation events, which are required for the successful formation of functional β-cells *in vitro*. However, the use of *in vitro* models is indispensable to study early pancreas development, since the access to primary human material is very limited.

3.1.1 Consistency of PDX1 Binding Sites Across Model Systems

One of the key TFs governing pancreas and β-cell development is PDX1. While its important role during the different stages of pancreas development has been extensively studied in mice, the stage specific functions of PDX1 in early pancreatic progenitors and adult β-cells is not well understood. We combined mRNA profiling with an integrated analysis of genome-wide PDX1 binding sites and H3K27ac enrichment in pancreatic progenitors (PPs) derived from a novel human iPSC line, in order to identify downstream target genes and to characterize the role of PDX1 in early pancreagenesis.

To our knowledge, this is the first profile of genome-wide PDX1 binding sites in iPSC derived PPs. Our analysis identified a multitude of target sites and associated genes, including developmental regulatory factors, signaling molecules and factors important for mature β-cell function. Importantly, many of the genes, differentially expressed between PPs and iPSCs, were bound by PDX1.

We compared our results with previously published PDX1 datasets from hESC-derived PPs of similar stages [302, 311] and found significant differences in PDX1 binding. The reasons that could explain the observed discrepancies are diverse and include differences in the efficiency of differentiation, the factors used in the *in vitro* differentiation protocol and the exact developmental stage at the time of the experiments. The latter could explain many of the differences between our data and that from Wang et al. [302], as their cells already expressed NKX6.1 and were thus slightly further in development. Moreover, cell lines differ in their competence to differentiate into specific lineages [358]. This likely influences the differentiation process in a given cell line and might favor a set of target sites over another.

Despite the differences between experiments from various labs using diverse cell lines and protocols, it is meaningful to compare the datasets. We used the overlap of PDX1 binding sites from these different experiments in order to identify a set of high confidence PDX1 binding sites that is more likely to represent the true PDX1 binding patterns in early PPs than those identified by a single experiment.

3.1.2 Identification of Novel PDX1 Target Genes in Humans

The specific function of a TF is defined by the target genes it regulates. However, most studies addressing the target genes of Pdx1 were conducted in mice [122, 359, 360] and it is not clear if PDX1 regulates the same genes in humans. Our analysis identified 5664 putative PDX1 target genes in human PPs, including *RFX3* and *DLL1*, which are important for pancreas development in mice [67, 308, 309]. *RFX3* is a member of the evolutionary conserved regulatory factor X gene family and encodes a TF, which contains a highly conserved winged helix DNA binding domain and a C-terminal dimerization domain [361]. In mice, Rfx3 is involved in cilia formation and is important for the differentiation of endocrine cells in the pancreas [308, 362, 363]. The deletion of the gene causes ciliary abnormalities, compromising the differentiation of pancreatic endocrine cells and resulting in reduced numbers of defective β-cells [308, 363]. Furthermore, RFX3 has been shown to bind to two well-known regulatory sequences of the glucokinase gene (*GCK*) in Min6 cells and human islets [308]. DLL1 is an important canonical Notch ligand and is a member of the delta/serrate/jagged family. DLL1 activates Notch signaling in neighboring cells, which in turn represses the tip and later the endocrine fate during pancreas development (see 1.4). Mice deficient of *Dll1* show increased commitment to the endocrine lineage and a concomitant depletion of the progenitor pool due to increased expression of Ngn3, which ultimately leads to pancreatic hypoplasia [67, 309]. Whether *RFX3* and *DLL1* play similarly important roles in human pancreas development has not been addressed so far, but our results suggest an evolutionary conserved function of these PDX1 target genes. In contrast, other genes bound by PDX1 in human PPs, such as *TIMP2*, *MAFB* and *SIX2*, are known to be specifically expressed in human but not in mouse pancreata [364, 365]. To study the function of novel PDX1 target genes and other aspects of pancreas development, the XM001 iPSC cell line has been modified to express CRISPR/Cas9 under the control of an inducible promoter, which will allow gene functional studies in the future [366].

3.1.3 Diverging Functions of PDX1 in Development & Adulthood

The expression of PDX1 follows a specific spatio-temporal pattern. During embryogenesis, it is expressed throughout the developing pancreas and duodenum, but becomes gradually restricted to β- and δ-cells in the pancreatic islets, as well as some cells in the duodenum and stomach [41, 58, 125, 126, 129]. Therefore, it is likely that the function of PDX1 and the set of bound genes differs in PPs and mature β- and δ-cells. The comparison of PDX1 binding profiles from stem cell-derived PPs and primary adult islets revealed that only a minor fraction of ∼15% of the binding sites are shared between development and adulthood. This strongly indicates diverging functions of PDX1, depending on the developmental stage. Based on the GO-term analysis of PDX1-bound genes in PPs and adult islets, we found that the function of genes near common and specific binding sites differed markedly and reflected cellular function. In islets, PDX1-bound genes were strongly related to β-cell specific functions, such as the regulation of insulin secretion. Genes with adult specific PDX1 binding sites include the K^+ channel *KCNJ11* and the glucose transporter *SLC2A2* (Glut2), essential for adult β-cell function, as well as *NKX6-1* and *MAFA*, which are important for β-cell identity and maturation. On the other hand, in PPs, PDX1 is bound to genes like *GATA4*, *ONECUT1* or *HNF1A*, which are involved in epithelial cell differentiation and the regulation of a PP-specific gene program. This shows that PDX1 binding is highly dynamic and depends on the cellular context and the developmental stage. Thus, the same TF can regulate multiple molecular programs, required for stage-specific functions.

3.1.4 Early Pancreas Development and Diabetes Risk

SNPs are a common form of genetic variation, and in most cases they have no negative impact on development and health. Some SNPs, however, can increase the risk of developing a particular disease or may even be causal for a given disease. In the last decade, many disease associated SNPs and their corresponding genes have been identified, of which more than 300 genes are linked to T2DM [314]. Severe mutations in key genes, regulating β-cell development and function result in monogenetic forms of diabetes or even pancreatic agenesis [13–16, 134, 135]. However, most SNPs associated with T2DM occur in non-coding regions and increase the susceptibility to T2DM. A recent study showed that active regulatory regions in adult islets are enriched in diabetes-associated SNPs, suggesting that the increased risk of T2DM is caused by disturbed regulatory processes [313]. Interestingly, we also found that the regulatory regions, active during early pancreas development (PP stage), are enriched in T2DM-associated SNPs.

One of the SNPs identified in our analysis, rs11263763, lies in a PP-specific PDX1 binding site within the first intron of the *HNF1B* gene, and has been shown to reduce *HNF1B* expression [367]. In addition to rs11263763, *HFN1B* harbors several SNPs in the first and second intron, which are associated with T2DM. During murine pancreas development, Hnf1β is involved in the regulation of multipotent pancreatic progenitor proliferation and survival, and the loss of *Hnf1b* results in pancreatic hypoplasia. Hnf1β also regulates the expression of *Neurog3* in endocrine progenitors and is thus important for specification of the endocrine lineage [368]. Moreover, certain mutations of the human *HNF1B* gene cause monogenic diabetes (MODY5) [369]. We also identified two SNPs, rs12762233 and rs7087006, which are located in PP-specific PDX1 binding

sites near the *TCF7L2* gene. TCF7L2, also known as TCF4, is a member of the TCF/LEF family of TFs, which are characterized by a high mobility group domain, facilitating DNA binding, and act as downstream effectors of the canonical Wnt pathway [370]. TFC7L2 has been shown to regulate β-cell function and survival in humans [371], and its murine ortholog is an important regulator of β-cell development [372]. Genome-wide association studies (GWAS) have identified many T2DM-linked SNPs near *TCF7L2,* including those with the highest disease association. Therefore, the *TCF7L2* locus conveys the highest risk for developing T2DM in Europeans [373]. For instance, insulin secretion defects in the presence of normal incretin plasma levels have been described in patients carrying TCF7L2 SNPs [374].

The presence of the active histone mark H3K27ac at regions carrying T2DM-associated SNPs indicates that these regions are involved in the regulation of gene activity in the PP stage. Thus, variation in these regions potentially compromises the regulation of T2DM-relevant genes at the PP stage, and might therefore impair developmental programs. Eventually, disturbances in pancreatic development could result in reduced β-cell mass or function at birth, or impaired postnatal expansion of β-cells, which would consequently contribute to an increased susceptibility to T2DM. This is in agreement with the finding that in children the β-cell mass, and presumably the rate of β-cell growth, is highly variable between individuals [375], and that the β-cell mass is reduced in diabetic patients [376]. HNF1β and TCF7L2 are important TFs regulating pancreas and endocrine development. Misregulation of these genes due to SNPs occurring in their regulatory regions might impair early developmental processes that increase the risk of developing T2DM later in adulthood.

3.2 Modeling the Effect of PDX1 Mutations on β-Cell Development

Interspecies diversity in the development and function of organs is a major issue in the translation of promising findings from model organisms to humans, and successful preclinical reports from animal models often fail to succeed in humans. Thus, in order to develop novel treatment strategies for complex diseases such as diabetes, comprehensive mechanistic understanding of human endocrine cell formation, function and failure is necessary. As described above, the access to primary human material and the impossibility of performing longitudinal studies in humans call for the establishment of well-performing models to investigate human pancreas development and homeostasis [151]. Hence, the use of ESCs, and in particular patient-derived iPSCs, to model pancreas and β-cell development *in vitro* provides a unique platform to study the etiology of human disease [155–157]. Of particular importance is the possibility to study developmental programs and the effect of perturbations on the predisposition to diabetes, by reproducing the phenotypes of specific mutations in genes, such as the master regulator of pancreas development and MODY gene *PDX1*. Using endocrine differentiation culture of donor-derived iPSCs, we investigated two common mutations, P33T and C18R, in the *PDX1* gene, and their effect on human β-cell development and function. Using iPSC lines generated from two subjects carrying a P33T and C18R mutation of *PDX1*, respectively, as well as newly generated isogenic cell lines, we show that the development of endocrine progenitors and β-cells is impaired by these mutations. However, our data revealed that in donor-derived $PDX1^{P33T/+}$ and

PDX1$^{C18R/+}$cells, the early stages of pancreas development are not affected. In contrast, an impaired induction of the endocrine lineage at the PP2 stage is observed in the homozygous *PDX1*$^{P33T/P33T}$ mutation of the isogenic control cell line. These results suggest that the P33T mutation exerts a dose-dependent effect on the endocrine lineage decision.

3.2.1 Impact of P33T and C18R Mutations on *PDX1* Expression

The expression of *Pdx1* is tightly controlled by four regulatory elements located upstream of the *Pdx1* gene. These regions are highly conserved and harbor binding sites for several important TFs, including Hnf1β, Foxa1/2, Nkx2.2, Pax6, MafA as well as Pdx1 itself, all of which regulate the cell type specific expression of Pdx1 [60, 113, 114]. Thus, Pdx1 regulates its own expression via an auto-regulatory positive feedback mechanism [114, 377, 378].

Our results show that *PDX1* mRNA and protein levels were reduced in the isogenic *PDX1*$^{P33T/P33T}$cell line at the PP1 stage. Since the P33T mutation lies in subdomain B of the PDX1 transactivation domain, we speculate that the mutation disturbs protein–protein interactions, thereby impairing the recruitment of transcriptional co-activators, which are required for the induction of target gene expression, including *PDX1* itself. Interestingly, the long non-coding RNA (lncRNA) *PLUTO* is upregulated in the *PDX1*$^{P33T/P33T}$mutant PPs. *PLUTO*, which is the acronym for *PDX1 Locus Upstream Transcript*, is a gene with multiple exons, spanning a ∼100 kb region, which contains a cluster of enhancers that is active in adult islets. The TSS is located approximately three kb upstream of the *PDX1* locus, and it is transcribed from the opposite strand [379, 380]. *PLUTO* has a positive regulatory effect on the neighboring *PDX1* gene by promoting interactions between the *PDX1* promoter and the upstream enhancer cluster, and the knock-down of the *PLUTO* lncRNA resulted in decreased *PDX1* expression [379]. Intriguingly, the expression of both, *PLUTO* and *PDX1* was found to be reduced in islets from diabetic and glucose intolerant donors [379]. Thus, the upregulation of *PLUTO* in the *PDX1*$^{P33T/P33T}$PP cells might be a compensatory response, aiming to increase *PDX1* expression. The reduction of *PDX1* levels also correlated with the impaired differentiation at the PP2 stage. During endocrine formation, the maintenance of high *PDX1* levels is essential for the specification and differentiation of endocrine progenitors [381]. It is therefore likely that the low levels of *PDX1*, as observed in *PDX1*$^{P33T/P33T}$mutants, perturb the endocrine fate specification.

Interestingly, neither *PDX1* levels at the PP1 stage, nor the differentiation at the PP2 stage, were impaired in *PDX1*$^{C18R/C18R}$mutant cells. These differences between the two mutations might be explained by the impact of the mutations on the structure of the PDX1 protein. The replacement of proline with threonine in the P33T mutation could be more severe, as proline is important for the generation of tight turns in the protein structure. The replacement with any other amino acid could therefore induce significant structural alterations that impair protein-protein interactions, involving the transactivation domain of PDX1. Hence, the P33T mutation might reduce PDX1 function by disturbing the interaction with co-binding TFs. This notion is also supported by the similar phenotype of *PDX1*$^{P33T/P33T}$and *PDX1*$^{+/-}$cells, which suggests a significant loss-of-function caused by the P33T point mutation.

3.2.2 Deregulation of Diabetes-Associated Genes in Mutant PPs

The differential gene expression analysis of the isogenic and donor-derived cell lines revealed many differentially expressed genes, including interesting diabetes-associated genes. Among the PPs from all isogenic mutants, 21 genes were commonly downregulated. These included *MEG3, LARGE1, ANPEP* and *CES1,* all of which are implicated in diabetes. *Maternally expressed gene 3* (*Meg3*) is, as the name suggests, a maternally expressed imprinted gene, which codes for a lncRNA. In mice, it is located in the distal region of chromosome 12, and its human ortholog is found on the syntenic region on chromosome 14q [382]. The downregulation of *Meg3* results in impaired insulin synthesis and secretion, as well as increased apoptosis in murine β-cells, suggesting an important role in maintaining β-cell identity [383]. Furthermore, it has been shown that *MEG3* is downregulated in islets of T2DM patients, supporting its essential role for β-cell function, and indicating relevance for T2DM pathogenesis [384]. The like-acetylglucosaminyltransferase 1 (LARGE1) is involved in protein glycosylation in the Golgi apparatus, and mutations of the *LARGE1* gene are generally associated with congenital muscular dystrophy. However, a recent study has identified a SNP at the *LARGE1* locus, rs16993330, as a T2DM risk variant in the Greenlandic population [385]. An increased risk for T2DM has also been associated with SNPs at the Aminopeptidase N (*ANPEP*) locus, which increase the expression of the *ANPEP* gene [386]. Moreover, the carboxylesterase 1 (Ces1) has been linked to obesity, hepatic steatosis and hyperlipidemia, as the deletion of *Ces1* in mice results in the development of obesity, fatty liver, hyperinsulinemia, insulin insensitivity and decreased energy expenditure [387]. Even though *LARGE1, ANPEP* and *CES1* have been associated with diabetes or insulin resistance, it is not clear whether these genes play a role during human β-cell development and/or function. To address these questions further investigation is necessary. Another interesting gene we found to be downregulated in $PDX1^{P33T/P33T}$and $PDX1^{+/-}$PPs, is *MNX1,* which encodes the TF Motor neuron and pancreas homeobox 1 (MNX1). This TF plays an important role in β-cell differentiation and proliferation, and is essential for postnatal β-cell fate maintenance [388]. Furthermore, homozygous mutations of the *MNX1* gene in humans cause permanent neonatal diabetes [318], but the mechanisms by which this TF affects human β-cell development are not yet understood.

The comparison of the differentially expressed genes in $PDX1^{P33T/P33T}$and $PDX1^{P33T/+}$mutant cells, identified five genes that were consistently misregulated in the same direction. Among these genes were *MEG3, NNAT* and *POSTN,* all of which have a function in β-cells. As described above, *MEG3* regulates insulin secretion and synthesis and is downregulated in T2DM. *NNAT* is a paternally expressed imprinted gene, encoding the proteolipid NEURONATIN. It is expressed in brain, endocrine cells and the adipose tissue, and its expression is regulated by the metabolic status and glucose levels [320, 389–392]. Mutations of the *NNAT* gene are linked to extreme childhood obesity and its expression is reduced in the adipose tissue of obese children [391, 393]. Moreover, Nnat regulates proinsulin cleavage and the homozygous deletion of *Nnat* results in reduced insulin content and glucose-stimulated insulin secretion in pancreatic β-cells [392]. Thus, Nnat exerts glucose-sensitive control of β-cell function [392]. In addition, we found that POSTN is upregulated in both, $PDX1^{P33T/+}$and $PDX1^{P33T/P33T}$mutant PPs. Periostin, which is encoded by the *Postn* gene, is a secreted protein, which is involved in epithelial-mesenchymal

transformation, and plays a role in the development of teeth, bone, and heart [394]. In the pancreas, *Postn* expression is highly upregulated upon partial pancreatectomy and is involved in the regeneration of β-cells [322]. Moreover, when Postn is injected intraperitoneally, an increase in islet numbers and beneficial effects on glucose regulation can be observed [322]. Therefore, it is clear that *NNAT* and *POSTN* have important functions in the endocrine pancreas, but their role during pancreas development needs to be addressed by future studies.

3.3 Foxa2 Expression Resolves Endocrine Differentiation in the Intestine

In response to food intake, enteroendocrine cells (EECs) of the intestinal epithelium secrete a multitude of hormones, which regulate various aspects of energy homeostasis, including feeding behavior, by regulating appetite and satiety, as well as digestion, through stimulating the secretion of gastric acid, bile acids and digestive enzymes from the pancreas [161]. Importantly, the hormones Glp1 and Gip, secreted from L- and K-cells, respectively, mediate the incretin effect, which is characterized by the amplification of insulin secretion from pancreatic β-cells. This effect accounts for up to 70% of the total insulin response, and is thus an integral part of the glycemic control [176]. These features generated great interest in EECs and the intestinal hormones, as they provide novel routes to treat metabolic diseases such as T2DM. Indeed, the first generation of incretin-based drugs, including long-acting Glp1 analogues or Dpp4 inhibitors, are already successfully used to treat diabetic patients [27, 161, 177].

EECs constantly differentiate from intestinal stem cells (ISCs) and the principles underlying their differentiation are remarkably similar to those seen in pancreatic endocrine cell formation and involve partially the same transcriptional regulators. Even though great advances have been made recently, the mechanisms governing EEC differentiation and subtype specification are still poorly understood [162, 191, 192]. However, it is of particular interest to decipher these regulatory mechanisms, as the intestinal epithelium holds great potential for cell-based therapies. For instance, the high similarity to the pancreatic endocrine system provides an interesting perspective as it allows the transdifferentiation of EECs into insulin producing β-like cells [395, 396]. Moreover, a recent study has shown that the hormonal program of EECs is dynamic and can change in response to BMP signaling [192], suggesting possibilities for drug-based modulations of the hormonal program. However, exact knowledge of the signals and transcriptional programs, which control EEC development, is necessary to manipulate EECs in order to produce specific hormones for the treatment of metabolic diseases.

The pioneer TF Foxa2 is essential for endoderm specification during embryogenesis and the development of endoderm-derived organs [288]. In the intestine, Foxa2 and its family member Foxa1 have been shown to be important for the differentiation of GCs and a subset of EECs, namely Glp1/2 producing L-cells and Sst producing D-cells [272]. In this thesis, we discovered that Foxa2 is expressed at different levels within the crypt by using a previously established fluorescent Foxa2-Venus fusion reporter mouse line. Transcriptome analysis revealed that ISCs, mature goblet cells (GCs) and Paneth cells (PCs) expressed low levels of Foxa2, while EECs and GC progenitors expressed Foxa2 at high levels. In contrast, tuft cells (TCs) and enterocytes did

not express Foxa2. Hence, the FVF mouse model enabled us to enrich for ISCs and secretory cells and to study EEC differentiation. Using a combination of bulk and single-cell transcriptomics, together with ChIP-seq and ATAC-seq, we then dissected the EECs lineage decision and described the role of Foxa2 and its impact on the chromatin landscape during EEC differentiation.

3.3.1 Similarity of the Enhancer Landscape in FVF^{low} and FVF^{high} Cells

Several studies comparing the epigenetic landscape in ISCs, progenitors and mature villus cells, have described a high similarity between the intestinal cell populations on the level of histone modifications and DNA methylation. Moreover, it was reported that the chromatin is generally in an open, permissive state that is maintained throughout differentiation [222–226]. For instance, the active histone marks H3K4me2 and H3K27ac are present on the same loci in secretory and absorptive progenitors, even though gene expression from these loci is markedly different [222, 224]. In our ChIP-seq data from FVF^{low} and FVF^{high} cells, we observed the same nearly identical configuration of the active histone modifications H3K4me3 and H3K27ac that were also present on lineage specific loci, regardless of gene expression. Interestingly, as in previous reports [222], some exceptions were observed at the loci of important enteroendocrine lineage determining genes, such as the TFs *Neurog3* or *Nkx2-2*, which showed pronounced differences in H3K4me3 and H3K27ac enrichment between FVF^{low} and FVF^{high} cells. Despite the general similarity, we found that the enrichment of H3K27ac is subject to slight changes between the two FVF populations, and that these changes correlate with the expression of the corresponding genes. However, due to the low numbers of cells obtained from the FVF^{high} population, it was not feasible to perform the appropriate number of replicate experiments required for reliable quantification of enrichment differences. Therefore, these results should be seen as a qualitative, rather than a quantitative, evaluation. The low number of available cells, together with the high similarity of histone modifications, necessitated a different approach to assess lineage commitment on the level of epigenetics.

To overcome the present limitations, we performed ATAC-seq to profile chromatin accessibility as approximation to chromatin activity [331]. The assay requires two orders of magnitude lower numbers input of cells and could therefore be performed in replicates, allowing a quantitative analysis. Similar to the histone modification data, chromatin opening was largely concordant between FVF^{low} and FVF^{high} cells, but differential accessibility analysis identified a set of 1344 regions with significant differences. The vast majority of these sites was found in putative enhancer regions, which is in agreement with previous observations, showing that cell type-specific chromatin modifications are mainly found at enhancer regions, while promoter chromatin modifications are generally cell type-invariant [330]. More importantly, regions with differentially open chromatin were found near lineage specific genes, such as the endocrine genes *Neurog3*, *Neurod1* or *Fev* in FVF^{high} cells or the ISC-specific genes *Olfm4* and *Ascl2*. Thus, the ATAC-seq data reflected the lineage-specificity observed in the expression data and suggests that the different cell types of the crypt can be distinguished on the level of chromatin accessibility.

3.3.2 Dose-Dependent Binding of Foxa2

FoxA TFs have been shown to occupy their target sites in a cell type and developmental stage-specific mode [60, 348, 397, 398]. The mechanisms allowing the specific binding of FoxA factors are not understood, though results from invertebrates have suggested a possible model. In *C. elegans*, the Foxa2 homolog defective pharyngeal development protein 4 (PHA-4) is, as the name suggests, essential for pharyngeal development [399]. It regulates many genes during multiple stages of pharyngeal development, and the onset of target gene expression is regulated by the affinity of PHA-4 to its target sites. During early stages of development, PHA-4 is expressed at low levels and binds to high affinity binding sites. As the expression of PHA-4 rises in later stages of pharyngeal development, PHA-4 binds to low affinity binding sites. Thus, PHA-4 regulates its target genes in a dose-dependent manner, and the affinity to the target sequence seems to be the main determinant for DNA binding [400]. Similarly, analysis of promoter and enhancer sequences bound by Foxa2 in the adult mouse liver, revealed that genes with weaker consensus motifs were more liver-specific than those with stronger binding motifs [401]. However, the situation in vertebrates appears to be more complex, as the strength of the consensus sequence does not determine Foxa2 occupancy [344].

Analysis of the Foxa2 binding data from FVF^{low} and FVF^{high} cells showed that also in the intestine Foxa2 binds its target genes in a cell type and dose-specific manner. Upon the $\sim 2\times$ increase in Foxa2 expression in the FVF^{high} population, the number of Foxa2-bound sites almost doubled compared to the FVF^{low} population. Using the ChIP-seq signal strength as approximation for the amount of bound Foxa2, and thus as measure for the overall affinity of Foxa2 to a given target site, we found that population-specific sites were bound with a lower affinity. In our binding profile, the mean enrichment of Foxa2 in FVF^{high} cells at common binding sites was $1.8\times$ higher compared to binding sites exclusively bound in FVF^{high} cells (data not shown). Evaluation of electrophoretic mobility shift assay data [402] showed a good correlation between the motif score of a given sequence and the binding strength of Foxa2 *in vitro* ($\rho = 0.73$). However, we could not confirm this correlation in our *in vivo* data. Extensive analysis of the consensus motifs present in Foxa2 binding sites in FVF^{high} and FVF^{low} cells did not reveal any correlation between the motif scores and the occupancy or binding strength of Foxa2 in both FVF populations. Thus, it seems that FVF^{high}-specific sites are bound with lower affinity, but that the consensus sequence is not the main determinant of Foxa2 binding strength, suggesting that other features are more important in the determination of Foxa2 binding. A possible mechanism would be that co-binding TFs influence the stability of Foxa2 at its target sites. This is supported by the observation that sites bound by Foxa2 and Atoh1 tend to show a stronger signal in FVF^{high} cells. However, it should be taken into account that it is not clear to which extent the ChIP-seq signal strength reflects the actual amount of bound protein, especially in a mixed cell population assayed by bulk ChIP-seq.

3.3.3 Foxa2 as Pioneering Factor of the Endocrine and Goblet Cell Fates

Inactive chromatin is packed with nucleosomes, which in general prevent binding of TFs. However, a subset of TFs possess the ability to bind their target sequences even when they are

bound by a nucleosome, and induce local chromatin decompaction, which allows other TFs to bind [292, 344]. Since these TFs are the first to bind to closed inactive chromatin, they have been termed pioneer TFs [289]. FoxA and GATA4 were the first pioneer TFs to be described, as they were able to open compact chromatin during hepatic development [289]. The ability of FoxA factors to open condensed chromatin is conferred by the C-terminal domain, which can bind to the core histones H3 and H4, and more importantly by the structural similarity of the DNA binding domain to the linker histones H1 and H5 [285, 289]. This feature allows FoxA factors to displace linker histones by direct competition for DNA binding, which in turn increases the accessibility of target nucleosomes [292]. Thus, FoxA pioneer factors can act as master regulators of cell fate decisions as they are able to bind inactive chromatin of cell type-specific genes and impart the induction of cell type-specific gene expression [343].

Our data shows that upon increased expression in FVF^{high} cells, Foxa2 binds more than 9000 additional sites, not previously bound in FVF^{low} cells, and that the accessibility of the surrounding chromatin increases with Foxa2 binding. Strikingly, the forkhead motif had the second highest association with FVF^{high}-specific chromatin opening, and we found that 70% of the sites specifically open in FVF^{high} cells, were actually bound by Foxa2. Moreover, the increase of chromatin accessibility was strongest at FVF^{high}-only Foxa2 binding sites, and vice versa, Foxa2 binding showed the strongest increase at sites with FVF^{high}-specific accessibility. Together, several lines of evidence show a strong correlation between Foxa2 binding and chromatin opening in FVF^{high} cells. Therefore, our results suggest that Foxa2 acts as pioneering factor in FVF^{high} cells. Upon increased Foxa2 expression, cell type-specific target sites are bound and chromatin accessibility is promoted in order for other TFs to bind. According to our motif analysis, basic helix-loop-helix (bHLH) TFs, such as Atoh1, Neurod1 or Ngn3, which are important for EEC and GC differentiation, might be the main FVF^{high}-specific co-binding TFs. This indicates, that chromatin opening facilitated by Foxa2 binding is an important step in the initiation of EEC and GC differentiation, setting the stage for lineage-specific TFs.

3.3.4 Modulation of Foxa2 Function by Cooperating Transcription Factors

The cell-type specific regulation of transcription is achieved through the coordinated action of co-binding TFs at enhancer elements [330, 346]. It has been shown that Foxa2 cooperates with multiple TFs in different tissues. In the pancreas, for instance, Foxa2 works together with Pdx1, Ngn3, Insm1 and Neurod1 to regulate pancreas and endocrine development, as well as β-cell function [87, 347, 348]. In the liver, on the other hand, Foxa2 cooperates with Hnf4a to regulate a liver-specific gene program [348]. The combined analysis of Foxa2 occupancy and motif sequences at the binding sites revealed a relation between the identified motifs and the ChIP-seq signal strength. Nuclear receptor motifs, such as the Hnf4a and Pparg motifs, were associated with the FVF^{low} population. In contrast, bHLH and RFX motifs, corresponding to secretory and endocrine TFs like Ngn3, Neurod1, Atoh1 or Rfx6, were associated with increased Foxa2 signals in the FVF^{high} population. This suggests that Foxa2 cooperates with different partners in FVF^{low} and FVF^{high} cells, respectively. In agreement with this hypothesis, we found a significant overlap of Foxa2-bound sites with Atoh1, Hnf4a and Cdx2 binding sites, and that the Foxa2 signal

intensity at shared binding sites differed between the potential binding partners. While sites co-occupied with Foxa2 and Atoh1 showed increased Foxa2 enrichment in FVF^{high} cells, sites shared between Foxa2 and Hnf4a or Foxa2 and Cdx2 did not show changes in Foxa2 signal intensity, suggesting that the cooperation between Foxa2 and Atoh1 is specific to the FVF^{high} cells. This notion is also supported by chromatin accessibility data, which showed the strongest increase in chromatin opening at sites co-bound by Foxa2 and Atoh1. Moreover, pathway analysis indicates that genes near Foxa2–Atoh1 co-occupied sites are involved in EEC and GC differentiation. On the other hand, Foxa2–Hnf4a and Foxa2–Cdx2 co-binding appears to be present in both populations and these sites are among the most accessible in both FVF populations, with only minor changes between FVF^{low} and FVF^{high} cells. Genes near Foxa2–Hnf4a sites were associated with metabolic processes, and those near Foxa2–Cdx2 sites with the differentiation of enterocytes. The latter finding seems to contradict the observation that enterocyte progenitors cease Foxa2 expression. However, Cdx2 is an important driver of enterocyte differentiation [403, 404] and is required for the maintenance of cell type-specific chromatin accessibility [345, 405]. As Cdx2 binding relies on open chromatin [405], the pioneering activity of Foxa2 might be required to induce chromatin accessibility, which is then maintained by Cdx2. This is also in agreement with the finding that FoxA factors are necessary for the induction of the liver fate during development but not for the maintenance of the liver identity [295, 406, 407]. Taken together, our results indicate that Foxa2 cooperates with different TFs in a cell type-specific manner, and that the binding partner directs the function of Foxa2. With respect to cell type specification in the intestine, our data suggests that Foxa2 and Atoh1 together regulate EEC and GC differentiation, while Foxa2 together with Cdx2 initiates enterocyte differentiation. The evidence shown in this work, however, is solely based on the overlap of detected binding sites and we did not attempt to show the physical interaction of these factors at the shared binding sites. Hence, further experiments using sequential ChIP are required to prove the concurrent binding, and thus the cooperation of Foxa2 and its potential partners.

3.3.5 New Roles of Foxa2 in the Intestinal Epithelium

The conditional deletion of *Foxa1* and *Foxa2* in the intestinal epithelium results in a reduction of Sst and Pyy producing D- and L-cells, and the loss of Glp1/2 producing L-cells. On the mRNA level, reduced expression of the endocrine TFs *Isl1* and *Pax6* was detected in *Foxa1/2* knockouts. This suggests that in EEC differentiation, Foxa2 acts downstream of Ngn3 and upstream of Isl1 and Pax6 to regulate D- and L-cell differentiation [272]. In contrast, during pancreatic endocrine differentiation, Foxa2 acts upstream of Ngn3, and cooperates with Ngn3 to regulate *Neurog3* expression [87]. As Ngn3 is the master regulator of endocrine differentiation, its expression is tightly controlled by the cooperative action of multiple TFs, including Pdx1, Onecut1, Hnf1b, Myt1, Sox9, Ngn3 itself, and Foxa2. Since the TF networks guiding endocrine differentiation in the pancreas and intestine are very similar, we hypothesized that also during EEC differentiation Foxa2 acts upstream of Ngn3. Indeed, our analysis showed that Foxa2 binds to two sites in a conserved region from -5000 to -3500 bp upstream of the Ngn3 promoter, which also contains the Foxa2 binding site active in pancreatic endocrine progenitors [87, 347]. Moreover, these sites are only occupied in the FVF^{high} population, suggesting that they require the presence

of high Foxa2 levels in order to be bound. This is also in agreement with the observation of increased Foxa2 levels in pancreatic endocrine progenitors that correlate with the expression of Ngn3 [87]. Thus, our results indicate that Foxa2, when expressed at high levels, is involved in the regulation of *Neurog3* expression. Furthermore, our ChIP-seq, microarray and scRNA-seq data show that upregulation of Foxa2 precedes the onset of expression of many known regulators of endocrine differentiation, including *Neurod1, Nkx2-2, Insm1* and *Arx*, which are bound by Foxa2 and upregulated in FVF^{high} cells. Therefore, we suggest a broad role for Foxa2 in onset and progression of EEC differentiation.

Similarly, we identified the GC TFs *Spdef, Klf4* and *Creb3l1* as potential Foxa2 target genes, suggesting that Foxa2 is involved in the differentiation of GCs. Our single-cell and bulk transcriptome data revealed that GC progenitors express high levels of Foxa2, indicating that also the onset of GC differentiation depends on high Foxa2 expression. Moreover, we show that the expression of the GC TFs *Spdef, Klf4* and *Creb3l1* follows the upregulation of Foxa2 in GC progenitors, placing Foxa2 upstream of these factors. This is in accordance with the notion that GC differentiation is impaired in conditional *Foxa1/2* knockout mice [272]. However, it shows that Foxa2 not only regulates *Mucins* and *Tff3*, but is also important for the onset of GC differentiation.

In addition to the roles of Foxa2 in EEC and GC differentiation, our data suggests that Foxa2 might also be important in ISCs. Foxa2 is expressed in non-PC cells in the stem cell zone, and lineage tracings of Foxa2-positive crypt cells using a $Foxa2^{nEGFP\text{-}CreERT2/+}$;$Gt(ROSA)26^{mTmG/+}$ mouse line showed that Foxa2 expression cells are capable of generating all intestinal cell types and the lineage tracing patterns resemble those observed for $Lgr5^{+}$ ISCs [326]. Moreover, we could show that Foxa2 binds to the majority of the ISC signature genes [342], suggesting that Foxa2 is involved in the regulation of an ISC gene program. However, the deletion of *Foxa1/2* from the intestinal epithelium did not reveal obvious defects in the ISC compartment [272]. A possible explanation for this results is that in the *Foxa1/2* knockout, the third family member Foxa3 might be able to compensate for the loss of Foxa1 and Foxa2. In fact, Foxa3 expression levels in the intestine exceed the expression levels of both, Foxa1 and Foxa2, and it has been shown that Foxa3 is able to bind to Foxa1/2-specific binding sites in the absence of both TFs [292]. Thus, a scenario in which compensation by Foxa3 prevents a severe phenotype in ISCs of *Foxa2/1* null mice seems to be likely. To test this idea, a triple knockout of all FoxA factors would be required.

3.4 Conclusion

Hormones secreted by endocrine cells of the pancreas and intestine are crucial for the homeostatic control of blood glucose and energy intake. Disturbances of the hormonal balance are implicated in the development of metabolic diseases, such as T2DM. Both organs are derived from the endoderm and share common mechanisms of cell fate decisions. Better understanding of TF networks, controlling the development of pancreatic and intestinal endocrine cells, is required to develop novel treatment strategies for complex diseases such as diabetes.

In the first part of this thesis, we assessed the role of PDX1 during early pancreatic development and how PDX1 function is affected by two common point mutations. Using PPs derived from a novel iPSC line, we characterized genome-wide PDX1 binding sites. We were able to show that PDX1 acts in a stage-specific manner, and targets a PP-specific program, governing early pancreatic differentiation, during development, while it switches to β-cell specific targets, regulating β-cell function and identity, in adulthood. Furthermore, we showed that T2DM-associated SNPs are enriched within active regulatory regions of PPs, indicating that early pancreatic development affects the susceptibility to T2DM. We then studied PPs generated from two donor-derived iPSCs carrying heterozygous C18R and P33T mutations in the PDX1 transactivation domain, respectively, and the corresponding homozyogous isogenic control cell lines, and showed that these mutations contribute to the predisposition to T2DM. Furthermore, our data suggests that PDX1 affects pancreas development and β-cell function in a dose-dependent fashion and that developmental and compensatory β-cell formation and expansion depends on fully functional PDX1.

The second part of the thesis addresses the role of Foxa2 in the onset of EEC differentiation in the intestine. We found that Foxa2 has cell type-specific functions in ISCs and secretory progenitor cells, which is mediated through the expression levels of Foxa2 and cooperating TFs. In ISCs, low levels of Foxa2 are involved in the regulation of metabolic processes and, presumably, the initiation of enterocyte differentiation, while in GC and EEC progenitors high levels of Foxa2 pioneer the TF networks that control the differentiation into GCs and EECs. These results suggest an important role of Foxa2 in the initiation of EEC differentiation and provide insight into the dose-dependent action of Foxa2. Future experiments will show if this mechanism can be used to induce the differentiation of Glp1 producing L-cell, through the stimulation of Foxa2 expression. This could be achieved by the inhibition of ROCK signaling, as previous work from our lab has shown that the ROCK inhibitors Fasudil and RKI-1447 can induce *Foxa2* expression in ESCs [408].

4 Materials & Methods

4.1 Antibodies

4.1.1 Primary Antibodies

Antibody	Host	Company	Catalog Number	Application
ChgA polyclonal	Rabbit	Abcam	ab15160	IF (1:200)
ChgA polyclonal	Goat	Santa Cruz	sc-1488	IF (1:200)
Dclk1 polyclonal	Rabbit	Abcam	ab37994	IF (1:500)
Foxa2 polyclonal	Goat	Santa Cruz	sc-6554	ChIP-seq
GFP polyclonal	Chicken	Aves Labs	GFP-1020	IF (1:500)
GFP polyclonal	Rabbit	Abcam	ab290	ChIP-seq
H3K4me3 polyclonal	Rabbit	Diagenode	C15410003-50	ChIP-seq
H3K27ac polyclonal	Rabbit	Diagenode	C15410174	ChIP-seq
Lyz1 polyclonal	Rabbit	DAKO A0099	IF (1:1000)	
Muc2 polyclonal	Rabbit	Santa Cruz	sc-7314	IF (1:1000)
PDX1 polyclonal	Goat	kindly provided by	C. Wright	ChIP-seq

4.1.2 Secondary Antibodies

Antibody	Host	Company	Catalog Number	Application
Anti-chicken Alexa Fluor 488	Donkey	Dianova	703-225-155	IF (1:800)
Anti-goat Alexa Fluor 555	Donkey	Invitrogen	A21432	IF (1:800)
Anti-goat Alexa Fluor 633	Donkey	Invitrogen	A21082	IF (1:800)
Anti-mouse Cy5	Donkey	Dianova	715-175-151	IF (1:800)
Anti-rabbit Alexa Fluor 488	Donkey	Invitrogen	A21206	IF (1:800)
Anti-rabbit Alexa Fluor 555	Donkey	Invitrogen	A31572	IF (1:800)
Anti-rabbit Alexa Fluor 649	Donkey	Dianova	711-605-152	IF (1:800)

4.2 Buffers & Media

4.2.1 Immunofluorescence

10× PBS (pH7.4)	1.37 M NaCl, 26.8 mM KCl, 0,1 M Na_2HPO_4, 13.8 mM KH_2PO_4
PBST	1× PBS, 0.1% Tween20, pH7.4
4% PFA	1.3 M PFA in 1× PBS, pH7.2-7.4
Permeabilisation	0.25% TritonX-100, 100 mM glycin in dH_2O
Blocking solution	10% FCS, 1% BSA, 3% donkey serum in PBST
DAPI	5 mg DAPI in 25 ml 1× PBS
Elvanol	0.015 mM Polyvinyl-alcohol, 24 mM Tris pH6.0, 2 g DABCO in 90 ml H_2O and 37.8 ml glycerol

4.2.2 Crypt Isolation & FACS

Crypt isolation buffer	2 mM EDTA in DPBS
Basal Crypt Medium	AdvancedDMEM/F12 containing 100 U/ml Penicillin, 100 U/ml Streptomycin and 10 mM HEPES
Complete Crypt Medium	AdvancedDMEM/F12 containing 100 U/ml Penicillin, 100 U/ml Streptomycin, 10 mM HEPES and 10% FCS
FACS buffer	1× DPBS, 2% FCS and 2 mM EDTA
FACS buffer (scRNA-seq)	1× DPBS, 1% FCS and 0.1 mM EDTA

4.2.3 ChIP-seq

Lysis-Buffer 0.5% SDS	50 mM Tris-HCl (pH 8.0), 10 mM EDTA, 0.5% SDS
IP-Buffer	10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.5 mM EGTA, 1% Triton X-100, 0.1% SDS, 0.1% Na-Desoxycholate, 140 mM NaCl, H_2O, Protease Inhibitors
Washing-Buffer 1 (High Salt Buffer)	500 mM NaCl, 50 mM Tris-HCl (pH 8.0), 0.1% SDS, 1% NP-40
Washing-Buffer 2 (LiCl Buffer)	250 mM LiCl, 50 mM Tris-HCl (pH 8.0), 0.5% Na-Deoxycholate, 1% NP-40
Washing-Buffer 3	10 mM Tris-HCl (pH 8.0), 10 mM EDTA
Elution Buffer	50 mM Tris-HCl (pH 8.0), 10 mM EDTA, 1% SDS
TE Buffer	10 mM Tris-HCl (pH 8.0), 1 mM EDTA

4.2.4 ATAC-seq

Lysis-Buffer	10 mM Tris-HCl, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% (v/v) IGEPAL CA-630
Elution-Buffer	10 mM Tris-HCl, pH 8.0

4.3 Kits & Master Mixes

Agilent High Sensitivity DNA Kit	Agilent	5067-4626
Agilent RNA 6000 Pico Kit	Agilent	5067-1513
Chromium Single Cell 3' Library & Gel Bead Kit v2	10× Genomics	120237
Chromium Single Cell A Chip Kit, 48 rxns	10× Genomics	120236
Encore Biotin Module	NuGEN	4200-12
GeneChip Human Gene 2.0 ST Array	Affymetrix	902113
GeneChip Mouse Gene 1.0 ST Array	Affymetrix	901168
High-Capacity cDNA Reverse Transcription Kit	Applied Biosystems	4368814
MicroPlex Library Preparation Kit	Diagenode	C05010010
MinElute PCR Purification Kit	Qiagen	28004
miRNeasy Mini Kit	Qiagen	217004
NEBNext High-Fidelity 2X PCR Master Mix	New England Biolabs	M0541S
Nextera DNA Library Preparation Kit	Ilumina	discontinued
Ovation PicoSL WTA System V2	NuGEN	3312-24
QIAquick PCR Purification Kit	Qiagen	28104
Qubit dsDNA HS Assay Kit	Molecular Probes	Q32854
RNase-Free DNase Set	Qiagen	79254
SuperScript VILO cDNA Synthesis Kit	Invitrogen	11754250
SYBR Green PCR Master Mix	Applied Biosystems	4364346
TaqMan Fast Advanced Master Mix	Applied Biosystems	4444965
TruSeq Stranded mRNA Library Prep	Illumina	20020594

4.4 Reagents & Media

Advanced DMEM/F12	Gibco	12634028
AMPure XP	Beckman Coulter	A63880
DNase I	Roche	10104159001
Donkey serum	Millipore	S30-100ML
DPBS, no calcium, no magnesium	Gibco	14190250
FBS Good	PAN-Seratech	ST40-37500
HEPES (1 M)	Gibco	15630080
Nuclease-free water	Promega	P119C
OCT Tissue freezing medium	Leica Biosystems	14020108926
Penicillin-Streptomycin (10,000 U/mL)	Gibco	15140122
SPRIselect	Beckman Coulter	B23318
TrypLE Express Enzyme (1×)	Gibco	12605010

4.5 Chemicals

Unless indicated otherwise, chemicals were purchased from Merck, Thermo Fisher Scientific or Carl Roth.

10N HCl		
7-AAD Viability Staining Solution	eBioscience	00-6993-50
BSA		
Calcium chloride		
DABCO		
DAPI		
EDTA		
EGTA		
Ethanol, 96%		
Ethanol, Pure (200 Proof, anhydrous)		
EvaGreen Dye, 20×in Water	Biotium	#31000
Formaldehyde solution (37%)		
Glycerol, 50% (v/v) Aqueous Solution	Ricca Chemical Company	3290-32
Glycin		
IGEPAL CA-630		
Lithium chloride		
$MgCl_2$	New England Biolabs	B0510A
NP-40	Life Technologies	
Paraformaldehyde		
Polyvinyl-alcohol		
Potassium chloride (KCl)		
Potassium hydrogenphosphate (KH_2PO_4)		
QIAzol Lysis Reagent	Qiagen	79306
RNaseZAP		
Sodium chloride		
Sodium desoxycholate		
Sodium dodecylsulphate (SDS)		
Sodium hydrogenic phosphate (Na_2HPO_4)		
TE Buffer	Invitrogen	12090015
Tris		
Triton X-100		
Trypan Blue solution		
Tween-20		

4.6 Primers & TaqMan Probes

4.6.1 ATAC-seq Sequencing Adapter

Ad1	AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATGTG
Ad2.1	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTCGGAGATGT
Ad2.2	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGGAGATGT
Ad2.3	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGGAGATGT
Ad2.4	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGGAGATGT
Ad2.5	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGGAGATGT
Ad2.6	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGGAGATGT
Ad2.7	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCTCGGAGATGT
Ad2.8	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTCTCGTGGGCTCGGAGATGT
Ad2.9	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTCTCGTGGGCTCGGAGATGT
Ad2.10	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTCTCGTGGGCTCGGAGATGT
Ad2.11	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTCTCGTGGGCTCGGAGATGT
Ad2.12	CAAGCAGAAGACGGCATACGAGATTCCTCTACGTCTCGTGGGCTCGGAGATGT
Ad2.13	CAAGCAGAAGACGGCATACGAGATATCACGACGTCTCGTGGGCTCGGAGATGT
Ad2.14	CAAGCAGAAGACGGCATACGAGATACAGTGGTGTCTCGTGGGCTCGGAGATGT
Ad2.15	CAAGCAGAAGACGGCATACGAGATCAGATCCAGTCTCGTGGGCTCGGAGATGT
Ad2.16	CAAGCAGAAGACGGCATACGAGATACAAACGGGTCTCGTGGGCTCGGAGATGT
Ad2.17	CAAGCAGAAGACGGCATACGAGATACCCAGCAGTCTCGTGGGCTCGGAGATGT
Ad2.18	CAAGCAGAAGACGGCATACGAGATAACCCCTCGTCTCGTGGGCTCGGAGATGT
Ad2.19	CAAGCAGAAGACGGCATACGAGATCCCAACCTGTCTCGTGGGCTCGGAGATGT
Ad2.20	CAAGCAGAAGACGGCATACGAGATCACCACACGTCTCGTGGGCTCGGAGATGT
Ad2.21	CAAGCAGAAGACGGCATACGAGATGAAACCCAGTCTCGTGGGCTCGGAGATGT
Ad2.22	CAAGCAGAAGACGGCATACGAGATTGTGACCAGTCTCGTGGGCTCGGAGATGT
Ad2.23	CAAGCAGAAGACGGCATACGAGATAGGGTCAAGTCTCGTGGGCTCGGAGATGT
AP2.24	CAAGCAGAAGACGGCATACGAGATAGGAGTGGGTCTCGTGGGCTCGGAGATGT

4.6.2 qPCR TaqMan Probes

Human Probes	
18S	Hs99999901_s1
GAPDH	Hs02758991_g1
HNF1B	Hs01001602_m1
NANOG	Hs04260366_g1
NKX2.2	Hs00159616_m1
NKX6.1	Hs01055914_m1
OCT4	Hs00999632_g1
PDX1	Hs00236830_m1
PTF1A	Hs00603586_g1
SOX9	Hs01001343_g1

Mouse Probes	
Chga	Mm00514341_m1
Fabp1	Mm00444340_m1
Foxa1	Mm00484713_m1
Foxa2	Mm01976556_s1
Foxa3	Mm00484714_m1
Gapdh	Mm99999915_g1
Isl1	Mm00517585_m1
Lgr5	Mm00438890_m1
Neurod1	Mm01280117_m1
Neurog3	Mm00437606_s1
Nkx2-2	Mm00839794_m1
Olfm4	Mm01320260_m1
Pax6	Mm00443081_m1
Rfx3	Mm00803303_m1
Rfx6	Mm00624115_m1
RN18S	Mm03928990_g1
Rpl37	Mm00782745_s1

4.6.3 SYBR Green qPCR Primer

ACTB forward	`CCCAGAGCAAGAGAGG`
ACTB reverse	`GTCCAGACGCAGGATG`
HHEX forward	`ACGGTGAACGACTACACGC`
HHEX reverse	`CTTCTCCAGCTCGATGGTCT`
MEIS1 forward	`GGGCATGGATGGAGTAGGC`
MEIS1 reverse	`GGGTACTGATGCGAGTGCAG`
ONECUT1 forward	`GAACATGGGAAGGATAGAGGCA`
ONECUT1 reverse	`GTAGAGTTCGACGCTGGACAT`
RFX6 forward	`AAGCAGCGGATCAATACCTGT`
RFX6 reverse	`ACCGTGGTAAGCAAACTCCTT`

4.7 Instruments & Equipment

2100 Bioanalyzer	Agilent
Applied Biosystems ViiA 7	Thermo Fisher Scientific
C1000 Touch Thermal Cycler	Bio Rad
Centrifuge 5417 R	Eppendorf
Centrifuge 5424 R	Eppendorf
Centrifuge 5804 R	Eppendorf
Chromium Controller	10× Genomics
FACSAria III	BD

Pipet-Lite LTS Pipette L-1000XLS+	Rainin
Pipet-Lite LTS Pipette L-100XLS+	Rainin
Pipet-Lite LTS Pipette L-10XLS+	Rainin
Pipet-Lite LTS Pipette L-200XLS+	Rainin
Pipet-Lite LTS Pipette L-20XLS+	Rainin
Pipet-Lite LTS Pipette L-2XLS+	Rainin
Pipet-Lite Multi Pipette L8-200XLS+	Rainin
Pipet-Lite Multi Pipette L8-50XLS+	Rainin
PIPETMAN Neo P1000N	Gilson
PIPETMAN Neo P10N	Gilson
PIPETMAN Neo P200N	Gilson
PIPETMAN Neo P20N	Gilson
PX1 PCR Plate Sealer	Bio Rad
Qubit Fluorometer	Molecular Probes
S220 Focused-ultrasonicator	Covaris
ThermoMixer C	Eppendorf

4.8 Consumables

10 cm bacterial plates	BD
50 ml/ 15 ml Falcon tubes	Corning
5PRIME Phase Lock Gel - Heavy	Quantabio
96- and 384-Well plates	Thermo Scientific
DNA LoBind Tubes, 1.5 ml	Eppendorf
Falcon 5 mL Round Bottom High Clarity PP Test Tube, with Snap Cap	Corning
Falcon 5 ml Round Bottom Polystyrene Tube, with Cell Strainer Snap Cap	Corning
Falcon Cell Strainers, 70 µm	Corning
PCR Tubes 0.2 ml 8-tube strips	Eppendorf
Pierceable Foil Heat Seal	Bio-Rad
Tips LTS 1ML Filter RT-L1000FLR	Rainin
Tips LTS 200UL Filter RT-L200FLR	Rainin
Tips LTS 20UL Filter RT-L10FLR	Rainin
twin.tec 96-Well PCR Plate Semi-skirted	Eppendorf

4.9 General Mouse Handling & Animal Studies Approvements

All animal experiments were carried out in compliance with the German Animal Protection Act, the guidelines of the Society of Laboratory Animals (GV-SOLAS) and Federation of Laboratory Animal Science Associations (FELASA). Mice were housed in groups of two to four animals at 23±1°C, with constant humidity and on a 12 hours light–dark cycle with free access to water and food.

4.10 Genetic Lineage Tracing

For genetic lineage tracing experiments, Cre recombinase activity was induced in $Foxa2^{nEGFP\text{-}CreERT2/+}$;$Gt(ROSA)26^{mTmG/+}$ mice by a single dose of orally administered tamoxifen at 0.3 mg/g body weight in sunflower oil (Sigma-Aldrich) after a 3 hours fasting period. Mice were sacrificed 48 hours or 30 days after tamoxifen gavage.

4.11 Immunofluorescence

For immunofluorescence stainings on cryo-preserved tissue sections, intestinal tissue was dissected and rinsed in ice-cold PBS. To remove the luminal content, the intestinal lumen was carefully flushed with PBS. Cleaned tissues were fixed with 4% paraformaldehyde (PFA) for 3 h at 4°C and incubated in a sucrose gradient (1 h in 5% sucrose/PBS, followed by 1 h in 15% sucrose/PBS at RT, and 30% sucrose/PBS overnight at 4°C). Prior to embedding with OCT (optimal cutting temperature) embedding medium (Leica), tissue samples were incubated in a 1:1 mixture of 30% sucrose/PBS and OCT for 3 h at RT. Subsequently, tissue samples were placed into embedding molds, frozen on dry ice and stored at -80°C. Tissue samples were sectioned on a cryostat (Leica CM 1860) at a thickness of 10–16 µm. Cryo-sections were then dried for ∼40 min at RT, rehydrated with PBS for 20 min and permeabilised with 0.4% Triton X-100/PBS for 20 min. After washing once with PBST (0.1% Tween-20/PBS) samples were blocked in using a blocking solution containing 10% FCS, 0.1% BSA and 3% donkey or goat serum/PBST for 1–2 h at RT. Sections were then incubated with primary antibodies, diluted in blocking solution, overnight at 4°C. Sections were washed 3× with PBS before incubation with secondary antibodies in blocking solution for 1 h at RT. Section were again washed 3× with PBS and nuclei were stained with DAPI (1:1000) in PBS for 15 min at RT. Following nuclear staining, sections were washed at least 3× and subsequently embedded with self-made Elvanol antifade agent. Sections were kept overnight at RT for drying.

4.12 Crypt Isolation

Small intestines were dissected and remaining adipose tissue was removed in ice-cold PBS. After that, intestines were opened longitudinally using scissors, and villi and the lunimal content were carefully scraped off using a cover slip. The opened intestines were then cut into ∼2

cm long pieces and washed 2–4× in ice-cold PBS until supernatant was clear. Subsequently, tissue fragments were incubated in 2 mM EDTA in ice-cold PBS for 35 min at 4 °C on a benchtop roller. Crypts were then collected in fresh PBS by rigorous shaking (3–4 rounds). To remove remaining large villus fragments, isolated crypts were passed through a 70 µm cell strainer and subsequently pelleted at 300×g for 5 min at 4°C. To prepare single-cell suspensions, isolated crypts were incubated with 1–2 ml TrypLE (Life technologies) for 5 min at 37 °C in a water bath. To stop the TrypLE reaction and to digest released DNA, 5–7 ml of complete crypt medium (basal medium supplemented with 10% FCS) supplemented with 10 µg/ml DNase I was added and cells were incubated for 5–10 min at 37 °C in a water bath, while being resuspended once every ∼2 min. After DNA was digested completely, cells were resuspended ∼20 times, pelleted at 300x g for 5 min at 4 °C, and washed 2–3× with ice-cold FACS buffer. Following the last washing step, cells were resuspended in 1–2 ml FACS buffer and passed through the 40 µm cell strainer cap of the FACS tubes. For single-cell RNA sequencing, isolated single cells were resuspended in a compatible FACS buffer, containing 1× PBS, 1% FCS, 0.1 mM EDTA.

4.13 FACS

Single-cell suspensions, prepared as described above, were sorted using a FACS-Aria III (BD Bioscience). Cell were gated according to their FSC-A (front scatter area) and SSC-A (side scatter area) values, and doublets were excluded based on the FSC-W (forward scatter width) and FSC-H (forward scatter height). 7-AAD (eBioscience) was used to exclude dead cells. FVF populations were discriminated by fluorescence emission excited using a 488 nm laser. For microarray and qPCR analysis, cells were directly sorted in QIAzol lysis reagent (Qiagen) using the 85 µm nozzle. For single-cell transcriptomics, cells were sorted in FACS buffer, containing 1× PBS, 1% FCS, 0.1 mM EDTA using the 100 µm nozzle. For ChIP-seq and ATAC-seq experiments, cells were sorted in regular FACS buffer using the 85 µm nozzle. Data analysis was performed using FlowJo and FACS Diva software.

4.14 Microarray

For Affymetrix microarray experiments, total RNA was extracted using the miRNeasy Mini kit (Qiagen). Subsequently, RNA integrity was assessed using the Agilent 2100 Bioanalyzer in conjunction with the RNA 6000 Pico Kit (Agilent). Total RNA was then reverse transcribed and amplified using the Ovation PicoSL WTA System V2 (Nugen) in combination with the Encore Biotin Module (Nugen). Amplified cDNA was hybridized on either GeneChip Human Gene 2.0 ST arrays or GeneChip Mouse Gene 1.0 ST arrays. Expression console (version 1.3.0.187, Affymetrix) was then used for quality control. All subsequent computational analyses were performed in R using Bioconductor packages. Raw expression values were RMA normalized using the oligo package (version 1.38.0) and probe sets were annotated either using the hugene20sttranscriptcluster.db package (version 8.5.0) or the mogene10sttranscriptcluster.db package (version 8.5.0).

After this step, analyses were performed differently for the individual projects, as described in the following subsections.

4.14.1 Analysis of XM001 iPSC & PP Microarrays

Differential expression was analyzed using the limma package (version 3.30.7) and p-values were adjusted for multiple testing by Benjamini-Hochberg correction. Genes were considered differentially expressed if the adjusted p-value (FDR) was below a threshold of 0.05 and the fold-change was greater than or equal to 2. Functional enrichment analyses were performed using HOMER (version 4.9) [409] and gene set enrichment analysis was performed using GSEA 3.0 [410, 411] with genes ranked by $\log_2$ ratio between XM001 PPs and XM001 iPSCs.

4.14.2 Joint Analysis of XM001 & *PDX1*$^{P33T/+}$ iPSC & PP Microarrays

In order to compare *PDX1*$^{P33T/+}$PPs to WT XM001 PP, XM001 microarray data was reanalyzed together with the *PDX1*$^{P33T/+}$data. Differential expression analysis was performed on a prefiltered dataset, containing the 30562 probe sets with the highest mean expression values across all samples, using the limma package (version 3.30.7). P-values were adjusted using the Benjamini-Hochberg correction and genes were considered as differentially expressed if the adjusted p-value (FDR) was below a threshold of 0.1 and the fold-change was greater than or equal to 2. Functional enrichment analyses were performed using HOMER (version 4.10) [409] and functional annotations were based on literature research and the term affiliation provided by HOMER (version 4.10).

4.14.3 Analysis of FVF Microarrays

Differential expression analysis was performed on a prefiltered dataset, containing the 19200 probe sets with the highest mean expression values across all samples, using the limma package (version 3.30.7). P-values were adjusted using the Benjamini-Hochberg correction and genes were considered as differentially expressed, if the adjusted p-value (FDR) was below a threshold of 0.05 and the fold-change was greater than or equal to 2. Functional enrichment analyses were performed using HOMER (version 4.10) [409] or metascape (version 3.0) (http://metascape.org), and gene set enrichment analysis was performed using the fgsea package (version 1.8.0).

4.15 qPCR

4.15.1 XM001 iPSC & PPs

Total RNA was extracted using the miRNeasy Mini kit (Qiagen) and cDNA was synthesized using the High-Capacity cDNA Reverse Transcription kit (Applied Biosystems). TaqMan probe based qPCRs were performed using the TaqMan Fast Advanced Master Mix (Applied Biosystems) under default cycling conditions on a Viia7 real-time PCR cycler (Applied Biosystems). For

a list of TaqMan probes see 4.6.2. SYBR Green based qPCR was performed using the SYBR Green PCR Master Mix under default cycling conditions on a Viia7 real-time PCR cycler (Applied Biosystems). For a list of SYBR Green primers see 4.6.3. Relative expression was calculated using the delta-delta Ct method and normalized to the reference genes *RN18S* and *GAPDH* (TaqMan assays), or *ACTB* (SYBR Green assays).

4.15.2 FVF Crypt Cells

Total RNA was extracted using the miRNeasy Mini kit (Qiagen) and cDNA was synthesized using the SuperScript VILO cDNA Synthesis kit (Invitrogen). qPCRs were performed using TaqMan probes and the TaqMan Fast Advanced Master Mix (Applied Biosystems) under default cycling conditions on a Viia7 real-time PCR cycler (Applied Biosystems). For a list of TaqMan probes see 4.6.2. Relative expression was calculated using the qBase method [412] and normalized to reference genes.

4.16 RNA-seq

Total RNA was extracted from *PDX1*$^{+/-}$, *PDX1*$^{P33T/P33T}$, *PDX1*$^{C18R/C18R}$and XM001 lines were collected at the PP1 stage and the iPSC stage (XM001 cell only) using the miRNeasy Mini kit (Qiagen). RNA integrity was assessed using Agilent 2100 Bioanalyzer together with the Agilent RNA 6000 Pico Kit. Libraries were prepared using the TruSeq Stranded mRNA Library Prep (Illumina). Libraries were sequenced on an Illumina HiSeq 4000 (150 bp, paired end). RNA-seq data was quantified using salmon (version 0.11.3) `quant` with options `–validateMappings –rangeFactorizationBins 4 –numBootstraps 100 –seqBias –gcBias` [413].Salmon quantification results were imported into R using the tximport package (version 1.10.1) and genes with low read count (< 6 reads) were discarded prior to differential expression analysis. DESeq2 (version 1.20) [333] was used for differential expression analysis with default parameters and results were calculated with independent filtering and independent hypothesis weighting [414]. For the comparison of PPs against iPSCs, all datasets were used and all PP samples (i.e. XM001, *PDX1*$^{C18R/C18R}$, *PDX1*$^{P33T/P33T}$and *PDX1*$^{+/-}$) were compared with XM001 iPSCs using the `lfcThreshold=1.5` option in the `results` function. For the comparison of the PPs derived from the different cell lines, only data from PPs were loaded into DESeq2 and the results for these comparisons were generated without the `lfcThreshold` option. For all analyses, fold changes were shrunken using the apeglm method [415] using the results from independent hypothesis weighting. Genes were considered as differentially expressed adjusted p-value (FDR) was > 0.05 (PP vs iPSC) or 0.1 (PDX1 mutations vs control) and the absolute shrunken $\log_2$ fold change was > 2 (PP vs iPSC) or > 1 (PDX1 mutations vs control). GO-term and pathway analysis were performed using HOMER (version 4.10).

4.17 ChIP-seq

4.17.1 Sample Preparation for iPSC Derived Pancreatic Progenitors

XM001 and *PDX1*$^{P33T/+}$PP cells (2×10^6 cells) were cross-linked in 1% formaldehyde in cell culture media for 10 min at RT. The cross-linking reaction was terminated by the addition of glycine to a final concentration of 125 mM. Cell were washed with PBS containing 10% FCS, snap frozen and stored at -80°C. For chromatin fragmentation, cells were thawed and resuspended in Lysis-Buffer and sonicated for 20 min in a Covaris S220 sonicator, using a duty cycle of 2%, a peak incident power of 105 W and 200 cycles per burst. The fragmented chromatin was diluted 1:5 in IP-Buffer and directly used for immunoprecipitation. For PDX1 ChIP, 60% of the chromatin (equivalent of 1.2×10^6 cells) was processed in two parallel ChIPs using Dynabeads Protein G magnetic beads, preloaded with 3 µl goat anti PDX1 antibody (kindly provided by C. Wright). For H3K27ac ChIP, 40% of the chromatin (equivalent 0.4×10^6 cells) was processed in three parallel ChIPs using Dynabeads magnetic beads, preloaded with 3 µg anti H3K27ac antibody (Diagenode, C15410174). Dynabeads were incubated with the antibody for 2–3 h at 4°C on a lab rotator at ∼35 rpm. ChIP was then performed as described in 4.17.4.

4.17.2 Sample Preparation for Histone ChIP-seq in FVF$^+$ Cells

After dissociation of crypt cells into single cells, cells were cross-linked in 1% formaldehyde in Advanced DMEM/F12 supplemented with 10% FCS for 10 min at RT. The cross-linking reaction was stopped by the addition of glycine to a final concentration of 125 mM. Cells were then washed with PBS and used for FACS as described in 4.13 but without 7AAD staining. After FACS, cells were washed with PBS with 10% FCS, snap frozen and stored at -80°C. For chromatin fragmentation, $\sim1\times10^6$ cells were pooled from several aliquots originating from multiple mice and resuspended in Lysis-Buffer. Samples were sonicated for 20 min in a Covaris S220 sonicator, using a duty cycle of 2%, a peak incident power of 105 W and 200 cycles per burst. The fragmented chromatin was diluted 1:4 in IP-Buffer and directly used for immunoprecipitation. For each histone modification ChIP experiment, chromatin equivalent to $\sim2\times10^5$ cells was processed in 2 parallel ChIPs using Dynabeads magnetic beads , preloaded with 3 µg anti H3K27ac antibody (Diagenode, C15410174) or 3 µg anti H3K4me3 antibody (Diagenode, C15410003-50). Dynabeads were incubated with the antibody for 2–3 h at 4°C on a lab rotator at ∼35 rpm. ChIP was then performed as described in 4.17.4.

4.17.3 Sample Preparation for Foxa2 ChIP-seq in FVF$^+$ Cells

After FACS (4.13), cells were collected by centrifugation and cross-linked in 1% formaldehyde in PBS for 10 min at RT. The cross-linking reaction was stopped by the addition of Glycine to a final concentration of 125 mM. Subsequently cells were washed with PBS supplemented with 10% FCS, snap frozen and stored at -80°C. For chromatin fragmentation, $\sim5\times10^6$ cells were pooled from several aliquots originating from multiple mice and resuspended in Lysis-Buffer. Samples were sonicated for 40 min in a Covaris S220 sonicator, using a duty cycle of 2%, a peak incident

power of 105 W and 200 cycles per burst. The fragmented chromatin was diluted 1:4 in IP-Buffer and directly used for immunoprecipitation. Each ChIP experiment was processed in 4 parallel ChIPs using Dynabeads magnetic beads , preloaded with 3 µl anti GFP antibody (Abcam, ab290). Dynabeads were incubated with the antibody for 2–3 h at 4°C on a lab rotator at ∼35 rpm. ChIP was then performed as described in 4.17.4.

4.17.4 General ChIP-seq Procedure

For all ChIPs, antibody incubation was performed at 4°C for 5 h on a lab rotator at ∼35 rpm. Beads were then washed 5 times for 10 minutes at 4°C on a lab rotator at ∼35 rpm, using (1.) IP-Buffer, (2.) Washing-Buffer 1, (3.) Washing-Buffer 2 and (4. & 5.) Washing-Buffer 3. Subsequently, protein-DNA complexes were eluted from the beads in Elution-Buffer at 65°C for 20 min. Cross-links were reversed at 65°C overnight and DNA was purified by ethanol precipitation for library construction using the MicroPlex kit (Diagenode, C05010010). Libraries were sequenced on an Illumina HiSeq 1500 (50 bp, single end).

4.18 ChIP-seq Analysis

4.18.1 iPSC Derived Pancreatic Progenitors

Preprocessing & Alignment To remove low quality bases and potential adapter contamination, raw reads from PDX1 and H3K27ac ChIP-seq were processed with Trimmomatic (version 0.35). Reads were subsequently aligned to the hg19 genome using bowtie2 (version 2.2.6) using the `–very–sensitive` option and duplicate reads were removed using samtools (version 1.3). BAM files for PDX1 and H3K27ac from *PDX1*$^{P33T/+}$cells were subsampled to match the read depth of the BAM files from XM001 cells using samtools (version 1.3).

Peak Calling For the analysis of PDX1 in XM001 PPs, peaks were called using GEM (version 3.2) [416] using default settings and filtered, after visual inspection, using a q-value cut-off of 10^{-4}. Then, overlapping regions were merged. For the joint analysis of PDX1 in XM001 and *PDX1*$^{P33T/+}$PPs, PDX1 binding sites were called using GEM in multi condition mode, filtered, after visual inspection, using a q-value cut-offs of 10^{-4} (XM001) and $10^{-4.05}$ (*PDX1*$^{P33T/+}$) and overlapping regions were subsequently merged. H3K27ac enriched regions were called using BCP-HM [417] with default parameters. For the joint analysis of H3K27ac in XM001 and *PDX1*$^{P33T/+}$cells, reads from both samples were merged and enriched regions were called using HOMER (4.10) [409] with the options `–style histone –size 500 –minDist 2000`. All PDX1 peaks and H3K27ac enriched regions were further filtered by removing blacklist regions [418].

Functional Enrichement & Motif Analysis To map regions and binding sites to putative target genes, peaks within 20 kb of a TSS or within a gene body were annotated with the respective

gene using bedtools (version 2.26.0). Motif analysis, annotations with genomic features and pathway enrichment analysis were performed using HOMER.

Clustering To cluster H3K27ac profiles, the reads, falling in the intersection of enriched regions from all samples to be compared were counted and normalization factors were calculated using DESeq2 (version 1.20) [333]. After that, all regions from all samples to be compared were merged and reads falling into these union regions were counted and normalized with DESeq2 using the normalization factors calculated before. Normalized read counts were then used to calculate Pearson correlation coefficients that were subsequently subjected to hierarchical clustering.

Normalization of H3K27ac H3K27ac read counts from XM001 and *PDX1*$^{P33T/+}$cells were normalized using a set of promoters from housekeeping genes [419]. Housekeeping genes were derived from the microarray data (4.14) and defined by an absolute $\log_2$ fold change (*PDX1*$^{P33T/+}$vs XM001) < $\log_2(1.1)$ and a linear coefficient of variation < 0.15. H3K27ac reads were counted in a regions from 1000 bp upstream to 100 bp downstream of the housekeeping genes' TSSs and DESeq2 (version 1.20) was used to estimate size factors for normalization. Normalized read counts where used when indicated in the figures.

H3K27ac Enrichment To calculate enrichment of H3K27ac at PDX1-bound sites, three sets of random sites were generated by shuffling PDX1 binding sites from either XM001 cells alone or from the joint analysis of XM001 and *PDX1*$^{P33T/+}$cells using bedtools (version 2.26.0) shuffle. Next, H3K27ac reads were counted and ratio of the mean read count at PDX1 sites over the mean read count at shuffled sites was calculated. P-vales were calculated using the Wilcoxon rank sum test with continuity correction.

High Confidence Binding Sites High confidence binding sites from PPs and adult islets were generated from the overlap of available PDX1 binding profiles in progenitor cells or islets, where sites found in two out of three progenitor data sets or three out of four islet data sets were considered as high confidence binding sites.

Visualization BIGWIG files, normalized to 1×sequencing depth (Reads Per Genomic Content) using deepTools bamCoverage (3.1.0), were used for visualization of ChIP-seq data. Tracks from H3K27ac and PDX1 in adult islets were generated from merged BAM files containing reads from all available experiments. For visualization of jointly analyzed H3K27ac data, downsampled BAM files were used to gnerate BIGWIG files, which were also normalized using the size factors previously determined by DESeq2.

Public Data For publicly available datasets, raw reads were downloaded from online repositories and processed as described above (GSE54471 [302], GSE58686 [311], E-MTAB-1143 [312], E-MTAB-1919 [313]).

4.18.2 FVF^{+} Intestinal Crypt Cells

Preprocessing & Alignment To remove low quality bases and potential adapter contamination, raw reads from Foxa2, H3K4me3 and H3K27ac ChIP-seq experiments were processed with Trimmomatic (version 0.35). Reads were subsequently aligned to the mm10 genome using bowtie2 (version 2.2.6) using the `–very-sensitive` option and duplicate reads were removed using samtools (version 1.3).

Custom Blacklist In order to remove false positive peaks, a custom blacklist of region with strong read accumulation in input samples was generated. To that end, BAM files from input samples for histone and GFP ChIP-seq were merged and peaks were called using MACS2 with the parameters `–nomodel –nolambda –keep-dup all –broad -q 1e-03`. Subsequently, all peaks closer than 2 kb were merged using bedtools (version 2.26.0). The resulting list of peaks was then combined with a publicly available blacklist for the mm10 genome obtained from `https://www.encodeproject.org/annotations/ENCSR636HFF/` using bedtools (version 2.26.0).

Peak Calling For Foxa2 ChIP-seq, peaks from FVF^{low} and FVF^{high} cells were called together using GEM (version 3.2) [416] in multi condition mode with default settings and filtered, after visual inspection, using a q-value cut-offs of $10^{-3.6}$ (FVF^{low}) and $10^{-5.6}$ (FVF^{high}). Overlapping regions were subsequently merged using bedtools (version 2.26.0). For the analysis of H3K27ac in FVF^{low} and FVF^{high} cells, reads from both samples were merged and enriched regions were called using HOMER (version 4.10) [409] with the options `-style histone -size 500 -minDist 2000`. All Foxa2 peaks and H3K27ac enriched regions were further filtered by removing blacklist regions (see 4.18.2).

Functional Enrichement & Motif Analysis To map regions and binding sites to putative target genes, peaks within 20 kb of a TSS or within a gene body were annotated with the respective gene using bedtools (version 2.26.0). Motif analysis, annotations with genomic features and pathway enrichment analysis were performed using HOMER (version 4.10). Analysis of the associations between motif occurrences and Foxa2 ChIP-seq were performed using chromVar [336] and 1:1 scaling was forced by setting a uniform read depth.

Normalization of H3K27ac & H3K4me3 H3K27ac and H3K4me3 read counts from FVF^{low} and FVF^{high} cells were normalized using a set of promoters from housekeeping genes [419]. Housekeeping genes were derived from the microarray data (4.14) and defined by an absolute $\log_2$

fold change (FVF^{high} vs FVF^{low}) < $\log_2(1.1)$ and a linear coefficient of variation < 0.15. H3K27ac and H3K4me3 reads were counted in a regions from 1000 bp upstream to 100 bp downstream of the housekeeping genes' TSSs and DESeq2 (version 1.20) was used to estimate size factors for normalization. Normalized read counts where used when indicated in the figures.

Enrichment of Active Chromatin Marks To calculate enrichment of H3K27ac H3K4me3 and ATAC-seq at Foxa2-bound sites, three sets of random sites were generated by shuffling the union of Foxa2 binding sites from the joint analysis of FVF^{low} and FVF^{high} cells using bedtools (version 2.26.0) shuffle. H3K27ac H3K4me3 and ATAC-seq reads were then counted and the ratio of the mean read count at Foxa2-bound sites over the mean read count at shuffled sites was calculated. P-vales were calculated using the Wilcoxon rank sum test with continuity correction.

Visualization For visualization of ChIP-seq data, BIGWIG files generated from the corresponding BAM files using deepTools bamCoverage (3.1.0). Reads were extended to the mean fragment length, determined by HOMER. H3K27ac and H3K4me3 tracks were scaled using the size factors previously determined by DESeq2.

Public Data or publicly available datasets, raw reads were downloaded from online repositories and processed as described above (GSE51464 [222], GSE34568 [345]).

4.19 ATAC-seq

ATAC-seq was performed in four replicates following the original ATAC-seq protocol developed by Buenrostro et al. 2013 [331] with minor modifications during cell preparation.

Cell Preparation For each ATAC-seq experiment, 60 000 FVF^{low} and FVF^{high} cells (FACS events) were sorted by FACS and immediately used for ATAC reactions. Cells were collected by centrifugation at 900×g for 5 min and 4 °C in a fixed angle rotor to ensure minimal cell loss during aspiration of supernatants. Cells were washed once with 50 µl cold PBS, resuspended in 50 µl cold Lysis-Buffer and immediately pelleted at 900×g for 10 min and 4 °C to proceed with the transposition reaction.

Transposition & Purification Cell pellet was kept on ice while the transposition reaction mix was prepared, combining the following:

25 µl	2× TD Buffer
2.5 µl	Tn5 Transposase
22.5 µl	Nuclease free H_2O
50 µl	Total

Cells were then gently resuspended in the transposition reaction mix and incubated for transposition at 37 °C for 30 min. Immediately after the transposition reaction, DNA was purified from samples using the Qiagen MinElute PCR Purification Kit. Transposed DNA was eluted in 10 µl Elution-Buffer and DNA concentration was determined using the Qubit dsDNA HS assay. Purified DNA was stored at -20°C.

PCR Amplification DNA fragments were then amplified by PCR, using the following reagent mix and cycling conditions. A complete list of adapter sequences and indices can be found in 4.6.1.

10 µl	Transposed DNA
10 µl	Nuclease free H_2O
2.5 µl	25 µM Sequencing adapter Ad1
2.5 µl	25 µM Sequencing adapter Ad2.x (Index)
25 µl	NEBNext High-Fidelity 2× PCR Master Mix
50 µl	Total

Cycling conditions:

72°C	5 min	
98°C	30 sec	
98°C	10 sec	
63°C	30 sec	4×
72°C	1 min	
4°C	hold	

After this pre-amplification step, the PCR reaction was monitored by qPCR to avoid over-amplification of the transposed DNA fragments. To this end, a pPCR side reaction using the following reagent mix and cycling conditions was performed.

5 µl	Pre-amplified DNA from previous step
2.25 µl	Nuclease free H_2O
0.75 µl	20× EvaGreen dye
1 µl	6.25 µM Sequencing adapter Ad1
1 µl	6.25 µM Sequencing adapter Ad2.x (Index)
5 µl	NEBNext High-Fidelity 2× PCR Master Mix
15 µl	Total

Cycling conditions:

98°C	30 sec	
98°C	10 sec	
63°C	30 sec	19×
72°C	1 min	
4°C	hold	

To calculate the number of additional cycles, required for the amplification of the remaining 45 µl of PCR reaction, the qPCR cycle number, corresponding to ¼ of the maximum florescence intensity, was determined ($C_{t\ 1/4}$). The main PCR reaction was then continued by running the appropriate number of additional cycles.

Cycling conditions:

98°C	30 sec	
98°C	10 sec	
63°C	30 sec	$C_{t\ 1/4}\times$
72°C	1 min	
4°C	hold	

After the amplification step, the library was purified using the QIAquick PCR Purification Kit and eluted in 20 µl Elution-Buffer. After purification, 1 µl was set aside for Bioanalyzer analysis.

SPRI Size Selection To enrich for fragments, from 100–600 bp double sided SPRI based size selection (0.6×/1.3×) was performed using AMPure XP beads according to the SPRIselect user guide. Purified and size selected libraries were eluted in 17 µl Elution-Buffer and stored at -20°C. Quality and concentrations of the libraries before and after size selection were assessed on the 2100 Bioanalyzer platform using the Agilent High Sensitivity DNA Kit. Libraries were sequenced on an Illumina HiSeq 1500 (50 bp, paired end).

4.20 ATAC-seq Analysis

Preprocessing & Alignment To remove low quality bases and potential adapter contamination, raw reads from ATAC-seq experiments were processed with Trimmomatic (version 0.35). Reads were subsequently aligned to the mm10 genome using bowtie2 (version 2.2.6) with the `–very-sensitive` option. Duplicate reads were removed using samtools (version 1.3) and reads were shifted +4/-5 bp to account for Tn5 cut properties [331].

Peak Calling ATAC-seq peaks were called on nucleosome free fragments (fragment length < 150 bp) of merged BAM files from all replicates, combined from FVF^{low} and FVF^{high} datasets, using MACS2 with the parameters `–nomodel –nolambda –keep-dup all –call-summits`. Peaks overlapping blacklist regions for the mm10 genome obtained from `https://www.encodeproject.org/annotations/ENCSR636HFF/` were removed using bedtools (version 2.26.0).

Functional Enrichement & Motif Analysis To map regions of open chromatin to genes, peaks within 20 kb of a TSS or within a gene body were annotated with the respective gene using bedtools (version 2.26.0). Annotations with genomic features were performed using HOMER

(version 4.10). Analysis of the associations between motif occurrences and ATAC-seq variation were performed using chromVar [336]. Functional enrichment analyses were performed using HOMER (version 4.10) [409] or metascape (version 3.0) (http://metascape.org).

Normalization of ATAC-seq Data ATAC-seq data from FVF^{low} and FVF^{high} cells were normalized using a set of promoters from housekeeping genes [419]. Housekeeping genes were derived from the microarray data (4.14) and defined by an absolute $\log_2$ fold change (FVF^{high} vs FVF^{low}) < $\log_2(1.1)$ and a linear coefficient of variation < 0.15. ATAC-seq reads were counted in the regions from 1000 bp upstream to 100 bp downstream of the housekeeping genes' TSSs and DESeq2 (version 1.20) was used to estimate size factors for normalization. Normalized read counts where used for differential enrichment analysis, and when indicated in the figures.

Differential Enrichment Analysis For differential enrichment analysis, ATAC-seq reads were counted in the identified peak regions and normalized using the size factors determined using the housekeeping genes, as described above. Regions with low read coverage (< 5 reads) were removed and remaining regions were processed with DESeq2 (version 1.20) [333] and fold changes were shrunken using the apeglm method [415] with the `lfcThreshold=0.4` option in the `results` function. Regions were considered as differentially enriched when s-value (local false sign rate) [420] was > 0.1 and the absolute shrunken $\log_2$ fold change was > 0.5.

Visualization For visualization of ATAC-seq data, BIGWIG files generated from each BAM file using deepTools bamCoverage (3.1.0). Reads were extended to the actual fragment length, determined by the read pairs, and tracks were scaled using the size factors previously determined by DESeq2. Subsequently, replicates from FVF^{low} and FVF^{high} were merged and the mean signal was calculated to obtain a single track for FVF^{low} and FVF^{high}, respectively.

4.21 Single-Cell RNA-seq

For scRNA-seq, we made use of the FVF reporter, and enriched FVF^{+} cells in order to increase the fraction of secretory cell types. To that end, crypts were isolated and single cell suspension were generated as described in 4.12. Single cells were then purified by FACS as described in 4.13. To obtain a FVF-enriched crypt cell sample, we combined 30,000 FACS purified live FVF^{+} cells, sorted by their Venus fluorescence and consisting of both, FVF^{low} and FVF^{high} cells, with 30,000 FACS purified life cells, not sorted by their Venus fluorescence and thus consisting of FVF^{neg}, FVF^{low} and FVF^{high} cells. Sorted cells were counted to determine the exact cell number and the fraction of dead cells was estimated by trypan blue staining. Samples were used for scRNA-seq if the fraction of dead cells was below 20%. The final cell number was adjusted to recover a target cell number of 10.000 cells per sample. scRNA-seq libraries were then generated in 3 replicates using the 10x Chromium Single Cell 3' Library & Gel Bead Kit v2 (10× Genomics) following the manufacturer's protocol. Libraries were sequenced on an Illumina HiSeq 4000 (150 bp, paired end).

4.22 scRNA-seq Analysis

Primary Analysis & Preprocessing For primary data analysis, the CellRanger pipeline (version 2.0.0) provided by 10× Genomics was used for demultiplexing of raw base call (BCL) files, alignment to the mm10 genome, read filtering, as well as barcode and UMI counting. All subsequent analysis steps were performed in python using the SCANPY API [421]. First, cells and genes were filtered, in order to retain only genes with an UMI count > 1 and only cells expression > 1000 genes. Subsequently, all cells with a large fraction of counts from mitochondrial genes (> 10%) were excluded. Count matrices from individual replicates were merged and expression values were log-transformed (log(count+1)). We then further filtered the dataset and excluded genes, expressed in <20 cells. In the next step, batch correction was performed using a python implementation of ComBat [422] with default parameters and defining each sample as one batch. To assess contamination from non-epithelial cells, an initial round of graph-based clustering (louvain) was performed and a clearly distinct cluster of cells, expression high levels of immune cell marker genes, was identified and excluded from further analysis.

Clustering & Cell Type Annotation The top 2000 variable genes were determined based on the normalized dispersion of genes, using `pp.filter_genes_dispersion` with default parameters. To compute a neighborhood graph of the individual cells, `pp.neighbors` function was used with the parameters `n_neighbors=25,n_pcs=50,method='gauss'` and a Uniform Manifold Approximation and Projection (UMAP) was calculated for visualization [423]. To identify clusters of cell types, luvain clustering was performed using the `tl.louvain` function in multiple iterations with varying resolutions in a „split-and-merge“ approach. The final clusters of cell types and subtypes were annotated based on the expression of known marker genes. Novel marker genes for each populations were identified using the `tl.rank_genes_groups` function in an pairwise approach comparing each cluster individually against every other cluster, retaining only genes identified in each comparison (score > 8). In order to put the cells in a pseudotemporal order, reflecting the differentiation progress, `tl.dpt` function was used to calculate the diffusion pseudotime, rooted to a randomly selected cycling cells (> 1 count for *Mki67* and *Pcna*) from the ISC cluster.

Denoising To get a finer resolution of the expression values of lowly expressed genes, expression data was denoised by imputation of missing values using the Deep Count Autoencoder (DCA) package [356]. The *dca* function was used with the parameters `ae_type='nb', hidden_size=[1024,512,1024], scale=False, batchnorm=False` on a subset containing all identified marker genes for every cell population and all putative Foxa2 target genes. Genes with expression patterns similar to *Foxa2* were identified by calculating the Pearson product-moment correlation coefficients from the denoised data for *Foxa2*–gene pair across all cells.

References

[1] International Diabetes Federation: *IDF Diabetes Atlas, 8th edn.* Technical Report, International Diabetes Federation, Brussels, Belgium, 2017.

[2] Roglic, Gojka and World Health Organization: *Global report on diabetes*. 2016.

[3] You, Wen-Peng and Maciej Henneberg: *Type 1 diabetes prevalence increasing globally and regionally: the role of natural selection and life expectancy at birth*. BMJ Open Diabetes Research & Care, 4(1):e000161, mar 2016.

[4] Daneman, Denis: *Type 1 diabetes*. The Lancet, 367(9513):847–858, mar 2006.

[5] Maahs, David M., Nancy A. West, Jean M. Lawrence and Elizabeth J. Mayer-Davis: *Epidemiology of Type 1 Diabetes*. Endocrinology and Metabolism Clinics of North America, 39(3):481–497, sep 2010.

[6] Evans, J. M. M., R. W. Newton, D. A. Ruta, T. M. MacDonald and A. D. Morris: *Socio-economic status, obesity and prevalence of Type 1 and Type 2 diabetes mellitus*. Diabetic Medicine, 17(6):478–480, jun 2000.

[7] Bruno, G., C. Runzo, P. Cavallo-Perin, F. Merletti, M. Rivetti, S. Pinach, G. Novelli, M. Trovati, F. Cerutti and G. Pagano: *Incidence of Type 1 and Type 2 Diabetes in Adults Aged 30-49 Years: The population-based registry in the province of Turin, Italy*. Diabetes Care, 28(11):2613–2619, nov 2005.

[8] Holman, N., B. Young and R. Gadsby: *Current prevalence of Type 1 and Type 2 diabetes in adults and children in the UK*. Diabetic Medicine, 32(9):1119–1120, sep 2015.

[9] Weir, Gordon C and Susan Bonner-weir: *Progression to Diabetes*. Diabetes, 53(December):16–21, 2004.

[10] Prentki, M.: *Islet cell failure in type 2 diabetes*. Journal of Clinical Investigation, 116(7):1802–1812, jul 2006.

[11] Mozaffarian, Dariush: *Dietary and Policy Priorities for Cardiovascular Disease, Diabetes, and Obesity*. Circulation, 133(2):187–225, jan 2016.

[12] Elks, C. E., K. K. Ong, R. A. Scott, Y. T. van der Schouw, J. S. Brand, P. A. Wark, P. Amiano, B. Balkau, A. Barricarte, H. Boeing, A. Fonseca-Nunes, P. W. Franks, S. Grioni, J. Halkjaer, R. Kaaks, T. J. Key, K. T. Khaw, A. Mattiello, P. M. Nilsson, K. Overvad, D. Palli, J. R. Quiros, S. Rinaldi, O. Rolandsson, I. Romieu, C. Sacerdote, M.-J. Sanchez, A. M. W. Spijkerman, A. Tjonneland, M.-J. Tormo, R. Tumino, D. L. van der A, N. G. Forouhi, S. J. Sharp, C. Langenberg, E. Riboli and N. J. Wareham: *Age at Menarche and Type 2 Diabetes Risk: The EPIC-InterAct study*. Diabetes Care, 36(11):3526–3534, nov 2013.

[13] McDonald, Tim J. and Sian Ellard: *Maturity onset diabetes of the young: Identification and diagnosis*. Annals of Clinical Biochemistry, 50(5):403–415, 2013.

[14] Ellard, S., C. Bellanné-Chantelot, A. T. Hattersley, C. Carette, L. Castano Gonzalez, G. De Nanclares Leal, R. Elles, G. Gaspar, D. Gasperikova, T. Hansen, M. Herr, O. Kamarainen, C. Kannengiesser, I. Klimes, G. Lacape, M. Losekoot, M. Malecki, P. Meyer, P. Njolstad, T. Predragovic, S. Pruhova and W. Wuyts: *Best practice guidelines for the molecular genetic diagnosis of maturity-onset diabetes of the young*. Diabetologia, 51(4):546–553, 2008.

[15] Stoffers, D a, N T Zinkin, V Stanojevic, W L Clarke and J F Habener: *Pancreatic agenesis attributable to a single nucleotide deletion in the human IPF1 gene coding sequence.* Nature genetics, 15(1):106–110, 1997.

[16] Stoffers, D A, Jorge Ferrer, William L. Clarke and Joel F. Habener: *Early-onset type-II diabetes mellitus (MODY4) linked to IPF1*. Nature genetics, 17(2):138–9, oct 1997.

[17] Shih, Hung Ping, Allen Wang and Maike Sander: *Pancreas Organogenesis: From Lineage Determination to Morphogenesis*. Annual Review of Cell and Developmental Biology, 29(1):81–105, 2013.

[18] Pallagi, Petra, Péter Hegyi and Zoltán Rakonczay: *The Physiology and Pathophysiology of Pancreatic Ductal Secretion*. Pancreas, 44(8):1211–1233, nov 2015.

[19] Hauge-Evans, A. C., A. J. King, D. Carmignac, C. C. Richardson, I. C.A.F. Robinson, M. J. Low, M. R. Christie, S. J. Persaud and P. M. Jones: *Somatostatin Secreted by Islet δ-Cells Fulfills Multiple Roles as a Paracrine Regulator of Islet Function*. Diabetes, 58(2):403–411, feb 2009.

[20] Kojima, Shinya, Naohiko Ueno, Akihiro Asakawa, Kenichirou Sagiyama, Tetsuro Naruo, Shigeto Mizuno and Akio Inui: *A role for pancreatic polypeptide in feeding and body weight regulation*. Peptides, 28(2):459–463, feb 2007.

[21] Khandekar, Neeta, Britt A. Berning, Amanda Sainsbury and Shu Lin: *The role of pancreatic polypeptide in the regulation of energy homeostasis*. Molecular and Cellular Endocrinology, 418:33–41, dec 2015.

[22] Tan, Tricia M. and Stephen R. Bloom: *Pancreatic Polypeptide*. In *Handbook of Biologically Active Peptides*, pages 1294–1299. Elsevier, 2013.

[23] Assmann, Anke, Charlotte Hinault and Rohit N Kulkarni: *Growth factor control of pancreatic islet regeneration and function*. Pediatric Diabetes, 10(1):14–32, feb 2009.

[24] Peeters, Theo L.: *Ghrelin*. In *Handbook of Biologically Active Peptides*, pages 1236–1240. Elsevier, 2013.

[25] Burger, Kyle S. and Laura A. Berner: *A functional neuroimaging review of obesity, appetitive hormones and ingestive behavior*. Physiology & Behavior, 136:121–127, sep 2014.

[26] Yada, T., B. Damdindorj, R. S. Rita, T. Kurashina, A. Ando, M. Taguchi, M. Koizumi, H. Sone, M. Nakata, M. Kakei and K. Dezaki: *Ghrelin signalling in β -cells regulates insulin secretion and blood glucose*. Diabetes, Obesity and Metabolism, 16(S1):111–117, sep 2014.

[27] Röder, Pia V, Bingbing Wu, Yixian Liu and Weiping Han: *Pancreatic regulation of glucose homeostasis*. Experimental & Molecular Medicine, 48(3):e219, mar 2016.

[28] Slack, J M: *Developmental biology of the pancreas.* Development (Cambridge, England), 121(6):1569–80, jun 1995.

[29] Gittes, George K: *Developmental biology of the pancreas: a comprehensive review.* Developmental biology, 326(1):4–35, feb 2009.

[30] Oliver-Krasinski, Jennifer M and Doris A Stoffers: *On the origin of the beta cell.* Genes & Development, 22(15):1998–2021, aug 2008.

[31] Wessells, Norman K. and Julia H. Cohen: *Early pancreas organogenesis: Morphogenesis, tissue interactions, and mass effects*. Developmental Biology, 15(3):237–270, mar 1967.

[32] Golosow, Nikolas and Clifford Grobstein: *Epitheliomesenchymal interaction in pancreatic morphogenesis*. Developmental Biology, 4(2):242–255, apr 1962.

[33] Lammert, E.: *Induction of Pancreatic Differentiation by Signals from Blood Vessels*. Science, 294(5542):564–567, oct 2001.

[34] Pan, Fong Cheng and Chris Wright: *Pancreas organogenesis: From bud to plexus to gland*. Developmental Dynamics, 240(3):530–565, mar 2011.

[35] Bastidas-Ponce, Aimée, Sara S. Roscioni, Ingo Burtscher, Erik Bader, Michael Sterr, Mostafa Bakhti and Heiko Lickert: *Foxa2 and Pdx1 cooperatively regulate postnatal maturation of pancreatic β-cells*. Molecular Metabolism, 6(6):524–534, 2017.

[36] Kesavan, Gokul, Fredrik Wolfhagen Sand, Thomas Uwe Greiner, Jenny Kristina Johansson, Sune Kobberup, Xunwei Wu, Cord Brakebusch and Henrik Semb: *Cdc42-Mediated Tubulogenesis Controls Cell Specification*. Cell, 139(4):791–801, nov 2009.

[37] Villasenor, A., D. C. Chong, M. Henkemeyer and O. Cleaver: *Epithelial dynamics of pancreatic branching morphogenesis*. Development, 137(24):4295–4305, dec 2010.

[38] Zhou, Qiao, Anica C. Law, Jayaraj Rajagopal, William J. Anderson, Paul A. Gray and Douglas A. Melton: *A Multipotent Progenitor Domain Guides Pancreatic Organogenesis*. Developmental Cell, 13(1):103–114, jul 2007.

[39] Reichert, Maximilian and Anil K. Rustgi: *Pancreatic ductal cells in development, regeneration, and neoplasia*. Journal of Clinical Investigation, 121(12):4572–4578, dec 2011.

[40] Schwitzgebel, V M, D W Scheel, J R Conners, J Kalamaras, J E Lee, D J Anderson, L Sussel, J D Johnson and M S German: *Expression of neurogenin3 reveals an islet cell precursor population in the pancreas*. Development, 127(16):3533 LP – 3542, aug 2000.

[41] Gu, Guoqiang, Jolanta Dubauskaite and Douglas A Melton: *Direct evidence for the pancreatic lineage: NGN3+ cells are islet progenitors and are distinct from duct progenitors.* Development (Cambridge, England), 129(10):2447–57, may 2002.

[42] Pictet, R. and W. J. Rutter: *Development of the embryonic endocrine pancreas*. Handbook of physiology, sec. 7, Vol. 1, 1972.

[43] Nikolova, Ganka, Normund Jabs, Irena Konstantinova, Anna Domogatskaya, Karl Tryggvason, Lydia Sorokin, Reinhard Fässler, Guoqiang Gu, Hans-Peter Gerber, Napoleone Ferrara, Douglas A. Melton and Eckhard Lammert: *The Vascular Basement Membrane: A Niche for Insulin Gene Expression and β Cell Proliferation*. Developmental Cell, 10(3):397–405, mar 2006.

[44] Reinert, R. B., Q. Cai, J.-Y. Hong, J. L. Plank, K. Aamodt, N. Prasad, R. Aramandla, C. Dai, S. E. Levy, A. Pozzi, P. A. Labosky, C. V. E. Wright, M. Brissova and A. C. Powers: *Vascular endothelial growth factor coordinates islet innervation via vascular scaffolding*. Development, 141(7):1480–1491, apr 2014.

[45] Rodriguez-Diaz, Rayner and Alejandro Caicedo: *Neural control of the endocrine pancreas*. Best Practice & Research Clinical Endocrinology & Metabolism, 28(5):745–756, oct 2014.

[46] Thorens, B.: *Neural regulation of pancreatic islet cell mass and function*. Diabetes, Obesity and Metabolism, 16(S1):87–95, sep 2014.

[47] Steiner, Donald J., Abraham Kim, Kevin Miller and Manami Hara: *Pancreatic islet plasticity: Interspecies comparison of islet architecture and composition*. Islets, 2(3):135–145, may 2010.

[48] Bliss, C. R. and G. W. Sharp: *Glucose-induced insulin release in islets of young rats: time-dependent potentiation and effects of 2-bromostearate*. American Journal of Physiology-Endocrinology and Metabolism, 263(5):E890–E896, nov 1992.

[49] Aguayo-Mazzucato, C., A. Koh, I. El Khattabi, W. C. Li, E. Toschi, A. Jermendy, K. Juhl, K. Mao, G. C. Weir, A. Sharma and S. Bonner-Weir: *Mafa expression enhances glucose-responsive insulin secretion in neonatal rat beta cells*. Diabetologia, 2011.

[50] Gu, Chunyan, Gretchen H. Stein, Ning Pan, Sandra Goebbels, Hanna Hörnberg, Klaus-Armin Nave, Pedro Herrera, Peter White, Klaus H. Kaestner, Lori Sussel and Jacqueline E. Lee: *Pancreatic β Cells Require NeuroD to Achieve and Maintain Functional Maturity*. Cell Metabolism, 11(4):298–310, apr 2010.

[51] Guo, Lili, Akari Inada, Cristina Aguayo-Mazzucato, Jennifer Hollister-Lock, Yoshio Fujitani, Gordon C. Weir, C. V. E. Wright, Arun Sharma and Susan Bonner-Weir: *PDX1 in Ducts Is Not Required for Postnatal Formation of -Cells but Is Necessary for Their Subsequent Maturation*. Diabetes, 62(10):3459–3468, oct 2013.

[52] Stolovich-Rain, Miri, Jonatan Enk, Jonas Vikesa, Finn Cilius Nielsen, Ann Saada, Benjamin Glaser and Yuval Dor: *Weaning Triggers a Maturation Step of Pancreatic β Cells*. Developmental Cell, 32(5):535–545, 2015.

[53] Burlison, Jared S., Qiaoming Long, Yoshio Fujitani, Christopher V.E. Wright and Mark A. Magnuson: *Pdx-1 and Ptf1a concurrently determine fate specification of pancreatic multipotent progenitor cells*. Developmental Biology, 2008.

[54] Guz, Y, M R Montminy, R Stein, J Leonard, L W Gamer, C V Wright and G Teitelman: *Expression of murine STF-1, a putative insulin gene transcription factor, in beta cells of pancreas, duodenal epithelium and pancreatic exocrine and endocrine progenitors during ontogeny.* Development (Cambridge, England), 121(1):11–18, 1995.

[55] Kawaguchi, Yoshiya, Bonnie Cooper, Maureen Gannon, Michael Ray, Raymond J. MacDonald and Christopher V.E. Wright: *The role of the transcriptional regulator Ptf1a in converting intestinal to pancreatic progenitors*. Nature Genetics, 32(1):128–134, sep 2002.

[56] Jonsson, J, L Carlsson, T Edlund and H Edlund: *Insulin-promoter-factor 1 is required for pancreas development in mice.*, 1994.

[57] Ahlgren, U, J Jonsson and H Edlund: *The morphogenesis of the pancreatic mesenchyme is uncoupled from that of the pancreatic epithelium in IPF1/PDX1-deficient mice.* Development (Cambridge, England), 122(5):1409–16, 1996.

[58] Offield, M F, T L Jetton, P A Labosky, M Ray, R W Stein, M A Magnuson, B L Hogan and C V Wright: *PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum.* Development (Cambridge, England), 122(3):983–95, 1996.

[59] Marty-Santos, L. and O. Cleaver: *Pdx1 regulates pancreas tubulogenesis and E-cadherin expression*. Development, 143(6):1056–1056, 2016.

[60] Gao, Nan, John LeLay, Marko Z Vatamaniuk, Sebastian Rieck, Joshua R Friedman and Klaus H Kaestner: *Dynamic regulation of Pdx1 enhancers by Foxa1 and Foxa2 is essential for pancreas development.* Genes & development, 22(24):3435–48, dec 2008.

[61] Haumaitre, C., E. Barbacci, M. Jenny, M. O. Ott, G. Gradwohl and S. Cereghini: *Lack of TCF2/vHNF1 in mice leads to pancreas agenesis*. Proceedings of the National Academy of Sciences, 102(5):1490–1495, feb 2005.

[62] Lynn, F. C., S. B. Smith, M. E. Wilson, K. Y. Yang, N. Nekrep and M. S. German: *Sox9 coordinates a transcriptional network in pancreatic progenitor cells*. Proceedings of the National Academy of Sciences, 104(25):10500–10505, jun 2007.

[63] Shih, Hung Ping, Philip A. Seymour, Nisha A. Patel, Ruiyu Xie, Allen Wang, Patrick P. Liu, Gene W. Yeo, Mark A. Magnuson and Maike Sander: *A Gene Regulatory Network Cooperatively Controlled by Pdx1 and Sox9 Governs Lineage Allocation of Foregut Progenitor Cells*. Cell Reports, 13(2):326–336, 2015.

[64] Stanger, Ben Z., Akemi J. Tanaka and Douglas A. Melton: *Organ size is limited by the number of embryonic progenitor cells in the pancreas but not the liver*. Nature, 445(7130):886–891, 2007.

[65] Hebrok, M., S. K. Kim and D. A. Melton: *Notochord repression of endodermal Sonic hedgehog permits pancreas development*. Genes & Development, 12(11):1705–1713, jun 1998.

[66] Seymour, P. A., H. P. Shih, N. A. Patel, K. K. Freude, R. Xie, C. J. Lim and M. Sander: *A Sox9/Fgf feed-forward loop maintains pancreatic organ identity*. Development, 139(18):3363–3372, sep 2012.

[67] Apelqvist, A, Hao Li, Lukas Sommer, Paul Beatus, David J. Anderson, Tasuku Honjo, M Hrabe de Angelis, Urban Lendahl and Helena Edlund: *Notch signalling controls pancreatic cell differentiation.* Nature, 400(6747):877–81, aug 1999.

[68] Jensen, Jan, Erna Engholm Pedersen, Philip Galante, Jacob Hald, R. Scott Heller, Makoto Ishibashi, Ryoichiro Kageyama, Francois Guillemot, Palle Serup and Ole D. Madsen: *Control of endodermal endocrine development by Hes-1.* Nature Genetics, 24(1):36–44, jan 2000.

[69] Seymour, P. A., K. K. Freude, M. N. Tran, E. E. Mayes, J. Jensen, R. Kist, G. Scherer and M. Sander: *SOX9 is required for maintenance of the pancreatic progenitor cell pool.* Proceedings of the National Academy of Sciences, 104(6):1865–1870, feb 2007.

[70] Georgia, Senta, Rosemary Soliz, Min Li, Pumin Zhang and Anil Bhushan: *p57 and Hes1 coordinate cell cycle exit with self-renewal of pancreatic progenitors.* Developmental Biology, 298(1):22–31, oct 2006.

[71] Fujikura, Junji, Kiminori Hosoda, Hiroshi Iwakura, Tsutomu Tomita, Michio Noguchi, Hiroaki Masuzaki, Kenji Tanigaki, Daisuke Yabe, Tasuku Honjo and Kazuwa Nakao: *Notch/Rbp-j signaling prevents premature endocrine and ductal cell differentiation in the pancreas.* Cell Metabolism, 3(1):59–65, jan 2006.

[72] Afelik, Solomon, Xiaoling Qu, Edy Hasrouni, Michael a Bukys, Tye Deering, Stephan Nieuwoudt, William Rogers, Raymond J Macdonald and Jan Jensen: *Notch-mediated patterning and cell fate allocation of pancreatic progenitor cells.* Development (Cambridge, England), 139(10):1744–53, may 2012.

[73] Esni, F.: *Notch inhibits Ptf1 function and acinar cell differentiation in developing mouse and zebrafish pancreas.* Development, 131(17):4213–4224, sep 2004.

[74] Schaffer, Ashleigh E., Kristine K. Freude, Shelley B. Nelson and Maike Sander: *Nkx6 Transcription Factors and Ptf1a Function as Antagonistic Lineage Determinants in Multipotent Pancreatic Progenitors.* Developmental Cell, 18(6):1022–1029, jun 2010.

[75] Solar, Myriam, Carina Cardalda, Isabelle Houbracken, Mercè Martín, Miguel Angel Maestro, Nele De Medts, Xiaobo Xu, Vanessa Grau, Harry Heimberg, Luc Bouwens and Jorge Ferrer: *Pancreatic Exocrine Duct Cells Give Rise to Insulin-Producing β Cells during Embryogenesis but Not after Birth.* Developmental Cell, 17(6):849–860, dec 2009.

[76] Klinck, Rasmus, Ernst-Martin Füchtbauer, Jonas Ahnfelt-Rønne, Palle Serup, Jan Nygaard Jensen and Mette Christine Jørgensen: *A BAC transgenic Hes1-EGFP reporter reveals novel expression domains in mouse embryos.* Gene Expression Patterns, 11(7):415–426, oct 2011.

[77] Kopp, J. L., C. L. Dubois, A. E. Schaffer, E. Hao, H. P. Shih, P. A. Seymour, J. Ma and M. Sander: *Sox9+ ductal cells are multipotent progenitors throughout development but do not produce new endocrine cells in the normal or injured adult pancreas.* Development, 138(4):653–665, feb 2011.

[78] Wang, Junfeng, Gamze Kilic, Muge Aydin, Zoe Burke, Guillermo Oliver and Beatriz Sosa-Pineda: *Prox1 activity controls pancreas morphogenesis and participates in the production of "secondary transition" pancreatic endocrine cells.* Developmental Biology, 286(1):182–194, oct 2005.

[79] Krapp, A, M Knöfler, S Frutiger, G J Hughes, O Hagenbüchle and P K Wellauer: *The p48 DNA-binding subunit of transcription factor PTF1 is a new exocrine pancreas-specific basic helix-loop-helix protein.* The EMBO journal, 15(16):4317–29, aug 1996.

[80] Krapp, A, M Knöfler, B Ledermann, K Bürki, C Berney, N Zoerkler, O Hagenbüchle and P K Wellauer: *The bHLH protein PTF1-p48 is essential for the formation of the exocrine and the correct spatial organization of the endocrine pancreas.* Genes & development, 12(23):3752–63, dec 1998.

[81] Rose, Scott D., Galvin H. Swift, Michael J. Peyton, Robert E. Hammer and Raymond J. MacDonald: *The Role of PTF1-P48 in Pancreatic Acinar Gene Expression.* Journal of Biological Chemistry, 276(47):44018–44026, nov 2001.

[82] Masui, Toshihiko, Galvin H. Swift, Tye Deering, Chengcheng Shen, Ward S. Coats, Qiaoming Long, Hans–Peter Elsässer, Mark A. Magnuson and Raymond J. MacDonald: *Replacement of Rbpj With Rbpjl in the PTF1 Complex Controls the Final Maturation of Pancreatic Acinar Cells.* Gastroenterology, 139(1):270–280, jul 2010.

[83] Holmstrom, S. R., T. Deering, G. H. Swift, F. J. Poelwijk, D. J. Mangelsdorf, S. A. Kliewer and R. J. MacDonald: *LRH-1 and PTF1-L coregulate an exocrine pancreas-specific transcriptional network for digestive function.* Genes & Development, 25(16):1674–1679, aug 2011.

[84] Lee, J C, S B Smith, H Watada, J Lin, D Scheel, J Wang, R G Mirmira and M S German: *Regulation of the pancreatic pro-endocrine gene neurogenin3.* Diabetes, 50(5):928–36, may 2001.

[85] Qu, Xiaoling, Solomon Afelik, Jan Nygaard Jensen, Michael A. Bukys, Sune Kobberup, Martin Schmerr, Fan Xiao, Pia Nyeng, Maria Veronica Albertoni, Anne Grapin-Botton and Jan Jensen: *Notch-mediated post-translational control of Ngn3 protein stability regulates pancreatic patterning and cell fate commitment.* Developmental Biology, 376(1):1–12, apr 2013.

[86] Shih, H. P., J. L. Kopp, M. Sandhu, C. L. Dubois, P. A. Seymour, A. Grapin-Botton and M. Sander: *A Notch-dependent molecular circuitry initiates pancreatic endocrine and ductal cell differentiation.* Development, 139(14):2488–2499, 2012.

[87] Ejarque, Miriam, Sara Cervantes, Gemma Pujadas, Anna Tutusaus, Lidia Sanchez and Rosa Gasa: *Neurogenin3 cooperates with Foxa2 to autoactivate its own expression.* The Journal of biological chemistry, pages 1–23, mar 2013.

[88] Afelik, Solomon and Jan Jensen: *Notch signaling in the pancreas: Patterning and cell fate specification.* Wiley Interdisciplinary Reviews: Developmental Biology, 2(4):531–544, 2013.

[89] Magenheim, Judith, Allon M. Klein, Ben Z. Stanger, Ruth Ashery-Padan, Beatriz Sosa-Pineda, Guoqiang Gu and Yuval Dor: *Ngn3+ endocrine progenitor cells control the fate and morphogenesis of pancreatic ductal epithelium.* Developmental Biology, 359(1):26–36, nov 2011.

[90] Petri, Andreas, Jonas Ahnfelt-Rønne, Klaus Stensgaard Frederiksen, David George Edwards, Dennis Madsen, Palle Serup, Jan Fleckner and R Scott Heller: *The effect of neurogenin3 deficiency on pancreatic gene expression in embryonic mice.* Journal of Molecular Endocrinology, 37(2):301–316, oct 2006.

[91] Collombat, P.: *Opposing actions of Arx and Pax4 in endocrine pancreas development.* Genes & Development, 17(20):2591–2603, oct 2003.

[92] Collombat, P.: *The simultaneous loss of Arx and Pax4 genes promotes a somatostatin-producing cell fate specification at the expense of the - and -cell lineages in the mouse endocrine pancreas.* Development, 132(13):2969–2980, jul 2005.

[93] Bramswig, N. C. and K. H. Kaestner: *Transcriptional regulation of α-cell differentiation.* Diabetes, Obesity and Metabolism, 13:13–20, oct 2011.

[94] Sosa-Pineda, Beatriz, Kamal Chowdhury, Miguel Torres, Guillermo Oliver and Peter Gruss: *The Pax4 gene is essential for differentiation of insulin-producing β cells in the mammalian pancreas.* Nature, 386(6623):399–402, mar 1997.

[95] Gannon, Maureen, Elizabeth Tweedie Ables, Laura Crawford, David Lowe, Martin F. Offield, Mark A. Magnuson and Christopher V.E. Wright: *pdx-1 function is specifically required in embryonic β cells to generate appropriate numbers of endocrine cell types and maintain glucose homeostasis.* Developmental Biology, 2008.

[96] Henseleit, K. D.: *NKX6 transcription factor activity is required for - and -cell development in the pancreas.* Development, 2005.

[97] Schaffer, Ashleigh E., Brandon L. Taylor, Jacqueline R. Benthuysen, Jingxuan Liu, Fabrizio Thorel, Weiping Yuan, Yang Jiao, Klaus H. Kaestner, Pedro L. Herrera, Mark A. Magnuson, Catherine Lee May and Maike Sander: *Nkx6.1 Controls a Gene Regulatory Network Required for Establishing and Maintaining Pancreatic Beta Cell Identity.* PLoS Genetics, 9(1):e1003274, jan 2013.

[98] Yang, Y.-P., Fabrizio Thorel, Daniel F. Boyer, Pedro L. Herrera and C. V. E. Wright: *Context-specific -to--cell reprogramming by forced Pdx1 expression.* Genes & Development, 25(16):1680–1685, aug 2011.

[99] Mastracci, Teresa L., Keith R. Anderson, James B. Papizan and Lori Sussel: *Regulation of Neurod1 Contributes to the Lineage Potential of Neurogenin3+ Endocrine Precursor Cells in the Pancreas*. PLoS Genetics, 9(2):e1003278, feb 2013.

[100] Papizan, J. B., R. A. Singer, S.-I. Tschen, S. Dhawan, J. M. Friel, S. B. Hipkens, M. A. Magnuson, A. Bhushan and L. Sussel: *Nkx2.2 repressor complex regulates islet -cell specification and prevents -to- -cell reprogramming*. Genes & Development, 25(21):2291–2305, nov 2011.

[101] Artner, I., B. Blanchi, J. C. Raum, M. Guo, T. Kaneko, S. Cordes, M. Sieweke and R. Stein: *MafB is required for islet beta cell maturation*. Proceedings of the National Academy of Sciences, 104(10):3853–3858, mar 2007.

[102] Conrad, Elizabeth, Chunhua Dai, Jason Spaeth, Min Guo, Holly A. Cyphert, David Scoville, Julie Carroll, Wei-Ming Yu, Lisa V. Goodrich, David M. Harlan, Kevin L. Grove, Charles T. Roberts, Alvin C. Powers, Guoqiang Gu and Roland Stein: *The MAFB transcription factor impacts islet α-cell function in rodents and represents a unique signature of primate islet β-cells*. American Journal of Physiology-Endocrinology and Metabolism, 310(1):E91–E102, jan 2016.

[103] Nishimura, Wataru, Takuma Kondo, Therese Salameh, Ilham El Khattabi, Rikke Dodge, Susan Bonner-Weir and Arun Sharma: *A switch from MafB to MafA expression accompanies differentiation to pancreatic β-cells*. Developmental Biology, 293(2):526–539, may 2006.

[104] Arda, H. Efsun, Cecil M. Benitez and Seung K. Kim: *Gene Regulatory Networks Governing Pancreas Development*. Developmental Cell, 25(1):5–13, apr 2013.

[105] Gao, Tao, Brian McKenna, Changhong Li, Maximilian Reichert, James Nguyen, Tarjinder Singh, Chenghua Yang, Archana Pannikar, Nicolai Doliba, Tingting Zhang, Doris A. Stoffers, Helena Edlund, Franz Matschinsky, Roland Stein and Ben Z. Stanger: *Pdx1 Maintains β Cell Identity and Function by Repressing an α Cell Program*. Cell Metabolism, 19(2):259–271, feb 2014.

[106] Taylor, BrandonL, Fen Fen Liu and Maike Sander: *Nkx6.1 Is Essential for Maintaining the Functional State of Pancreatic Beta Cells*. Cell Reports, 4(6):1262–1275, 2013.

[107] Zhang, Chuan, Takashi Moriguchi, Miwako Kajihara, Ayako Harada, Homare Shimohata, Hisashi Oishi, Michito Hamada, Naoki Morito, Kazuteru Hasegawa, Takashi Kudo, James Douglas Engel, Masayuki Yamamoto, Ritsuko Esaki and Satoru Takahashi: *MafA Is a Key Regulator of Glucose-Stimulated Insulin Secretion MafA Is a Key Regulator of Glucose-Stimulated Insulin Secretion*. Molecular and cellular biology, 25(12):4969–76, 2005.

[108] Blum, Barak, Siniša Hrvatin, Christian Schuetz, Claire Bonal, Alireza Rezania and Douglas A Melton: *Functional beta-cell maturation is marked by an increased glucose threshold and by expression of urocortin 3*. Nature Biotechnology, 30(3):261–264, mar 2012.

[109] Doyle, Michelle J and Lori Sussel: *Nkx2 . 2 Regulates B- Cell Function in the Mature Islet*. Diabetes, 56(August):1999–2007, 2007.

[110] Hani, El Habib, Doris A. Stoffers, Jean-Claude Chèvre, Emmanuelle Durand, Violeta Stanojevic, Christian Dina, Joel F. Habener and Philippe Froguel: *Defective mutations in the insulin promoter factor-1 (IPF-1) gene in late-onset type 2 diabetes mellitus*. Journal of Clinical Investigation, 104(9):R41–R48, nov 1999.

[111] Brooke, Nina M, Jordi Garcia-Fernàndez and Peter W H Holland: *The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster*. Nature, 392:920, apr 1998.

[112] Fiedorek, Fred T. and Eric S. Kay: *Mapping of the Insulin Promoter Factor I Gene (Ipf1) to Distal Mouse Chromosome 5*. Genomics, 28(3):581–584, aug 1995.

[113] Zhou, G and FC Brunicardi: *PDX1 (pancreatic and duodenal homeobox 1)*. Atlas of Genetics and Cytogenetics in Oncology and Haematology, (6), nov 2011.

[114] Gerrish, Kevin, Jennifer C. Van Velkinburgh and Roland Stein: *Conserved Transcriptional Regulatory Domains of the pdx-1 Gene* . Molecular Endocrinology, 18(3):533–548, 2004.

[115] Ohneda, K, R G Mirmira, J Wang, J D Johnson and M S German: *The homeodomain of PDX-1 mediates multiple protein-protein interactions in the formation of a transcriptional activation complex on the insulin promoter.* Molecular and cellular biology, 20(3):900–11, feb 2000.

[116] Ashizawa, Satoshi, F Charles Brunicardi and Xiao-Ping Wang: *PDX-1 and the pancreas.* Pancreas, 28(2):109–20, mar 2004.

[117] Moede, Tilo, Barbara Leibiger, Hamedeh Ghanaat Pour, Per-Olof Berggren and Ingo B Leibiger: *Identification of a nuclear localization signal, RRMKWKK, in the homeodomain transcription factor PDX-1*. FEBS Letters, 461(3):229–234, nov 1999.

[118] Noguchi, H., H. Kaneto, G. C. Weir and S. Bonner-Weir: *PDX-1 Protein Containing Its Own Antennapedia-Like Protein Transduction Domain Can Transduce Pancreatic Duct and Islet Cells*. Diabetes, 52(7):1732–1737, jul 2003.

[119] Peers, B, S Sharma, T Johnson, M Kamps and M Montminy: *The pancreatic islet factor STF-1 binds cooperatively with Pbx to a regulatory element in the somatostatin promoter: importance of the FPWMK motif and of the homeodomain.* Molecular and Cellular Biology, 15(12):7091–7097, dec 1995.

[120] Claiborn, Kathryn C., Mira M. Sachdeva, Corey E. Cannon, David N. Groff, Jeffrey D. Singer and Doris A. Stoffers: *Pcif1 modulates Pdx1 protein stability and pancreatic β cell function and survival in mice*. Journal of Clinical Investigation, 120(10):3713–3721, oct 2010.

[121] Ostertag, Michael Sebastian, Ana Cristina Messias, Michael Sattler and Grzegorz Maria Popowicz: *The Structure of the SPOP-Pdx1 Interface Reveals Insights into the Phosphorylation-Dependent Binding Regulation*. Structure, 27(2):327–334.e3, feb 2019.

[122] Oliver-Krasinski, Jennifer M., Margaret T Kasner, Juxiang Yang, Michael F Crutchlow, Anil K Rustgi, Klaus H Kaestner and Doris A Stoffers: *The diabetes gene Pdx1 regulates the transcriptional network of pancreatic endocrine progenitor cells in mice*. Journal of Clinical Investigation, 119(7):1888–1898, jul 2009.

[123] Liu, Aihua, Jennifer Oliver-Krasinski and Doris A. Stoffers: *Two conserved domains in PCIF1 mediate interaction with pancreatic transcription factor PDX-1*. FEBS Letters, 580(28-29):6701–6706, 2006.

[124] Klein, Sabrina, Rui Meng, Mathias Montenarh and Claudia Götz: *The Phosphorylation of PDX-1 by Protein Kinase CK2 Is Crucial for Its Stability*. Pharmaceuticals, 10(4):2, dec 2016.

[125] Stoffers, Doris A., R. Scott Heller, Christopher P. Miller and Joel F. Habener: *Developmental Expression of the Homeodomain Protein IDX-1 in Mice Transgenic for an IDX-1 Promoter/ lacZ Transcriptional Reporter 1*. Endocrinology, 140(11):5374–5381, nov 1999.

[126] Larsson, Lars-Inge, Ole D. Madsen, Palle Serup, Jörgen Jonsson and Helena Edlund: *Pancreatic-duodenal homeobox 1 -role in gastric endocrine patterning*. Mechanisms of Development, 60(2):175–184, dec 1996.

[127] Perez-Villamil, Beatriz, Petra T. Schwartz and Mario Vallejo: *The Pancreatic Homeodomain Transcription Factor IDX1/IPF1 Is Expressed in Neural Cells during Brain Development*. Endocrinology, 140(8):3857–3857, aug 1999.

[128] Schwartz, Petra T., Beatriz Pérez-Villamil, Alicia Rivera, Rosario Moratalla and Mario Vallejo: *Pancreatic Homeodomain Transcription Factor IDX1/IPF1 Expressed in Developing Brain Regulates Somatostatin Gene Transcription in Embryonic Neural Cells*. Journal of Biological Chemistry, 275(25):19106–19114, jun 2000.

[129] Holland, Andrew M., Sonia Garcia, Gaetano Naselli, Raymond J. MacDonald and Leonard C. Harrison: *The Parahox gene Pdx1 is required to maintain positional identity in the adult foregut*. The International Journal of Developmental Biology, 57(5):391–398, 2013.

[130] Jepeal, Lisa I., Yoshio Fujitani, Michael O. Boylan, Cherrell N. Wilson, Christopher V. Wright and M. Michael Wolfe: *Cell-Specific Expression of Glucose-Dependent-Insulinotropic Polypeptide Is Regulated by the Transcription Factor PDX-1*. Endocrinology, 146(1):383–391, jan 2005.

[131] Brissova, Marcela, Masakazu Shiota, Wendell E. Nicholson, Maureen Gannon, Susan M. Knobel, David W. Piston, Christopher V E Wright and Alvin C. Powers: *Reduction in pancreatic transcription factor PDX-1 impairs glucose-stimulated insulin secretion*. Journal of Biological Chemistry, 277(13):11225–11232, 2002.

[132] Johnson, James D., Noreen T. Ahmed, Dan S. Luciani, Zhiqiang Han, Hung Tran, Jun Fujita, Stanley Misler, Helena Edlund and Kenneth S. Polonsky: *Increased islet apoptosis in Pdx1+/- mice*. Journal of Clinical Investigation, 111(8):1147–1160, 2003.

[133] Holland, Andrew M., L. J. Gonez, Gaetano Naselli, Raymond J. MacDonald and Leonard C. Harrison: *Conditional Expression Demonstrates the Role of the Homeodomain Transcription Factor Pdx1 in Maintenance and Regeneration of -Cells in the Adult Pancreas*. Diabetes, 54(9):2586–2595, sep 2005.

[134] Stoffers, Doris A., Violeta Stanojevic and Joel F. Habener: *Insulin promoter factor-1 gene mutation linked to early-onset type 2 diabetes mellitus directs expression of a dominant negative isoprotein*. Journal of Clinical Investigation, 102(1):232–241, 1998.

[135] Cockburn, Brian N., Giovanna Bermano, Laura Lee G. Boodram, Surujpal Teelucksingh, Takafumi Tsuchiya, Deepak Mahabir, Andrew B. Allan, Roland Stein, Kevin Docherty and Graeme I. Bell: *Insulin Promoter Factor-1 Mutations and Diabetes in Trinidad: Identification of a Novel Diabetes-Associated Mutation (E224K) in an Indo-Trinidadian Family*. Journal of Clinical Endocrinology and Metabolism, 89(2):971–978, 2004.

[136] Macfarlane, Wendy M., Timothy M. Frayling, Sian Ellard, Julie C. Evans, Lisa I.S. Allen, Michael P. Bulman, Susan Ayres, Maggie Shepherd, Penny Clark, Ann Millward, Andrew Demaine, Terence Wilkin, Kevin Docherty and Andrew T. Hattersley: *Missense mutations in the insulin promoter factor-1 gene predispose to type 2 diabetes*. Journal of Clinical Investigation, 104(9), 1999.

[137] Nicolino, M., K. C. Claiborn, V. Senee, A. Boland, D. A. Stoffers and C. Julier: *A Novel Hypomorphic PDX1 Mutation Responsible for Permanent Neonatal Diabetes With Subclinical Exocrine Deficiency*. Diabetes, 59(3):733–740, mar 2010.

[138] Weng, J., W. M. Macfarlane, M. Lehto, H. F. Gu, L. M. Shepherd, S. A. Ivarsson, L. Wibell, T. Smith and L. C. Groop: *Functional consequences of mutations in the MODY4 gene (IPF1) and coexistence with MODY3 mutations*. Diabetologia, 44(2):249–258, feb 2001.

[139] Gragnoli, Claudia, Violeta Stanojevic, Antonio Gorini, Guido Menzinger Von Preussenthal, Melissa K. Thomas and Joel F. Habener: *IPF-1/MODY4 gene missense mutation in an Italian family with type 2 and gestational diabetes*. Metabolism, 54(8):983–988, aug 2005.

[140] Jennings, R. E., A. A. Berry, J. P. Strutt, D. T. Gerrard and N. A. Hanley: *Human pancreas development*. Development, 142(18):3126–3137, 2015.

[141] Jennings, R. E., A. A. Berry, R. Kirkwood-Wilson, N. A. Roberts, T. Hearn, R. J. Salisbury, J. Blaylock, K. Piper Hanley and N. A. Hanley: *Development of the Human Pancreas From Foregut to Endocrine Commitment*. Diabetes, 62(10):3514–3522, oct 2013.

[142] Pan, Fong Cheng and Marcela Brissova: *Pancreas development in humans*. Current Opinion in Endocrinology & Diabetes and Obesity, 21(2):77–82, apr 2014.

[143] Salisbury, Rachel J, Jennifer Blaylock, Andrew A Berry, Rachel E Jennings, Ronald De Krijger, Karen Piper Hanley and Neil A Hanley: *The window period of NEUROGENIN3 during human gestation*. Islets, 6(3):e954436, may 2014.

[144] Piper, K: *Beta cell differentiation during early human pancreas development*. Journal of Endocrinology, 181(1):11–23, apr 2004.

[145] Piper Hanley, Karen, Tom Hearn, Andrew Berry, Melanie J Carvell, Ann-Marie Patch, Louise J Williams, Sarah A Sugden, David I Wilson, Sian Ellard and Neil A Hanley: *In vitro expression of NGN3 identifies RAB3B as the predominant Ras-associated GTP-binding protein 3 family member in human islets.* Journal of Endocrinology, 207(2):151–161, nov 2010.

[146] Meier, Juris J, Christina U Köhler, Bacel Alkhatib, Consolato Sergi, Theresa Junker, Harald H Klein, Wolfgang E Schmidt and Helga Fritsch: *β-cell development and turnover during prenatal life in humans.* European Journal of Endocrinology, 162(3):559–568, mar 2010.

[147] Jeon, Jongmin, Mayrin Correa-Medina, Camillo Ricordi, Helena Edlund and Juan A. Diez: *Endocrine Cell Clustering During Human Pancreas Development.* Journal of Histochemistry & Cytochemistry, 57(9):811–824, sep 2009.

[148] Otonkoski, T., S. Andersson, M. Knip and O. Simell: *Maturation of insulin response to glucose during human fetal and neonatal development. Studies with perifusion of pancreatic isletlike cell clusters.* Diabetes, 1988.

[149] Otonkoski, Timo, Mikael Knip, Inés Wong and Olli Simell: *Lack of glucose-induced functional maturation during long-term culture of human fetal islet cells.* Life Sciences, 1991.

[150] Shapiro, A.M. James, Camillo Ricordi, Bernhard J. Hering, Hugh Auchincloss, Robert Lindblad, R. Paul Robertson, Antonio Secchi, Mathias D. Brendel, Thierry Berney, Daniel C. Brennan, Enrico Cagliero, Rodolfo Alejandro, Edmond A. Ryan, Barbara DiMercurio, Philippe Morel, Kenneth S. Polonsky, Jo-Anna Reems, Reinhard G. Bretzel, Federico Bertuzzi, Tatiana Froud, Raja Kandaswamy, David E.R. Sutherland, George Eisenbarth, Miriam Segal, Jutta Preiksaitis, Gregory S. Korbutt, Franca B. Barton, Lisa Viviano, Vicki Seyfert-Margolis, Jeffrey Bluestone and Jonathan R.T. Lakey: *International Trial of the Edmonton Protocol for Islet Transplantation.* New England Journal of Medicine, 2006.

[151] Bakhti, Mostafa, Anika Böttcher and Heiko Lickert: *Modelling the endocrine pancreas in health and disease.* Nature Reviews Endocrinology, nov 2018.

[152] Assady, Suheir, Gila Maor, Michal Amit, Joseph Itskovitz-Eldor, Karl L. Skorecki and Maty Tzukerman: *Insulin Production by Human Embryonic Stem Cells.* Diabetes, 2001.

[153] D'Amour, Kevin A, Anne G Bang, Susan Eliazer, Olivia G Kelly, Alan D Agulnick, Nora G Smart, Mark A Moorman, Evert Kroon, Melissa K Carpenter and Emmanuel E Baetge: *Production of pancreatic hormone–expressing endocrine cells from human embryonic stem cells.* Nature Biotechnology, 24(11):1392–1401, 2006.

[154] Kroon, Evert, Laura A. Martinson, Kuniko Kadoya, Anne G. Bang, Olivia G. Kelly, Susan Eliazer, Holly Young, Mike Richardson, Nora G. Smart, Justine Cunningham, Alan D. Agulnick, Kevin A. D'Amour, Melissa K. Carpenter and Emmanuel E. Baetge: *Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells in vivo.* Nature Biotechnology, 2008.

[155] Rezania, Alireza, Jennifer E Bruin, Payal Arora, Allison Rubin, Irina Batushansky, Ali Asadi, Shannon O'Dwyer, Nina Quiskamp, Majid Mojibian, Tobias Albrecht, Yu Hsuan Carol Yang, James D Johnson and Timothy J Kieffer: *Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells.* Nature Biotechnology, 32(11):1121–1133, 2014.

[156] Pagliuca, Felicia W., Jeffrey R. Millman, Mads Gürtler, Michael Segel, Alana Van Dervort, Jennifer Hyoje Ryu, Quinn P. Peterson, Dale Greiner and Douglas A. Melton: *Generation of functional human pancreatic β cells in vitro.* Cell, 159(2):428–439, 2014.

[157] Russ, Holger a, Audrey V Parent, Jennifer J Ringler, Thomas G Hennings, Gopika G Nair, Mayya Shveygert, Tingxia Guo, Sapna Puri, Leena Haataja, Vincenzo Cirulli, Robert Blelloch, Greg L Szot, Peter Arvan and Matthias Hebrok: *Controlled induction of human pancreatic progenitors produces functional beta-like cells in vitro.* The EMBO journal, 34(13):e201591058, 2015.

[158] Haumaitre, Cécile, Mélanie Fabre, Sarah Cormier, Clarisse Baumann, Anne Lise Delezoide and Silvia Cereghini: *Severe pancreas hypoplasia and multicystic renal dysplasia in two human fetuses carrying novel HNF1β/MODY5 mutations.* Human Molecular Genetics, 2006.

[159] Mayer, Christof, Yvonne Böttcher, Peter Kovacs, Jan Halbritter and Michael Stumvoll: *Phenotype of a patient with a de novo mutation in the hepatocyte nuclear factor 1β/maturity-onset diabetes of the young type 5 gene.* Metabolism: Clinical and Experimental, 2008.

[160] Janssen, Sara and Inge Depoortere: *Nutrient sensing in the gut: new roads to therapeutics?* Trends in endocrinology and metabolism: TEM, 24(2):92–100, feb 2013.

[161] Gribble, Fiona M. and Frank Reimann: *Enteroendocrine Cells: Chemosensors in the Intestinal Epithelium.* Annual Review of Physiology, 78(1):277–299, 2015.

[162] Haber, Adam L., Moshe Biton, Noga Rogel, Rebecca H. Herbst, Karthik Shekhar, Christopher Smillie, Grace Burgin, Toni M. Delorey, Michael R. Howitt, Yarden Katz, Itay Tirosh, Semir Beyaz, Danielle Dionne, Mei Zhang, Raktima Raychowdhury, Wendy S. Garrett, Orit Rozenblatt-Rosen, Hai Ning Shi, Omer Yilmaz, Ramnik J. Xavier and Aviv Regev: *A single-cell survey of the small intestinal epithelium.* Nature, 551(7680):333–339, nov 2017.

[163] Habib, Abdella M., Paul Richards, Lynne S. Cairns, Gareth J. Rogers, Christopher A. M. Bannon, Helen E. Parker, Tom C. E. Morley, Giles S. H. Yeo, Frank Reimann and Fiona M. Gribble: *Overlap of Endocrine Hormone Expression in the Mouse Intestine Revealed by Transcriptional Profiling and Flow Cytometry.* Endocrinology, 153(7):3054–3065, jul 2012.

[164] Schubert, Mitchell L. and David A. Peura: *Control of Gastric Acid Secretion in Health and Disease.* Gastroenterology, 134(7):1842–1860, jun 2008.

[165] Dupre, J., S.A. Ross, D. Watson and J.C. Brown: *STIMULATION OF INSULIN SECRETION BY GASTRIC INHIBITORY POLYPEPTIDE IN MAN. 1.* The Journal of Clinical Endocrinology & Metabolism, 37(5):826–828, nov 1973.

[166] Seino, Yutaka, Mitsuo Fukushima and Daisuke Yabe: *GIP and GLP-1, the two incretin hormones: Similarities and differences.* Journal of Diabetes Investigation, 1(1-2):8–23, feb 2010.

[167] Sekiguchi, Toshio: *Cholecystokinin.* In *Handbook of Hormones*, pages 177–e20B–3. Elsevier, 2016.

[168] Nakamachi, Tomoya: *Secretin.* In *Handbook of Hormones*, pages 142–e18A–2. Elsevier, 2016.

[169] Little, Tanya J, Michael Horowitz and Christine Feinle-Bisset: *Modulation by high-fat diets of gastrointestinal function and hormones associated with the regulation of energy intake: implications for the pathophysiology of obesity.* The American journal of clinical nutrition, 86(3):531–41, sep 2007.

[170] Park, Min Kyun: *Glucagon-Like Peptide-1.* In *Handbook of Hormones*, pages 135–e17C–5. Elsevier, 2016.

[171] Katsuura, Goro and Akio Inui: *Peptide YY.* In *Handbook of Hormones*, pages 217–218. Elsevier, 2016.

[172] Liddle, R A, E T Morita, C K Conrad and J A Williams: *Regulation of gastric emptying in humans by cholecystokinin.* Journal of Clinical Investigation, 77(3):992–996, mar 1986.

[173] Nauck, Michael A., Ulrich Niedereichholz, Rainer Ettler, Jens Juul Holst, Cathrine Ørskov, Robert Ritzel and Wolff H. Schmiegel: *Glucagon-like peptide 1 inhibition of gastric emptying outweighs its insulinotropic effects in healthy humans.* American Journal of Physiology-Endocrinology and Metabolism, 273(5):E981–E988, nov 1997.

[174] Savage, A P, T E Adrian, G Carolan, V K Chatterjee and S R Bloom: *Effects of peptide YY (PYY) on mouth to caecum intestinal transit time and on the rate of gastric emptying in healthy volunteers.* Gut, 28(2):166–70, feb 1987.

[175] Batterham, Rachel L., Michael A. Cowley, Caroline J. Small, Herbert Herzog, Mark A. Cohen, Catherine L. Dakin, Alison M. Wren, Audrey E. Brynes, Malcolm J. Low, Mohammad A. Ghatei,

Roger D. Cone and Stephen R. Bloom: *Gut hormone PYY3-36 physiologically inhibits food intake.* Nature, 418(6898):650–654, aug 2002.

[176] Ahrén, B.: *Incretin dysfunction in type 2 diabetes: Clinical impact and future perspectives.* Diabetes & Metabolism, 39(3):195–201, may 2013.

[177] Drucker, D. J., S. I. Sherman, F. S. Gorelick, R. M. Bergenstal, R. S. Sherwin and J. B. Buse: *Incretin-Based Therapies for the Treatment of Type 2 Diabetes: Evaluation of the Risks and Benefits.* Diabetes Care, 33(2):428–433, feb 2010.

[178] Lieberkühn, Joannis Nathanael: *Dissertatio de fabrica et actione villorum intestinorum tenuium hominis.* Conrad et. Georg. Jac. Wishof, Leiden, 1745.

[179] Gehart, Helmuth and Hans Clevers: *Tales from the crypt: new insights into intestinal stem cells.* Nature Reviews Gastroenterology & Hepatology, 16(1):19–34, jan 2019.

[180] Barker, Nick and Hans Clevers: *Tracking down the stem cells of the intestine: strategies to identify adult stem cells.* Gastroenterology, 133(6):1755–60, dec 2007.

[181] Barker, Nick: *Adult intestinal stem cells: critical drivers of epithelial homeostasis and regeneration.* Nature reviews. Molecular cell biology, 15(1):19–33, jan 2014.

[182] Watson, Alastair J.M. and Kevin R. Hughes: *TNF-α-induced intestinal epithelial cell shedding: implications for intestinal barrier function.* Annals of the New York Academy of Sciences, 1258(1):1–8, jul 2012.

[183] Magney, J. E., S. L. Erlandsen, Matthew L. Bjerknes and Hazel Cheng: *Scanning electron microscopy of isolated epithelium of the murine gastrointestinal tract: Morphology of the basal surface and evidence for paracrinelike cells.* American Journal of Anatomy, 177(1):43–53, sep 1986.

[184] H Ross, Michael and Wojciech Pawlina: *Histology: A Text and Atlas: With Correlated Cell and Molecular Biology*, volume xvii. jan 2006.

[185] Mabbott, N A, D S Donaldson, H Ohno, I R Williams and A Mahajan: *Microfold (M) cells: important immunosurveillance posts in the intestinal epithelium.* Mucosal Immunology, 6(4):666–677, jul 2013.

[186] Birchenough, George M.H., Elisabeth E.L. Nystrom, Malin E.V. Johansson and Gunnar C. Hansson: *A sentinel goblet cell guards the colonic crypt by triggering Nlrp6-dependent Muc2 secretion.* Science, 352(6293):1535–1542, 2016.

[187] Gerbe, François, Johan H. van Es, Leila Makrini, Bénédicte Brulin, Georg Mellitzer, Sylvie Robine, Béatrice Romagnolo, Noah F. Shroyer, Jean-François Bourgaux, Christine Pignodel, Hans Clevers and Philippe Jay: *Distinct ATOH1 and Neurog3 requirements define tuft cells as a new secretory cell type in the intestinal epithelium.* The Journal of Cell Biology, 192(5):767–780, mar 2011.

[188] Moltke, Jakob von, Ming Ji, Hong-Erh Liang and Richard M. Locksley: *Tuft-cell-derived IL-25 regulates an intestinal ILC2–epithelial response circuit.* Nature, 529(7585):221–225, jan 2016.

[189] Gerbe, François, Emmanuelle Sidot, Danielle J. Smyth, Makoto Ohmoto, Ichiro Matsumoto, Valérie Dardalhon, Pierre Cesses, Laure Garnier, Marie Pouzolles, Bénédicte Brulin, Marco Bruschi, Yvonne Harcus, Valérie S. Zimmermann, Naomi Taylor, Rick M. Maizels and Philippe Jay: *Intestinal epithelial tuft cells initiate type 2 mucosal immunity to helminth parasites.* Nature, 529(7585):226–230, jan 2016.

[190] Sternini, Catia, Laura Anselmi and Enrique Rozengurt: *Enteroendocrine cells: a site of 'taste' in gastrointestinal chemosensing.* Current Opinion in Endocrinology, Diabetes and Obesity, 15(1):73–78, feb 2008.

[191] Gehart, Helmuth, Johan H. van Es, Karien Hamer, Joep Beumer, Kai Kretzschmar, Johanna F. Dekkers, Anne Rios and Hans Clevers: *Identification of Enteroendocrine Regulators by Real-Time Single-Cell Differentiation Mapping.* Cell, 176(5):1158–1173.e16, feb 2019.

[192] Beumer, Joep, Benedetta Artegiani, Yorick Post, Frank Reimann, Fiona Gribble, Thuc Nghi Nguyen, Hongkui Zeng, Maaike Van den Born, Johan H. Van Es and Hans Clevers: *Enteroendocrine cells switch hormone expression along the crypt-to-villus BMP signalling gradient*. Nature Cell Biology, 20(8):909–916, aug 2018.

[193] Batlle, Eduard, Jeffrey T. Henderson, Harry Beghtel, Maaike M.W. van den Born, Elena Sancho, Gerwin Huls, Jan Meeldijk, Jennifer Robertson, Marc van de Wetering, Tony Pawson and Hans Clevers: *β-Catenin and TCF Mediate Cell Positioning in the Intestinal Epithelium by Controlling the Expression of EphB/EphrinB*. Cell, 111(2):251–263, oct 2002.

[194] Genander, Maria, Michael M. Halford, Nan-Jie Xu, Malin Eriksson, Zuoren Yu, Zhaozhu Qiu, Anna Martling, Gedas Greicius, Sonal Thakar, Timothy Catchpole, Michael J. Chumley, Sofia Zdunek, Chenguang Wang, Torbjörn Holm, Stephen P. Goff, Sven Pettersson, Richard G. Pestell, Mark Henkemeyer and Jonas Frisén: *Dissociation of EphB2 Signaling Pathways Mediating Progenitor Cell Proliferation and Tumor Suppression*. Cell, 139(4):679–692, nov 2009.

[195] Ireland, Heather, Carol Houghton, Louise Howard and Douglas J. Winton: *Cellular inheritance of a Cre-activated reporter gene to determine paneth cell longevity in the murine small intestine*. Developmental Dynamics, 233(4):1332–1336, aug 2005.

[196] Gassler, Nikolaus: *Paneth cells in intestinal physiology and pathophysiology*. World Journal of Gastrointestinal Pathophysiology, 8(4):150–160, nov 2017.

[197] Allaire, Joannie M., Shauna M. Crowley, Hong T. Law, Sun-Young Chang, Hyun-Jeong Ko and Bruce A. Vallance: *The Intestinal Epithelium: Central Coordinator of Mucosal Immunity*. Trends in Immunology, 39(9):677–696, sep 2018.

[198] Sato, Toshiro, Johan H van Es, Hugo J Snippert, Daniel E Stange, Robert G Vries, Maaike van den Born, Nick Barker, Noah F Shroyer, Marc van de Wetering and Hans Clevers: *Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts.* Nature, 469(7330):415–8, jan 2011.

[199] Farin, Henner F., Johan H. Van Es and Hans Clevers: *Redundant Sources of Wnt Regulate Intestinal Stem Cells and Promote Formation of Paneth Cells*. Gastroenterology, 143(6):1518–1529.e7, dec 2012.

[200] Meran, Laween, Anna Baulies and Vivian S. W. Li: *Intestinal Stem Cell Niche: The Extracellular Matrix and Cellular Components*. Stem Cells International, 2017:1–11, 2017.

[201] Shoshkes-Carmel, Michal, Yue J Wang, Kirk J Wangensteen, Beáta Tóth, Ayano Kondo, Efi E Massasa, Shalev Itzkovitz and Klaus H Kaestner: *Subepithelial telocytes are an important source of Wnts that supports intestinal crypts*. Nature, 557(7704):242–246, may 2018.

[202] Aoki, Reina, Michal Shoshkes-Carmel, Nan Gao, Soona Shin, Catherine L. May, Maria L. Golson, Adam M. Zahm, Michael Ray, Caroline L. Wiser, Christopher V.E. Wright and Klaus H. Kaestner: *Foxl1-Expressing Mesenchymal Cells Constitute the Intestinal Stem Cell Niche*. Cellular and Molecular Gastroenterology and Hepatology, 2(2):175–188, mar 2016.

[203] Valenta, Tomas, Bahar Degirmenci, Andreas E. Moor, Patrick Herr, Dario Zimmerli, Matthias B. Moor, George Hausmann, Claudio Cantù, Michel Aguet and Konrad Basler: *Wnt Ligands Secreted by Subepithelial Mesenchymal Cells Are Essential for the Survival of Intestinal Stem Cells and Gut Homeostasis*. Cell Reports, 15(5):911–918, may 2016.

[204] Stzepourginski, Igor, Giulia Nigro, Jean-Marie Jacob, Sophie Dulauroy, Philippe J. Sansonetti, Gérard Eberl and Lucie Peduto: *CD34 + mesenchymal cells are a major component of the intestinal stem cells niche at homeostasis and after injury*. Proceedings of the National Academy of Sciences, 114(4):E506–E513, jan 2017.

[205] Snippert, Hugo J, Laurens G van der Flier, Toshiro Sato, Johan H van Es, Maaike van den Born, Carla Kroon-Veenboer, Nick Barker, Allon M Klein, Jacco van Rheenen, Benjamin D Simons and

Hans Clevers: *Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells.* Cell, 143(1):134–44, oct 2010.

[206] Lopez-Garcia, C., A. M. Klein, B. D. Simons and D. J. Winton: *Intestinal Stem Cell Replacement Follows a Pattern of Neutral Drift.* Science, 330(6005):822–825, nov 2010.

[207] Vermeulen, L., E. Morrissey, M. van der Heijden, A. M. Nicholson, A. Sottoriva, S. Buczacki, R. Kemp, S. Tavare and D. J. Winton: *Defining Stem Cell Dynamics in Models of Intestinal Tumor Initiation.* Science, 342(6161):995–998, nov 2013.

[208] Snippert, Hugo J., Arnout G. Schepers, Johan H. van Es, Benjamin D. Simons and Hans Clevers: *Biased competition between Lgr5 intestinal stem cells driven by oncogenic mutation induces clonal expansion.* EMBO reports, 15(1):62–69, jan 2014.

[209] Potten, Christopher S., Laszlo Kovacs and Elizabeth Hamilton: *CONTINUOUS LABELLING STUDIES ON MOUSE SKIN AND INTESTINE.* Cell Proliferation, 7(3):271–283, may 1974.

[210] Potten, Christopher S., William J. Hume, Peter Reid and John Cairns: *The segregation of DNA in epithelial stem cells.* Cell, 15(3):899–906, nov 1978.

[211] Potten, Christopher S, Gary Owen and Dawn Booth: *Intestinal stem cells protect their genome by selective segregation of template DNA strands.* Journal of Cell Science, 115(11):2381 LP – 2388, jun 2002.

[212] Sangiorgi, Eugenio and Mario R Capecchi: *Bmi1 is expressed in vivo in intestinal stem cells.* Nature Genetics, 40(7):915–920, jul 2008.

[213] Breault, D. T., I. M. Min, D. L. Carlone, L. G. Farilla, D. M. Ambruzs, D. E. Henderson, S. Algra, R. K. Montgomery, A. J. Wagers and N. Hole: *Generation of mTert-GFP mice as a model to identify and study tissue progenitor cells.* Proceedings of the National Academy of Sciences, 105(30):10420–10425, jul 2008.

[214] Montgomery, R. K., D. L. Carlone, C. A. Richmond, L. Farilla, M. E. G. Kranendonk, D. E. Henderson, N. Y. Baffour-Awuah, D. M. Ambruzs, L. K. Fogli, S. Algra and D. T. Breault: *Mouse telomerase reverse transcriptase (mTert) expression marks slowly cycling intestinal stem cells.* Proceedings of the National Academy of Sciences, 108(1):179–184, jan 2011.

[215] Takeda, N., R. Jain, M. R. LeBoeuf, Q. Wang, M. M. Lu and J. A. Epstein: *Interconversion Between Intestinal Stem Cell Populations in Distinct Niches.* Science, 334(6061):1420–1424, dec 2011.

[216] Powell, Anne E., Yang Wang, Yina Li, Emily J. Poulin, Anna L. Means, Mary K. Washington, James N. Higginbotham, Alwin Juchheim, Nripesh Prasad, Shawn E. Levy, Yan Guo, Yu Shyr, Bruce J. Aronow, Kevin M. Haigis, Jeffrey L. Franklin and Robert J. Coffey: *The Pan-ErbB Negative Regulator Lrig1 Is an Intestinal Stem Cell Marker that Functions as a Tumor Suppressor.* Cell, 149(1):146–158, mar 2012.

[217] Tetteh, Paul W., Henner F. Farin and Hans Clevers: *Plasticity within stem cell hierarchies in mammalian epithelia.* Trends in Cell Biology, 25(2):100–108, feb 2015.

[218] Es, Johan H van, Toshiro Sato, Marc van de Wetering, Anna Lyubimova, Annie Ng Yee Nee, Alex Gregorieff, Nobuo Sasaki, Laura Zeinstra, Maaike van den Born, Jeroen Korving, Anton C M Martens, Nick Barker, Alexander van Oudenaarden and Hans Clevers: *Dll1+ secretory progenitor cells revert to stem cells upon crypt damage.* Nature cell biology, 14(10):1099–104, oct 2012.

[219] Yan, Kelley S., Olivier Gevaert, Grace X.Y. Zheng, Benedict Anchang, Christopher S. Probert, Kathryn A. Larkin, Paige S. Davies, Zhuan fen Cheng, John S. Kaddis, Arnold Han, Kelly Roelf, Ruben I. Calderon, Esther Cynn, Xiaoyi Hu, Komal Mandleywala, Julie Wilhelmy, Sue M. Grimes, David C. Corney, Stéphane C. Boutet, Jessica M. Terry, Phillip Belgrader, Solongo B. Ziraldo, Tarjei S. Mikkelsen, Fengchao Wang, Richard J. von Furstenberg, Nicholas R. Smith, Parthasarathy Chandrakesan, Randal May, Mary Ann S. Chrissy, Rajan Jain, Christine A. Cartwright, Joyce C. Niland, Young Kwon Hong, Jill Carrington, David T. Breault, Jonathan Epstein, Courtney W. Houchen, John P. Lynch, Martin G. Martin, Sylvia K. Plevritis, Christina Curtis, Hanlee P. Ji,

Linheng Li, Susan J. Henning, Melissa H. Wong and Calvin J. Kuo: *Intestinal Enteroendocrine Lineage Cells Possess Homeostatic and Injury-Inducible Stem Cell Activity*. Cell Stem Cell, 21(1):78–90.e6, 2017.

[220] Tetteh, Paul W., Onur Basak, Henner F. Farin, Kay Wiebrands, Kai Kretzschmar, Harry Begthel, Maaike van den Born, Jeroen Korving, Frederic de Sauvage, Johan H. van Es, Alexander van Oudenaarden and Hans Clevers: *Replacement of Lost Lgr5-Positive Stem Cells through Plasticity of Their Enterocyte-Lineage Daughters*. Cell Stem Cell, 18(2):203–213, feb 2016.

[221] Buczacki, Simon J. a., Heather Ireland Zecchini, Anna M. Nicholson, Roslin Russell, Louis Vermeulen, Richard Kemp and Douglas J. Winton: *Intestinal label-retaining cells are secretory precursors expressing Lgr5*. Nature, 1, feb 2013.

[222] Kim, Tae-hee, Fugen Li, Isabel Ferreiro-Neira, Li-Lun Ho, Annouck Luyten, Kodandaramireddy Nalapareddy, Henry Long, Michael Verzi and Ramesh a Shivdasani: *Broadly permissive intestinal chromatin underlies lateral inhibition and cell plasticity*. Nature, 506(7489):511–515, feb 2014.

[223] Jadhav, Unmesh, Kodandaramireddy Nalapareddy, Madhurima Saxena, Nicholas K. O'Neill, Luca Pinello, Guo-Cheng Yuan, Stuart H. Orkin and Ramesh A. Shivdasani: *Acquired Tissue-Specific Promoter Bivalency Is a Basis for PRC2 Necessity in Adult Cells*. Cell, 165(6):1389–1400, jun 2016.

[224] Jadhav, Unmesh, Madhurima Saxena, Nicholas K. O'Neill, Assieh Saadatpour, Guo-Cheng Yuan, Zachary Herbert, Kazutaka Murata and Ramesh A. Shivdasani: *Dynamic Reorganization of Chromatin Accessibility Signatures during Dedifferentiation of Secretory Precursors into Lgr5+ Intestinal Stem Cells*. Cell Stem Cell, 21(1):65–77.e5, jul 2017.

[225] Kaaij, Lucas T.J., Marc van de Wetering, Fang Fang, Benjamin Decato, Antoine Molaro, Harmen J.G. van de Werken, Johan H. van Es, Jurian Schuijers, Elzo de Wit, Wouter de Laat, Gregory J. Hannon, Hans C. Clevers, Andrew D. Smith and René F. Ketting: *DNA methylation dynamics during intestinal stem cell differentiation reveals enhancers driving gene expression in the villus*. Genome Biology, 14(5):R50, 2013.

[226] Yu, Da-Hai, Manasi Gadkari, Quan Zhou, Shiyan Yu, Nan Gao, Yongtao Guan, Deborah Schady, Tony N. Roshan, Miao-Hsueh Chen, Eleonora Laritsky, Zhongqi Ge, Hui Wang, Rui Chen, Caroline Westwater, Lynn Bry, Robert A. Waterland, Chelsea Moriarty, Cindy Hwang, Alton G. Swennes, Sean R. Moore and Lanlan Shen: *Postnatal epigenetic regulation of intestinal stem cells requires DNA methylation and is guided by the microbiome*. Genome Biology, 16(1):211, dec 2015.

[227] Korinek, Vladimir, Nick Barker, Petra Moerer, Elly van Donselaar, Gerwin Huls, Peter J. Peters and Hans Clevers: *Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4*. Nature Genetics, 19(4):379–383, aug 1998.

[228] Es, J. H. van, A. Haegebarth, P. Kujala, S. Itzkovitz, B.-K. Koo, S. F. Boj, J. Korving, M. van den Born, A. van Oudenaarden, S. Robine and H. Clevers: *A Critical Role for the Wnt Effector Tcf4 in Adult Intestinal Homeostatic Self-Renewal*. Molecular and Cellular Biology, 32(10):1918–1927, may 2012.

[229] The Cancer Genome Atlas Network: *Comprehensive molecular characterization of human colon and rectal cancer*. Nature, 487(7407):330–337, jul 2012.

[230] Barker, Nick, Rachel A. Ridgway, Johan H. van Es, Marc van de Wetering, Harry Begthel, Maaike van den Born, Esther Danenberg, Alan R. Clarke, Owen J. Sansom and Hans Clevers: *Crypt stem cells as the cells-of-origin of intestinal cancer*. Nature, 457(7229):608–611, jan 2009.

[231] Moser, A.R, C. Luongo, K.A Gould, M.K McNeley, A.R Shoemaker and W.F Dove: *ApcMin: A mouse model for intestinal and mammary tumorigenesis*. European Journal of Cancer, 31(7-8):1061–1064, jul 1995.

[232] Alexandre, Cyrille, Alberto Baena-Lopez and Jean-Paul Vincent: *Patterning and growth control by membrane-tethered Wingless*. Nature, 505(7482):180–185, jan 2014.

[233] Farin, Henner F., Ingrid Jordens, Mohammed H. Mosa, Onur Basak, Jeroen Korving, Daniele V. F. Tauriello, Karin de Punder, Stephane Angers, Peter J. Peters, Madelon M. Maurice and Hans

Clevers: *Visualization of a short-range Wnt gradient in the intestinal stem-cell niche*. Nature, 530(7590):340–343, feb 2016.

[234] Lau, Wim de, Nick Barker, Teck Y. Low, Bon-Kyoung Koo, Vivian S W Li, Hans Teunissen, Pekka Kujala, Andrea Haegebarth, Peter J. Peters, Marc van de Wetering, D. E. Stange, J. van Es, Daniele Guardavaccaro, Richard B M Schasfoort, Yasuaki Mohri, Katsuhiko Nishimori, Shabaz Mohammed, Albert J R Heck and Hans Clevers: *Lgr5 homologues associate with Wnt receptors and mediate R-spondin signalling*. Nature, 476(7360):293–297, aug 2011.

[235] Glinka, Andrei, Christine Dolde, Nadine Kirsch, Ya-Lin Huang, Olga Kazanskaya, Dierk Ingelfinger, Michael Boutros, Cristina-Maria Cruciat and Christof Niehrs: *LGR4 and LGR5 are R-spondin receptors mediating Wnt/β-catenin and Wnt/PCP signalling*. EMBO reports, 12(10):1055–1061, sep 2011.

[236] Carmon, K. S., X. Gong, Q. Lin, A. Thomas and Q. Liu: *R-spondins function as ligands of the orphan receptors LGR4 and LGR5 to regulate Wnt/ -catenin signaling*. Proceedings of the National Academy of Sciences, 108(28):11452–11457, jul 2011.

[237] Koo, Bon-Kyoung, Maureen Spit, Ingrid Jordens, Teck Y. Low, Daniel E. Stange, Marc van de Wetering, Johan H. van Es, Shabaz Mohammed, Albert J R Heck, Madelon M. Maurice and Hans Clevers: *Tumour suppressor RNF43 is a stem-cell E3 ligase that induces endocytosis of Wnt receptors*. Nature, 488(7413):665–669, aug 2012.

[238] Hao, Huai-Xiang, Yang Xie, Yue Zhang, Olga Charlat, Emma Oster, Monika Avello, Hong Lei, Craig Mickanin, Dong Liu, Heinz Ruffner, Xiaohong Mao, Qicheng Ma, Raffaella Zamponi, Tewis Bouwmeester, Peter M. Finan, Marc W. Kirschner, Jeffery A. Porter, Fabrizio C. Serluca and Feng Cong: *ZNRF3 promotes Wnt receptor turnover in an R-spondin-sensitive manner*. Nature, 485(7397):195–200, apr 2012.

[239] Wong, Vivian W. Y., Daniel E. Stange, Mahalia E. Page, Simon Buczacki, Agnieszka Wabik, Satoshi Itami, Marc van de Wetering, Richard Poulsom, Nicholas A. Wright, Matthew W. B. Trotter, Fiona M. Watt, Doug J. Winton, Hans Clevers and Kim B. Jensen: *Lrig1 controls intestinal stem-cell homeostasis by negative regulation of ErbB signalling*. Nature Cell Biology, 14(4):401–408, apr 2012.

[240] Basak, Onur, Joep Beumer, Kay Wiebrands, Hiroshi Seno, Alexander van Oudenaarden and Hans Clevers: *Induced Quiescence of Lgr5+ Stem Cells in Intestinal Organoids Enables Differentiation of Hormone-Producing Enteroendocrine Cells*. Cell Stem Cell, 20(2):177–190.e4, 2017.

[241] Sancho, R., C. A. Cremona and A. Behrens: *Stem cell and progenitor fate in the mammalian intestine: Notch and lateral inhibition in homeostasis and disease*. EMBO reports, 16(5):571–581, may 2015.

[242] Tian, Hua, Brian Biehs, Cecilia Chiu, Christian W. Siebel, Yan Wu, Mike Costa, Frederic J. de Sauvage and Ophir D. Klein: *Opposing Activities of Notch and Wnt Signaling Regulate Intestinal Stem Cells and Gut Homeostasis*. Cell Reports, 11(1):33–42, apr 2015.

[243] Es, Johan H. van, Marielle E. van Gijn, Orbicia Riccio, Maaike van den Born, Marc Vooijs, Harry Begthel, Miranda Cozijnsen, Sylvie Robine, Doug J. Winton, Freddy Radtke and Hans Clevers: *Notch/γ-secretase inhibition turns proliferative cells in intestinal crypts and adenomas into goblet cells*. Nature, 435(7044):959–963, jun 2005.

[244] Milano, Joseph, Jenny McKay, Claude Dagenais, Linda Foster-Brown, Francois Pognan, Reto Gadient, Robert T. Jacobs, Anna Zacco, Barry Greenberg and Paul J. Ciaccio: *Modulation of Notch Processing by γ-Secretase Inhibitors Causes Intestinal Goblet Cell Metaplasia and Induction of Genes Known to Specify Gut Secretory Lineage Differentiation*. Toxicological Sciences, 82(1):341–358, nov 2004.

[245] VanDussen, Kelli L. and Linda C. Samuelson: *Mouse atonal homolog 1 directs intestinal progenitors to secretory cell rather than absorptive cell fate*. Developmental Biology, 346(2):215–223, oct 2010.

[246] Shroyer, Noah F., Michael A. Helmrath, Vincent Y.C. Wang, Barbara Antalffy, Susan J. Henning and Huda Y. Zoghbi: *Intestine-Specific Ablation of Mouse atonal homolog 1 (Math1) Reveals a Role in Cellular Homeostasis.* Gastroenterology, 132(7):2478–2488, jun 2007.

[247] Li, Hui Joyce, Archana Kapoor, Maryann Giel-Moloney, Guido Rindi and Andrew B. Leiter: *Notch signaling differentially regulates the cell fate of early endocrine precursor cells and their maturing descendants in the mouse pancreas and intestine.* Developmental Biology, 371(2):156–169, nov 2012.

[248] Massagué, Joan: *TGFβ signalling in context.* Nature Reviews Molecular Cell Biology, 13(10):616–630, oct 2012.

[249] Hardwick, James C.H., Gijs R. Van Den Brink, Sylvia A. Bleuming, Isabel Ballester, Jan.M.H. Van Den Brande, Josbert J. Keller, G.Johan A. Offerhaus, Sander J.H. Van Deventer and Maikel P. Peppelenbosch: *Bone morphogenetic protein 2 is expressed by, and acts upon, mature epithelial cells in the colon.* Gastroenterology, 126(1):111–121, jan 2004.

[250] Haramis, A.-P. G.: *De Novo Crypt Formation and Juvenile Polyposis on BMP Inhibition in Mouse Intestine.* Science, 303(5664):1684–1686, mar 2004.

[251] He, Xi C, Jiwang Zhang, Wei-Gang Tong, Ossama Tawfik, Jason Ross, David H. Scoville, Qiang Tian, Xin Zeng, Xi He, Leanne M. Wiedemann, Yuji Mishina and Linheng Li: *BMP signaling inhibits intestinal stem cell self-renewal through suppression of Wnt–β-catenin signaling.* Nature Genetics, 36(10):1117–1121, oct 2004.

[252] Kosinski, C., V. S. W. Li, A. S. Y. Chan, J. Zhang, C. Ho, W. Y. Tsui, T. L. Chan, R. C. Mifflin, D. W. Powell, S. T. Yuen, S. Y. Leung and X. Chen: *Gene expression patterns of human colon tops and basal crypts and BMP antagonists as intestinal stem cell niche factors.* Proceedings of the National Academy of Sciences, 104(39):15418–15423, sep 2007.

[253] Davis, Hayley, Shazia Irshad, Mukesh Bansal, Hannah Rafferty, Tatjana Boitsova, Chiara Bardella, Emma Jaeger, Annabelle Lewis, Luke Freeman-Mills, Francesc C. Giner, Pedro Rodenas-Cuadrado, Sreelakshmi Mallappa, Susan Clark, Huw Thomas, Rosemary Jeffery, Richard Poulsom, Manuel Rodriguez-Justo, Marco Novelli, Runjan Chetty, Andrew Silver, Owen J. Sansom, Florian R. Greten, Lai Mun Wang, James E. East, Ian Tomlinson and Simon J. Leedham: *Aberrant epithelial GREM1 expression initiates colonic tumorigenesis from cells outside the stem cell niche.* Nature Medicine, 21(1):62–70, jan 2015.

[254] Bjerknes, Matthew, Cyrus Khandanpour, Tarik Möröy, Tomoyuki Fujiyama, Mikio Hoshino, Tiemo J. Klisch, Qian Ding, Lin Gan, Jiafang Wang, Martín G. Martín and Hazel Cheng: *Origin of the brush cell lineage in the mouse intestinal epithelium.* Developmental Biology, 362(2):194–218, feb 2012.

[255] Yang, Q.: *Requirement of Math1 for Secretory Cell Lineage Commitment in the Mouse Intestine.* Science, 294(5549):2155–2158, dec 2001.

[256] Knoop, Kathryn A, Nachiket Kumar, Betsy R Butler, S. K. Sakthivel, Rebekah T Taylor, T. Nochi, Hisaya Akiba, Hideo Yagita, Hiroshi Kiyono and I. R. Williams: *RANKL Is Necessary and Sufficient to Initiate Development of Antigen-Sampling M Cells in the Intestinal Epithelium.* The Journal of Immunology, 183(9):5738–5747, nov 2009.

[257] Lau, Wim de, Pekka Kujala, Kerstin Schneeberger, Sabine Middendorp, Vivian S W Li, Nick Barker, Anton Martens, Frans Hofhuis, Rodney P DeKoter, Peter J Peters, Edward Nieuwenhuis and Hans Clevers: *Peyer's patch M cells derived from Lgr5(+) stem cells require SpiB and are induced by RankL in cultured "miniguts".* Molecular and cellular biology, 32(18):3639–47, sep 2012.

[258] Kanaya, Takashi, Koji Hase, Daisuke Takahashi, Shinji Fukuda, Katsuaki Hoshino, Izumi Sasaki, Hiroaki Hemmi, Kathryn A. Knoop, Nachiket Kumar, Mayuko Sato, Tatsuro Katsuno, Osamu Yokosuka, Kiminori Toyooka, Kumiko Nakai, Ayako Sakamoto, Yuuki Kitahara, Toshi Jinnohara, Stephen J McSorley, Tsuneyasu Kaisho, Ifor R. Williams and Hiroshi Ohno: *The Ets transcription factor Spi-B is essential for the differentiation of intestinal microfold cells.* Nature Immunology, 13(8):729–736, aug 2012.

[259] Gracz, Adam D., Leigh Ann Samsa, Matthew J. Fordham, Danny C. Trotier, Bailey Zwarycz, Yuan-Hung Lo, Katherine Bao, Joshua Starmer, Jesse R. Raab, Noah F. Shroyer, R. Lee Reinhardt and Scott T. Magness: *Sox4 Promotes Atoh1-Independent Intestinal Secretory Differentiation Toward Tuft and Enteroendocrine Fates*. Gastroenterology, 155(5):1508–1523.e10, nov 2018.

[260] Heuberger, Julian, Frauke Kosel, Jingjing Qi, Katja S. Grossmann, Klaus Rajewsky and Walter Birchmeier: *Shp2/MAPK signaling controls goblet/paneth cell fate decisions in the intestine*. Proceedings of the National Academy of Sciences, 111(9):3472–3477, mar 2014.

[261] Noah, Taeko K., Avedis Kazanjian, Jeffrey Whitsett and Noah F. Shroyer: *SAM pointed domain ETS factor (SPDEF) regulates terminal differentiation and maturation of intestinal goblet cells*. Experimental Cell Research, 316(3):452–465, feb 2010.

[262] Gregorieff, Alex, Daniel E. Stange, Pekka Kujala, Harry Begthel, Maaike van den Born, Jeroen Korving, Peter J. Peters and Hans Clevers: *The Ets-Domain Transcription Factor Spdef Promotes Maturation of Goblet and Paneth Cells in the Intestinal Epithelium*. Gastroenterology, 137(4):1333–1345.e3, oct 2009.

[263] Es, Johan H. van, Philippe Jay, Alex Gregorieff, Marielle E. van Gijn, Suzanne Jonkheer, Pantelis Hatzis, Andrea Thiele, Maaike van den Born, Harry Begthel, Thomas Brabletz, Makoto M. Taketo and Hans Clevers: *Wnt signalling induces maturation of Paneth cells in intestinal crypts*. Nature Cell Biology, 7(4):381–386, apr 2005.

[264] Bastide, Pauline, Charbel Darido, Julie Pannequin, Ralf Kist, Sylvie Robine, Christiane Marty-Double, Frédéric Bibeau, Gerd Scherer, Dominique Joubert, Frédéric Hollande, Philippe Blache and Philippe Jay: *Sox9 regulates cell proliferation and is required for Paneth cell differentiation in the intestinal epithelium*. The Journal of Cell Biology, 178(4):635–648, aug 2007.

[265] Mori–Akiyama, Yuko, Maaike van den Born, Johan H. van Es, Stanley R. Hamilton, Henry P. Adams, Jiexin Zhang, Hans Clevers and Benoit de Crombrugghe: *SOX9 Is Required for the Differentiation of Paneth Cells in the Intestinal Epithelium*. Gastroenterology, 133(2):539–546, aug 2007.

[266] Jenny, Marjorie: *Neurogenin3 is differentially required for endocrine cell fate specification in the intestinal and gastric epithelium*. The EMBO Journal, 21(23):6338–6347, dec 2002.

[267] López-Díaz, Lymari, Renu N. Jain, Theresa M. Keeley, Kelli L. VanDussen, Cynthia S. Brunkan, Deborah L. Gumucio and Linda C. Samuelson: *Intestinal Neurogenin 3 directs differentiation of a bipotential secretory progenitor to endocrine cell rather than goblet cell fate*. Developmental Biology, 309(2):298–305, sep 2007.

[268] Beucher, Anthony, Elisabet Gjernes, Caitlin Collin, Monica Courtney, Aline Meunier, Patrick Collombat and Gérard Gradwohl: *The Homeodomain-Containing Transcription Factors Arx and Pax4 Control Enteroendocrine Subtype Specification in Mice*. PLoS ONE, 7(5):e36449, may 2012.

[269] Li, H. J., S. K. Ray, N. K. Singh, B. Johnston and A. B. Leiter: *Basic helix-loop-helix transcription factors and enteroendocrine cell differentiation*. Diabetes, Obesity and Metabolism, 13:5–12, oct 2011.

[270] May, Catherine Lee and Klaus H. Kaestner: *Gut endocrine cell development*. Molecular and Cellular Endocrinology, 323(1):70–75, jul 2010.

[271] Larsson, Lars-Inge, Luc St-Onge, David M Hougaard, Beatriz Sosa-Pineda and Peter Gruss: *Pax 4 and 6 regulate gastrointestinal endocrine cell development*. Mechanisms of Development, 79(1-2):153–159, dec 1998.

[272] Ye, Diana Z and Klaus H Kaestner: *Foxa1 and Foxa2 control the differentiation of goblet and enteroendocrine L- and D-cells in mice.* Gastroenterology, 137(6):2052–62, dec 2009.

[273] Stamataki, Despina, Maxine Holder, Christine Hodgetts, Rosemary Jeffery, Emma Nye, Bradley Spencer-Dene, Douglas J. Winton and Julian Lewis: *Delta1 Expression, Cell Cycle Exit, and Commitment to a Specific Secretory Fate Coincide within a Few Hours in the Mouse Intestinal Stem Cell System*. PLoS ONE, 6(9):e24484, sep 2011.

[274] Clevers, Hans: *The Intestinal Crypt, A Prototype Stem Cell Compartment*. Cell, 154(2):274–284, jul 2013.

[275] Bjerknes, Matthew and Hazel Cheng: *Cell Lineage metastability in Gfi1-deficient mouse intestinal epithelium*. Developmental Biology, 345(1):49–63, sep 2010.

[276] Grün, Dominic, Anna Lyubimova, Lennart Kester, Kay Wiebrands, Onur Basak, Nobuo Sasaki, Hans Clevers and Alexander van Oudenaarden: *Single-cell messenger RNA sequencing reveals rare intestinal cell types*. Nature, 525(7568):251–5, 2015.

[277] Grün, Dominic, Mauro J. Muraro, Jean-charles Boisset, Kay Wiebrands, Anna Lyubimova, Gitanjali Dharmadhikari, Maaike van den Born, Johan van Es, Erik Jansen, Hans Clevers, Eelco J.P. de Koning and Alexander van Oudenaarden: *De Novo Prediction of Stem Cell Identity using Single-Cell Transcriptome Data*. Cell Stem Cell, 19(2):266–277, aug 2016.

[278] Kim, Tae-Hee, Assieh Saadatpour, Guoji Guo, Madhurima Saxena, Alessia Cavazza, Niyati Desai, Unmesh Jadhav, Lan Jiang, Miguel N. Rivera, Stuart H. Orkin, Guo-Cheng Yuan and Ramesh A. Shivdasani: *Single-Cell Transcript Profiles Reveal Multilineage Priming in Early Progenitors Derived from Lgr5 + Intestinal Stem Cells*. Cell Reports, 16(8):2053–2060, aug 2016.

[279] Böttcher, Anika, Maren Büttner, Sophie Tritschler, Michael Sterr, Alexandra Aliluev, Lena Oppenländer, Ingo Burtscher, Steffen Sass, Martin Irmler, Johannes Beckers, Christoph Ziegenhain, Wolfgang Enard, Andrea C. Schamberger, Fien M. Verhamme, Oliver Eickelberg, Fabian J. Theis and Heiko Lickert: *Wnt/PCP-primed intestinal stem cells directly differentiate into enteroendocrine or Paneth cells*. Nature Cell Biology (in revision).

[280] Golson, Maria L. and Klaus H. Kaestner: *Fox transcription factors: from development to disease*. Development, 143(24):4558–4570, dec 2016.

[281] Weigel, Detlef, Gerd Jürgens, Frank Küttner, Eveline Seifert and Herbert Jäckle: *The homeotic gene fork head encodes a nuclear protein and is expressed in the terminal regions of the Drosophila embryo*. Cell, 57(4):645–658, may 1989.

[282] Pohl, Barbara S. and Walter Knöchel: *Of Fox and Frogs: Fox (fork head/winged helix) transcription factors in Xenopus development*. Gene, 344:21–32, jan 2005.

[283] Hannenhalli, Sridhar and Klaus H Kaestner: *The evolution of Fox genes and their role in development and disease.* Nature reviews. Genetics, 10(4):233–40, apr 2009.

[284] Kaestner, K. H., J. Katz, Y. Liu, D. J. Drucker and G. Schutz: *Inactivation of the winged helix transcription factor HNF3alpha affects glucose homeostasis and islet glucagon gene expression in vivo*. Genes & Development, 13(4):495–504, feb 1999.

[285] Clark, Kirk L., Elaine D Halay, Eseng Lai and Stephen K Burley: *Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5*. Nature, 364(6436):412–420, jul 1993.

[286] Li, Jun, Ana Carolina Dantas Machado, Ming Guo, Jared M. Sagendorf, Zhan Zhou, Longying Jiang, Xiaojuan Chen, Daichao Wu, Lingzhi Qu, Zhuchu Chen, Lin Chen, Remo Rohs and Yongheng Chen: *Structure of the Forkhead Domain of FOXA2 Bound to a Complete DNA Consensus Site*. Biochemistry, 56(29):3745–3753, 2017.

[287] Costa, R H, D R Grayson and J E Darnell: *Multiple hepatocyte-enriched nuclear factors function in the regulation of transthyretin and alpha 1-antitrypsin genes.* Molecular and Cellular Biology, 9(4):1415–1425, apr 1989.

[288] Friedman, J R and K H Kaestner: *The Foxa family of transcription factors in development and metabolism.* Cellular and molecular life sciences : CMLS, 63(19-20):2317–28, oct 2006.

[289] Cirillo, Lisa Ann, Frank Robert Lin, Isabel Cuesta, Dara Friedman, Michal Jarnik and Kenneth S Zaret: *Opening of Compacted Chromatin by Early Developmental Transcription Factors HNF3 (FoxA) and GATA-4*. Molecular Cell, 9(2):279–289, feb 2002.

[290] Li, Zhaoyu, Paul Gadue, Kaifu Chen, Yang Jiao, Geetu Tuteja, Jonathan Schug, Wei Li and Klaus H. Kaestner: *Foxa2 and H2A.Z Mediate Nucleosome Depletion during Embryonic Stem Cell Differentiation*. Cell, 151(7):1608–1616, dec 2012.

[291] Updike, Dustin L. and Susan E. Mango: *Temporal Regulation of Foregut Development by HTZ-1/H2A.Z and PHA-4/FoxA*. PLoS Genetics, 2(9):e161, 2006.

[292] Iwafuchi-Doi, Makiko, Greg Donahue, Akshay Kakumanu, Jason A. Watts, Shaun Mahony, B. Franklin Pugh, Dolim Lee, Klaus H. Kaestner and Kenneth S. Zaret: *The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation*. Molecular Cell, 62(1):79–91, 2016.

[293] Sund, N. J., S.-L. Ang, S. D. Sackett, W. Shen, N. Daigle, M. A. Magnuson and K. H. Kaestner: *Hepatocyte Nuclear Factor 3beta (Foxa2) Is Dispensable for Maintaining the Differentiated State of the Adult Hepatocyte*. Molecular and Cellular Biology, 20(14):5175–5183, jul 2000.

[294] Wan, Huajing, Klaus H Kaestner, Siew-Lan Ang, Machiko Ikegami, Fred D Finkelman, Mildred T Stahlman, Patricia C Fulkerson, Marc E Rothenberg and Jeffrey A Whitsett: *Foxa2 regulates alveolarization and goblet cell hyperplasia.* Development (Cambridge, England), 131(4):953–64, feb 2004.

[295] Lee, Catherine S., Joshua R. Friedman, James T. Fulmer and Klaus H. Kaestner: *The initiation of liver development is dependent on Foxa transcription factors.* Nature, 2005.

[296] Burtscher, Ingo, Wenke Barkey and Heiko Lickert: *Foxa2-venus fusion reporter mouse line allows live-cell analysis of endoderm-derived organ formation*. genesis, 51(8):596–604, aug 2013.

[297] Wang, Xianming, Michael Sterr, Ingo Burtscher, Shen Chen, Anja Hieronimus, Fausto Machicao, Harald Staiger, Hans-Ulrich Häring, Gabriele Lederer, Thomas Meitinger, Filippo M. Cernilogar, Gunnar Schotta, Martin Irmler, Johannes Beckers, Martin Hrabě de Angelis, Michael Ray, Christopher V.E. Wright, Mostafa Bakhti and Heiko Lickert: *Genome-wide analysis of PDX1 target genes in human pancreatic progenitors*. Molecular Metabolism, 9:57–68, mar 2018.

[298] Wang, Xianming, Shen Chen, Ingo Burtscher, Michael Sterr, Anja Hieronimus, Fausto Machicao, Harald Staiger, Hans-Ulrich Häring, Gabriele Lederer, Thomas Meitinger and Heiko Lickert: *Generation of a human induced pluripotent stem cell (iPSC) line from a patient with family history of diabetes carrying a C18R mutation in the PDX1 gene*. Stem Cell Research, 17(2):292–295, sep 2016.

[299] Wang, Xianming, Shen Chen, Ingo Burtscher, Michael Sterr, Anja Hieronimus, Fausto Machicao, Harald Staiger, Hans Ulrich Häring, Gabriele Lederer, Thomas Meitinger and Heiko Lickert: *Generation of a human induced pluripotent stem cell (iPSC) line from a patient carrying a P33T mutation in the PDX1 gene*. Stem Cell Research, 17(2):273–276, 2016.

[300] Stefan, Norbert: *Identification and Characterization of Metabolically Benign Obesity in Humans*. Archives of Internal Medicine, 168(15):1609, aug 2008.

[301] Wang, Xianming, Michael Sterr, Ansarullah, Ingo Burtscher, Anika Böttcher, Julia Beckenbauer, Johanna Siehler, Thomas Meitinger, Hans-Ulrich Häring, Harald Staiger, Filippo M. Cernilogar, Gunnar Schotta, Martin Irmler, Johannes Beckers, Christopher V.E. Wright, Mostafa Bakhti and Heiko Lickert: *Point mutations in the PDX1 transactivation domain impair human β-cell development and function*. Molecular Metabolism, mar 2019.

[302] Wang, Allen, Feng Yue, Yan Li, Ruiyu Xie, Thomas Harper, Nisha A. Patel, Kayla Muth, Jeffrey Palmer, Yunjiang Qiu, Jinzhao Wang, Dieter K. Lam, Jeffrey C. Raum, Doris A. Stoffers, Bing Ren and Maike Sander: *Epigenetic Priming of Enhancers Predicts Developmental Competence of hESC-Derived Endodermal Lineage Intermediates*. Cell Stem Cell, 16(4):386–399, apr 2015.

[303] Heintzman, Nathaniel D., Gary C. Hon, R. David Hawkins, Pouya Kheradpour, Alexander Stark, Lindsey F. Harp, Zhen Ye, Leonard K. Lee, Rhona K. Stuart, Christina W. Ching, Keith A. Ching,

Jessica E. Antosiewicz-Bourget, Hui Liu, Xinmin Zhang, Roland D. Green, Victor V. Lobanenkov, Ron Stewart, James A. Thomson, Gregory E. Crawford, Manolis Kellis and Bing Ren: *Histone modifications at human enhancers reflect global cell-type-specific gene expression*. Nature, 459(7243):108–112, may 2009.

[304] Creyghton, M. P., A. W. Cheng, G. G. Welstead, T. Kooistra, B. W. Carey, E. J. Steine, J. Hanna, M. A. Lodato, G. M. Frampton, P. A. Sharp, L. A. Boyer, R. A. Young and R. Jaenisch: *Histone H3K27ac separates active from poised enhancers and predicts developmental state*. Proceedings of the National Academy of Sciences, 107(50):21931–21936, 2010.

[305] Zentner, Gabriel E., Paul J. Tesar and Peter C. Scacheri: *Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions*. Genome Research, 21(8):1273–1283, aug 2011.

[306] Bonn, Stefan, Robert P Zinzen, Charles Girardot, E Hilary Gustafson, Alexis Perez-Gonzalez, Nicolas Delhomme, Yad Ghavi-Helm, Bartek Wilczyński, Andrew Riddell and Eileen E M Furlong: *Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development.* Nature genetics, 44(2):148–56, feb 2012.

[307] Calo, Eliezer and Joanna Wysocka: *Modification of Enhancer Chromatin: What, How, and Why?* Molecular Cell, 49(5):825–837, 2013.

[308] Ait-Lounis, A., Claire Bonal, Q. Seguin-Estevez, Christoph D Schmid, Philipp Bucher, Pedro L Herrera, B. Durand, Paolo Meda and Walter Reith: *The Transcription Factor Rfx3 Regulates -Cell Differentiation, Function, and Glucokinase Expression*. Diabetes, 59(7):1674–1685, jul 2010.

[309] Kim, Wook, Yu-Kyong Shin, Byung-Joon Kim and Josephine M Egan: *Notch signaling in pancreatic endocrine cell and diabetes.* Biochemical and biophysical research communications, 392(3):247–51, 2010.

[310] Cebola, Inês, Santiago a. Rodríguez-Seguí, Candy H-H. Cho, José Bessa, Meritxell Rovira, Mario Luengo, Mariya Chhatriwala, Andrew Berry, Joan Ponsa-Cobas, Miguel Angel Maestro, Rachel E. Jennings, Lorenzo Pasquali, Ignasi Morán, Natalia Castro, Neil a. Hanley, Jose Luis Gomez-Skarmeta, Ludovic Vallier and Jorge Ferrer: *TEAD and YAP regulate the enhancer network of human embryonic pancreatic progenitors*. Nature Cell Biology, 17(5):615–626, 2015.

[311] Teo, Adrian Kee Keong, Norihiro Tsuneyoshi, Shawn Hoon, Ee Kim Tan, Lawrence W. Stanton, Christopher V.E. Wright and N. Ray Dunn: *PDX1 Binds and Represses Hepatic Genes to Ensure Robust Pancreatic Commitment in Differentiating Human Embryonic Stem Cells*. Stem Cell Reports, 4(4):578–590, apr 2015.

[312] Khoo, Cynthia, Juxiang Yang, Samuel A Weinrott, Klaus H Kaestner, Ali Naji, Jonathan Schug and Doris A Stoffers: *Research resource: the pdx1 cistrome of pancreatic islets.* Molecular endocrinology (Baltimore, Md.), 26(3):521–33, 2012.

[313] Pasquali, Lorenzo, Kyle J Gaulton, Santiago A Rodríguez-Seguí, Loris Mularoni, Irene Miguel-Escalada, İldem Akerman, Juan J Tena, Ignasi Morán, Carlos Gómez-Marín, Martijn van de Bunt, Joan Ponsa-Cobas, Natalia Castro, Takao Nammo, Inês Cebola, Javier García-Hurtado, Miguel Angel Maestro, François Pattou, Lorenzo Piemonti, Thierry Berney, Anna L Gloyn, Philippe Ravassard, José Luis Gómez Skarmeta, Ferenc Müller, Mark I McCarthy and Jorge Ferrer: *Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants*. Nature Genetics, 46(2):136–143, 2014.

[314] MacArthur, Jacqueline, Emily Bowler, Maria Cerezo, Laurent Gil, Peggy Hall, Emma Hastings, Heather Junkins, Aoife McMahon, Annalisa Milano, Joannella Morales, Zoe MayPendlington, Danielle Welter, Tony Burdett, Lucia Hindorff, Paul Flicek, Fiona Cunningham and Helen Parkinson: *The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog)*. Nucleic Acids Research, 45(D1):D896–D901, 2017.

[315] Corradin, Olivia and Peter C Scacheri: *Enhancer variants: evaluating functions in common disease*. Genome Medicine, 6(10):85, 2014.

[316] Scott, Robert A, Laura J Scott, Reedik Mägi, Letizia Marullo, Kyle J Gaulton, Marika Kaakinen, Natalia Pervjakova, Tune H Pers, Andrew D Johnson, John D Eicher, Anne U Jackson, Teresa Ferreira, Yeji Lee, Clement Ma, Valgerdur Steinthorsdottir, Gudmar Thorleifsson, Lu Qi, Natalie R Van Zuydam, Anubha Mahajan, Han Chen, Peter Almgren, Ben F Voight, Harald Grallert, Martina Müller-Nurasyid, Janina S Ried, William N Rayner, Neil Robertson, Lennart C Karssen, Elisabeth M van Leeuwen, Sara M Willems, Christian Fuchsberger, Phoenix Kwan, Tanya M Teslovich, Pritam Chanda, Man Li, Yingchang Lu, Christian Dina, Dorothee Thuillier, Loic Yengo, Longda Jiang, Thomas Sparso, Hans A Kestler, Himanshu Chheda, Lewin Eisele, Stefan Gustafsson, Mattias Frånberg, Rona J Strawbridge, Rafn Benediktsson, Astradur B Hreidarsson, Augustine Kong, Gunnar Sigurðsson, Nicola D Kerrison, Jian'an Luan, Liming Liang, Thomas Meitinger, Michael Roden, Barbara Thorand, Tõnu Esko, Evelin Mihailov, Caroline Fox, Ching-Ti Liu, Denis Rybin, Bo Isomaa, Valeriya Lyssenko, Tiinamaija Tuomi, David J Couper, James S Pankow, Niels Grarup, Christian T Have, Marit E Jørgensen, Torben Jørgensen, Allan Linneberg, Marilyn C Cornelis, Rob M van Dam, David J Hunter, Peter Kraft, Qi Sun, Sarah Edkins, Katharine R Owen, John RB Perry, Andrew R Wood, Eleftheria Zeggini, Juan Tajes-Fernandes, Goncalo R Abecasis, Lori L Bonnycastle, Peter S Chines, Heather M Stringham, Heikki A Koistinen, Leena Kinnunen, Bengt Sennblad, Thomas W Mühleisen, Markus M Nöthen, Sonali Pechlivanis, Damiano Baldassarre, Karl Gertow, Steve E Humphries, Elena Tremoli, Norman Klopp, Julia Meyer, Gerald Steinbach, Roman Wennauer, Johan G Eriksson, Satu M⊠nnistö, Leena Peltonen, Emmi Tikkanen, Guillaume Charpentier, Elodie Eury, Stéphane Lobbens, Bruna Gigante, Karin Leander, Olga McLeod, Erwin P Bottinger, Omri Gottesman, Douglas Ruderfer, Matthias Blüher, Peter Kovacs, Anke Tonjes, Nisa M Maruthur, Chiara Scapoli, Raimund Erbel, Karl-Heinz Jöckel, Susanne Moebus, Ulf de Faire, Anders Hamsten, Michael Stumvoll, Panagiotis Deloukas, Peter J Donnelly, Timothy M Frayling, Andrew T Hattersley, Samuli Ripatti, Veikko Salomaa, Nancy L Pedersen, Bernhard O Boehm, Richard N Bergman, Francis S Collins, Karen L Mohlke, Jaakko Tuomilehto, Torben Hansen, Oluf Pedersen, Inês Barroso, Lars Lannfelt, Erik Ingelsson, Lars Lind, Cecilia M Lindgren, Stephane Cauchi, Philippe Froguel, Ruth JF Loos, Beverley Balkau, Heiner Boeing, Paul W Franks, Aurelio Barricarte Gurrea, Domenico Palli, Yvonne T van der Schouw, David Altshuler, Leif C Groop, Claudia Langenberg, Nicholas J Wareham, Eric Sijbrands, Cornelia M van Duijn, Jose C Florez, James B Meigs, Eric Boerwinkle, Christian Gieger, Konstantin Strauch, Andres Metspalu, Andrew D Morris, Colin NA Palmer, Frank B Hu, Unnur Thorsteinsdottir, Kari Stefansson, Josée Dupuis, Andrew P Morris, Michael Boehnke, Mark I McCarthy and Inga Prokopenko: *An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans*. Diabetes, page db161253, 2017.

[317] Heuvel-Borsboom, H., H. W. de Valk, M. Losekoot and J. Westerink: *Maturity onset diabetes of the young: Seek and you will find*, 2016.

[318] Flanagan, Sarah E., Elisa De Franco, Hana Lango Allen, Michele Zerah, Majedah M. Abdul-Rasoul, Julie A. Edge, Helen Stewart, Elham Alamiri, Khalid Hussain, Sam Wallis, Liat De Vries, Oscar Rubio-Cabezas, Jayne A.L. Houghton, Emma L. Edghill, Ann Marie Patch, Sian Ellard and Andrew T. Hattersley: *Analysis of transcription factors key for mouse pancreatic development establishes NKX2-2 and MNX1 mutations as causes of neonatal diabetes in man*. Cell Metabolism, 19(1):146–154, 2014.

[319] Chu, Khoi and M.-J. Tsai: *Neuronatin, a Downstream Target of BETA2/NeuroD1 in the Pancreas, Is Involved in Glucose-Mediated Insulin Secretion*. Diabetes, 54(4):1064–1073, apr 2005.

[320] Joe, Myung Kuk, Hyo Jung Lee, Young Ho Suh, Kyu Lee Han, Joo Hyun Lim, Jihyun Song, Je Kyung Seong and Myeong Ho Jung: *Crucial roles of neuronatin in insulin secretion and high glucose-induced apoptosis in pancreatic β-cells*. Cellular Signalling, 20(5):907–915, may 2008.

[321] Kubosaki, Atsutaka, Steffen Gross, Junnosuke Miura, Keiichi Saeki, Min Zhu, Shinichiro Nakamura, Wiljan Hendriks and Abner Louis Notkins: *Targeted Disruption of the IA-2 Gene Causes Glucose Intolerance and Impairs Insulin Secretion but Does Not Prevent the Development of Diabetes in NOD Mice*. Diabetes, 53(7):1684–1691, jul 2004.

[322] Smid, Johnathan K., Sharlene Faulkes and Michael A. Rudnicki: *Periostin induces pancreatic regeneration.* Endocrinology, 156(3):824–836, 2015.

[323] Layden, B. T., V. Durai, M. V. Newman, A. M. Marinelarena, C. W. Ahn, G. Feng, S. Lin, X. Zhang, D. B. Kaufman, N. Jafari, G. L. Sørensen and W. L. Lowe: *Regulation of pancreatic islet gene expression in mouse islets by pregnancy.* Journal of Endocrinology, 207(3):265–279, 2010.

[324] Worthington, J. J., F. Reimann and F. M. Gribble: *Enteroendocrine cells-sensory sentinels of the intestinal environment and orchestrators of mucosal immunity.* Mucosal Immunology, 11(1):3–20, 2018.

[325] Kaestner, Klaus H: *The FoxA factors in organogenesis and differentiation.* Current opinion in genetics & development, 20(5):527–32, oct 2010.

[326] Barker, Nick, Johan H van Es, Jeroen Kuipers, Pekka Kujala, Maaike van den Born, Miranda Cozijnsen, Andrea Haegebarth, Jeroen Korving, Harry Begthel, Peter J Peters and Hans Clevers: *Identification of stem cells in small intestine and colon by marker gene Lgr5.* Nature, 449(7165):1003–7, oct 2007.

[327] Imuta, Yu, Hiroshi Kiyonari, Chuan Wei Jang, Richard R. Behringer and Hiroshi Sasaki: *Generation of knock-in mice that express nuclear enhanced green fluorescent protein and tamoxifen-inducible Cre recombinase in the notochord from Foxa2 and T loci.* Genesis, 51(3):210–218, 2013.

[328] Muzumdar, Mandar Deepak, Bosiljka Tasic, Kazunari Miyamichi, Ling Li and Liqun Luo: *A global double-fluorescent Cre reporter mouse.* genesis, 45(9):593–605, sep 2007.

[329] Heintzman, Nathaniel D., Rhona K. Stuart, Gary Hon, Yutao Fu, Christina W Ching, R. David Hawkins, Leah O. Barrera, Sara Van Calcar, Chunxu Qu, Keith A. Ching, Wei Wang, Zhiping Weng, Roland D. Green, Gregory E. Crawford and Bing Ren: *Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome.* Nature Genetics, 39(3):311–318, mar 2007.

[330] Heintzman, Nathaniel D., Gary C. Hon, R. David Hawkins, Pouya Kheradpour, Alexander Stark, Lindsey F. Harp, Zhen Ye, Leonard K. Lee, Rhona K. Stuart, Christina W. Ching, Keith A. Ching, Jessica E. Antosiewicz-Bourget, Hui Liu, Xinmin Zhang, Roland D. Green, Victor V. Lobanenkov, Ron Stewart, James A. Thomson, Gregory E. Crawford, Manolis Kellis and Bing Ren: *Histone modifications at human enhancers reflect global cell-type-specific gene expression.* Nature, 459(7243):108–112, may 2009.

[331] Buenrostro, Jason D, Paul G Giresi, Lisa C Zaba, Howard Y Chang and William J Greenleaf: *Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position.* Nature methods, 10(12):1213–8, dec 2013.

[332] Schones, Dustin E., Kairong Cui, Suresh Cuddapah, Tae Young Roh, Artem Barski, Zhibin Wang, Gang Wei and Keji Zhao: *Dynamic Regulation of Nucleosome Positioning in the Human Genome.* Cell, 132(5):887–898, 2008.

[333] Love, Michael I., Wolfgang Huber and Simon Anders: *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biology, 15(12):550, dec 2014.

[334] Thurman, Robert E., Eric Rynes, Richard Humbert, Jeff Vierstra, Matthew T. Maurano, Eric Haugen, Nathan C. Sheffield, Andrew B. Stergachis, Hao Wang, Benjamin Vernot, Kavita Garg, Sam John, Richard Sandstrom, Daniel Bates, Lisa Boatman, Theresa K. Canfield, Morgan Diegel, Douglas Dunn, Abigail K. Ebersol, Tristan Frum, Erika Giste, Audra K. Johnson, Ericka M. Johnson, Tanya Kutyavin, Bryan Lajoie, Bum-Kyu Kyu Lee, Kristen Lee, Darin London, Dimitra Lotakis, Shane Neph, Fidencio Neri, Eric D. Nguyen, Hongzhu Qu, Alex P. Reynolds, Vaughn Roach, Alexias Safi, Minerva E. Sanchez, Amartya Sanyal, Anthony Shafer, Jeremy M. Simon, Lingyun Song, Shinny Vong, Molly Weaver, Yongqi Yan, Zhancheng Zhuzhu Zhang, Zhancheng Zhuzhu Zhang, Boris Lenhard, Muneesh Tewari, Michael O. Dorschner, R. Scott Hansen, Patrick A. Navas, George Stamatoyannopoulos, Vishwanath R. Iyer, Jason D. Lieb, Shamil R. Sunyaev, Joshua M. Akey, Peter J. Sabo, Rajinder Kaul, Terrence S. Furey, Job Dekker, Gregory E. Crawford and John A.

Stamatoyannopoulos: *The accessible chromatin landscape of the human genome*. Nature, 489(7414):75–82, sep 2012.

[335] Sung, Myong-hee Hee, Songjoon Baek and Gordon L. Hager: *Genome-wide footprinting: ready for prime time?* Nature Methods, 13(3):222–228, mar 2016.

[336] Schep, Alicia N., Beijing Wu, Jason D. Buenrostro and William J. Greenleaf: *ChromVAR: Inferring transcription-factor-associated accessibility from single-cell epigenomic data*. Nature Methods, 14(10):975–978, aug 2017.

[337] Bosse, T., C. M. Piaseckyj, E. Burghard, J. J. Fialkovich, S. Rajagopal, W. T. Pu and S. D. Krasinski: *Gata4 Is Essential for the Maintenance of Jejunal-Ileal Identities in the Adult Mouse Small Intestine*. Molecular and Cellular Biology, 26(23):9060–9070, dec 2006.

[338] Battle, Michele A., Benjamin J. Bondow, Moriah A. Iverson, Scott J. Adams, Ronald J. Jandacek, Patrick Tso and Stephen A. Duncan: *GATA4 Is Essential for Jejunal Function in Mice*. Gastroenterology, 135(5):1676–1686.e1, nov 2008.

[339] Beuling, E., I. M. Kerkhof, G. A. Nicksa, M. J. Giuffrida, J. Haywood, D. J. a. de Kerk, C. M. Piaseckyj, W. T. Pu, T. L. Buchmiller, P. A. Dawson and S. D. Krasinski: *Conditional Gata4 deletion in mice induces bile acid absorption in the proximal small intestine*. Gut, 59(7):888–895, jul 2010.

[340] Aronson, B.E., S. Rabello Aronson, R.P. Berkhout, S.F. Chavoushi, A. He, W.T. Pu, M.P. Verzi and S.D. Krasinski: *GATA4 represses an ileal program of gene expression in the proximal small intestine by inhibiting the acetylation of histone H3, lysine 27*. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms, 1839(11):1273–1282, nov 2014.

[341] San Roman, Adrianna K., Boaz E. Aronson, Stephen D. Krasinski, Ramesh A. Shivdasani and Michael P. Verzi: *Transcription Factors GATA4 and HNF4A Control Distinct Aspects of Intestinal Homeostasis in Conjunction with Transcription Factor CDX2*. Journal of Biological Chemistry, 290(3):1850–1860, jan 2015.

[342] Muñoz, Javier, Daniel E Stange, Arnout G Schepers, Marc van de Wetering, Bon-Kyoung Koo, Shalev Itzkovitz, Richard Volckmann, Kevin S Kung, Jan Koster, Sorina Radulescu, Kevin Myant, Rogier Versteeg, Owen J Sansom, Johan H van Es, Nick Barker, Alexander van Oudenaarden, Shabaz Mohammed, Albert J R Heck and Hans Clevers: *The Lgr5 intestinal stem cell signature: robust expression of proposed quiescent '+4' cell markers.* The EMBO journal, 31(14):3079–91, jul 2012.

[343] Iwafuchi-doi, Makiko and Kenneth S Zaret: *Cell fate control by pioneer transcription factors*. Development, 143(11):1833–1837, jun 2016.

[344] Zaret, Kenneth S and Jason S Carroll: *Pioneer transcription factors: establishing competence for gene expression.* Genes & development, 25(21):2227–41, nov 2011.

[345] Verzi, M. P., H. Shin, A. K. San Roman, X. S. Liu and R. A. Shivdasani: *Intestinal Master Transcription Factor CDX2 Controls Chromatin Access for Partner Transcription Factor Binding*. Molecular and Cellular Biology, 33(2):281–292, 2013.

[346] Shlyueva, Daria, Gerald Stampfel and Alexander Stark: *Transcriptional enhancers: from properties to genome-wide predictions*. Nature Reviews Genetics, 15(4):272–286, apr 2014.

[347] Jia, Shiqi, Andranik Ivanov, Dinko Blasevic, T. Muller, B. Purfurst, Wei Sun, Wei Chen, Matthew N. Poy, Nikolaus Rajewsky, Carmen Birchmeier, Thomas Müller, Bettina Purfürst, Wei Sun, Wei Chen, Matthew N. Poy, Nikolaus Rajewsky, Carmen Birchmeier, T. Muller, B. Purfurst, Wei Sun, Wei Chen, Matthew N. Poy, Nikolaus Rajewsky and Carmen Birchmeier: *Insm1 cooperates with Neurod1 and Foxa2 to maintain mature pancreatic -cell function*. The EMBO Journal, 34(10):1417–1433, may 2015.

[348] Hoffman, Brad G, Gordon Robertson, Bogard Zavaglia, Mike Beach, Rebecca Cullum, Sam Lee, Galina Soukhatcheva, Leping Li, Elizabeth D Wederell, Nina Thiessen, Mikhail Bilenky, Timothee

Cezard, Angela Tam, Baljit Kamoh, Inanc Birol, Derek Dai, Yongjun Zhao, Martin Hirst, C Bruce Verchere, Cheryl D Helgason, Marco a Marra, Steven J M Jones and Pamela a Hoodless: *Locus co-occupancy, nucleosome positioning, and H3K4me1 regulate the functionality of FOXA2-, HNF4A-, and PDX1-bound loci in islets and liver.* Genome research, 20(8):1037–51, aug 2010.

[349] Alder, Olivia, Rebecca Cullum, Sam Lee, Arohumam C. Kan, Wei Wei, Yuyin Yi, Victoria C. Garside, Misha Bilenky, Malachi Griffith, A. Sorana Morrissy, Gordon A. Robertson, Nina Thiessen, Yongjun Zhao, Qian Chen, Duojia Pan, Steven J.M. Jones, Marco A. Marra and Pamela A. Hoodless: *Hippo signaling influences HNF4A and FOXA2 enhancer switching during hepatocyte differentiation.* Cell Reports, 9(1):261–271, 2014.

[350] Wallerman, Ola, Mehdi Motallebipour, Stefan Enroth, Kalicharan Patra, Madhu Sudhan Reddy Bysani, Jan Komorowski and Claes Wadelius: *Molecular interactions between HNF4a, FOXA2 and GABP identified at regulatory DNA elements through ChIP-sequencing.* Nucleic acids research, 37(22):7498–508, dec 2009.

[351] Motallebipour, Mehdi, Adam Ameur, Madhu Sudhan Reddy Bysani, Kalicharan Patra, Ola Wallerman, Jonathan Mangion, Melissa a Barker, Kevin J McKernan, Jan Komorowski and Claes Wadelius: *Differential binding and co-binding pattern of FOXA1 and FOXA3 and their relation to H3K4me3 in HepG2 cells revealed by ChIP-seq.* Genome Biology, 10(11):R129, 2009.

[352] Glass, Leslie L, Fernando J. Calero-Nieto, Wajid Jawaid, Pierre Larraufie, Richard G Kay, Berthold Göttgens, Frank Reimann and Fiona M Gribble: *Single-cell RNA-sequencing reveals a distinct population of proglucagon-expressing cells specific to the mouse upper small intestine.* Molecular Metabolism, 6(10):1296–1303, oct 2017.

[353] Moor, Andreas E, Yotam Harnik, Shani Ben-Moshe, Efi E Massasa, Milena Rozenberg, Raya Eilam, Keren Bahar Halpern and Shalev Itzkovitz: *Spatial Reconstruction of Single Enterocytes Uncovers Broad Zonation along the Intestinal Villus Axis.* Cell, 175(4):1156–1167.e15, nov 2018.

[354] Aliluev, Alexandra, Sophie Tritschler, Michael Sterr, Lena Oppenländer, Julia Hinterdobler, Martin Irmler, Na Sun, Axel K. Walch, Kerstin Stemmer, Matthias H. Tschöp, Fabian J. Theis, Heiko Lickert and Anika Böttcher: *High-fat diet alters intestinal stem cell turnover and lineage allocation.* Submitted.

[355] Haghverdi, Laleh, Maren Büttner, F. Alexander Wolf, Florian Buettner and Fabian J. Theis: *Diffusion pseudotime robustly reconstructs lineage branching.* Nature Methods, 13(10):845–848, 2016.

[356] Eraslan, Gökcen, Lukas M. Simon, Maria Mircea, Nikola S. Mueller and Fabian J. Theis: *Single-cell RNA-seq denoising using a deep count autoencoder.* Nature Communications, 10(1):390, dec 2019.

[357] Ramond, Cyrille, Nicolas Glaser, Claire Berthault, Jacqueline Ameri, Jeannette Schlichting Kirkegaard, Mattias Hansson, Christian Honoré, Henrik Semb and Raphaël Scharfmann: *Reconstructing human pancreatic differentiation by mapping specific cell populations during development.* eLife, 6, 2017.

[358] Kim, Sung-Eun, Byung-Kak Kim, Jung-Eun Gil, Suel-Kee Kim and Jong-Hoon Kim: *Comparative analysis of the developmental competence of three human embryonic stem cell lines in vitro.* Molecules and cells, 23(1):49–56, feb 2007.

[359] Ohlsson, H, K Karlsson and T Edlund: *IPF1, a homeodomain-containing transactivator of the insulin gene.* The EMBO journal, 12(11):4251–9, 1993.

[360] Svensson, Per, Cecilia Williams, Joakim Lundeberg, Patrik Rydén, Ingela Bergqvist and Helena Edlund: *Gene array identification of Ipf1/Pdx1-/- regulated genes in pancreatic progenitor cells.* BMC developmental biology, 7:129, 2007.

[361] Emery, P, B Durand, B Mach and W Reith: *RFX proteins, a novel family of DNA binding proteins conserved in the eukaryotic kingdom.* Nucleic acids research, 24(5):803–7, mar 1996.

[362] Bonnafe, E., M. Touka, A. AitLounis, D. Baas, E. Barras, C. Ucla, A. Moreau, F. Flamant, R. Dubruille, P. Couble, J. Collignon, B. Durand and W. Reith: *The Transcription Factor RFX3 Directs Nodal Cilium Development and Left-Right Asymmetry Specification.* Molecular and Cellular Biology, 24(10):4417–4427, may 2004.

[363] Ait-Lounis, A., D. Baas, E. Barras, C. Benadiba, A. Charollais, R. Nlend Nlend, D. Liegeois, P. Meda, B. Durand and W. Reith: *Novel Function of the Ciliogenic Transcription Factor RFX3 in Development of the Endocrine Pancreas.* Diabetes, 56(4):950–959, apr 2007.

[364] Segerstolpe, Åsa, Athanasia Palasantza, Pernilla Eliasson, Eva-Marie Andersson, Anne-Christine Andréasson, Xiaoyan Sun, Simone Picelli, Alan Sabirsh, Maryam Clausen, Magnus K. Bjursell, David M. Smith, Maria Kasper, Carina Ämmälä and Rickard Sandberg: *Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes.* Cell Metabolism, 24(4):593–607, 2016.

[365] Xin, Yurong, Jinrang Kim, Haruka Okamoto, George D Yancopoulos, Calvin Lin and Jesper Gromada Correspondence: *RNA Sequencing of Single Human Islet Cells Reveals Type 2 Diabetes Genes.* Cell Metabolism, 24:608–615, 2016.

[366] Yumlu, Saniye, Jürgen Stumm, Sanum Bashir, Anne-Kathrin Dreyer, Pawel Lisowski, Eric Danner and Ralf Kühn: *Gene editing and clonal isolation of human induced pluripotent stem cells using CRISPR/Cas9.* Methods, 121-122:29–44, may 2017.

[367] Painter, Jodie N., Tracy A. O'Mara, Jyotsna Batra, Timothy Cheng, Felicity A. Lose, Joe Dennis, Kyriaki Michailidou, Jonathan P. Tyrer, Shahana Ahmed, Kaltin Ferguson, Catherine S. Healey, Susanne Kaufmann, Kristine M. Hillman, Carina Walpole, Leire Moya, Pamela Pollock, Angela Jones, Kimberley Howarth, Lynn Martin, Maggie Gorman, Shirley Hodgson, Ma Magdalena Echeverry De Polanco, Monica Sans, Angel Carracedo, Sergi Castellvi-Bel, Augusto Rojas-Martinez, Erika Santos, Manuel R. Teixeira, Luis Carvajal-Carmona, Xiao Ou Shu, Jirong Long, Wei Zheng, Yong Bing Xiang, Grant W. Montgomery, Penelope M. Webb, Rodney J. Scott, Mark McEvoy, John Attia, Elizabeth Holliday, Nicholas G. Martin, Dale R. Nyholt, Anjali K. Henders, Peter A. Fasching, Alexander Hein, Matthias W. Beckmann, Stefan P. Renner, Thilo Dörk, Peter Hillemanns, Matthias Dürst, Ingo Runnebaum, Diether Lambrechts, Lieve Coenegrachts, Stefanie Schrauwen, Frederic Amant, Boris Winterhoff, Sean C. Dowdy, Ellen L. Goode, Attila Teoman, Helga B. Salvesen, Jone Trovik, Tormund S. Njolstad, Henrica M.J. Werner, Katie Ashton, Tony Proietto, Geoffrey Otton, Gerasimos Tzortzatos, Miriam Mints, Emma Tham, Per Hall, Kamila Czene, Jianjun Liu, Jingmei Li, John L. Hopper, Melissa C. Southey, Arif B. Ekici, Matthias Ruebner, Nicola Johnson, Julian Peto, Barbara Burwinkel, Frederik Marme, Hermann Brenner, Aida K. Dieffenbach, Alfons Meindl, Hiltrud Brauch, Annika Lindblom, Jeroen Depreeuw, Matthieu Moisse, Jenny Chang-Claude, Anja Rudolph, Fergus J. Couch, Janet E. Olson, Graham G. Giles, Fiona Bruinsma, Julie M. Cunningham, Brooke L. Fridley, Anne Lise Børresen-Dale, Vessela N. Kristensen, Angela Cox, Anthony J. Swerdlow, Nicholas Orr, Manjeet K. Bolla, Qin Wang, Rachel Palmieri Weber, Zhihua Chen, Mitul Shah, Juliet D. French, Paul D.P. Pharoah, Alison M. Dunning, Ian Tomlinson, Douglas F. Easton, Stacey L. Edwards, Deborah J. Thompson and Amanda B. Spurdle: *Fine-mapping of the HNF1B multicancer locus identifies candidate variants that mediate endometrial cancer risk.* Human Molecular Genetics, 24(5):1478–1492, 2015.

[368] De Vas, M. G., J. L. Kopp, C. Heliot, M. Sander, S. Cereghini and C. Haumaitre: *Hnf1b controls pancreas morphogenesis and the generation of Ngn3+ endocrine progenitors.* Development, 142(5):871–882, 2015.

[369] Horikawa, Yukio, Naoko Iwasaki, Manami Hara, Hiroto Furuta, Yoshinori Hinokio, Brian N. Cockburn, Tom Lindner, Kazuya Yamagata, Makiko Ogata, Osamu Tomonaga, Hiroyuki Kuroki, Tadasu

Kasahara, Yasuhiko Iwamoto and Graeme I. Bell: *Mutation in hepatocyte nuclear factor–1β gene (TCF2) associated with MODY*. Nature Genetics, 17(4):384–385, dec 1997.

[370] Lickert, H, C Domon, G Huls, C Wehrle, I Duluc, H Clevers, B I Meyer, J N Freund and R Kemler: *Wnt/(beta)-catenin signaling regulates the expression of the homeobox gene Cdx1 in embryonic intestine.* Development (Cambridge, England), 127(17):3805–13, sep 2000.

[371] Shu, Luan, Nadine S Sauter, Fabienne T Schulthess, Aleksey V Matveyenko, José Oberholzer and Kathrin Maedler: *Transcription factor 7-like 2 regulates beta-cell survival and function in human pancreatic islets.* Diabetes, 57(3):645–53, 2008.

[372] da Silva Xavier, G., A. Mondragon, G. Sun, L. Chen, J. A. McGinty, P. M. French and G. A. Rutter: *Abnormal glucose tolerance and insulin secretion in pancreas-specific Tcf7l2-null mice*. Diabetologia, 55(10):2667–2676, oct 2012.

[373] McCarthy, Mark I. and Eleftheria Zeggini: *Genome-wide association studies in type 2 diabetes*, 2009.

[374] Schäfer, S. A., O. Tschritter, F. Machicao, C. Thamer, N. Stefan, B. Gallwitz, J. J. Holst, J. M. Dekker, L. M. T'Hart, G. Nijpels, T. W. van Haeften, H. U. Häring and A. Fritsche: *Impaired glucagon-like peptide-1-induced insulin secretion in carriers of transcription factor 7-like 2 (TCF7L2) gene polymorphisms*. Diabetologia, 50(12):2443–2450, dec 2007.

[375] Meier, Juris J., Alexandra E. Butler, Yoshifumi Saisho, Travis Monchamp, Ryan Galasso, Anil Bhushan, Robert A. Rizza and Peter C. Butler: *β-cell replication is the primary mechanism subserving the postnatal expansion of β-cell mass in humans*. Diabetes, 57(6):1584–1594, 2008.

[376] Butler, Alexandra E, Juliette Janson, Susan Bonner-Weir, Robert Ritzel, Robert A Rizza and Peter C Butler: *Beta-cell deficit and increased beta-Cell apoptosis in humans with type 2 diabetes*. Diabetes, 52(January):102–10, 2003.

[377] Marshak, S., E. Benshushan, M. Shoshkes, L. Havin, E. Cerasi and D. Melloul: *Functional Conservation of Regulatory Elements in the pdx-1 Gene: PDX-1 and Hepatocyte Nuclear Factor 3beta Transcription Factors Mediate beta -Cell-Specific Expression*. Molecular and Cellular Biology, 20(20):7583–7590, oct 2000.

[378] Gerrish, Kevin, Michelle A. Cissell and Roland Stein: *The Role of Hepatic Nuclear Factor 1α and PDX-1 in Transcriptional Regulation of the pdx-1 Gene*. Journal of Biological Chemistry, 276(51):47775–47784, dec 2001.

[379] Akerman, Ildem, Zhidong Tu, Anthony Beucher, Delphine M.Y. Rolando, Claire Sauty-Colace, Marion Benazra, Nikolina Nakic, Jialiang Yang, Huan Wang, Lorenzo Pasquali, Ignasi Moran, Javier Garcia-Hurtado, Natalia Castro, Roser Gonzalez-Franco, Andrew F. Stewart, Caroline Bonner, Lorenzo Piemonti, Thierry Berney, Leif Groop, Julie Kerr-Conte, Francois Pattou, Carmen Argmann, Eric Schadt, Philippe Ravassard and Jorge Ferrer: *Human Pancreatic β Cell lncRNAs Control Cell-Specific Regulatory Networks*. Cell Metabolism, 25(2):400–411, 2017.

[380] Font-Cunill, Berta, Luis Arnes, Jorge Ferrer, Lori Sussel and Anthony Beucher: *Long Non-coding RNAs as Local Regulators of Pancreatic Islet Transcription Factor Genes*. Frontiers in Genetics, 9(2):400–411, nov 2018.

[381] Mamidi, Anant, Christy Prawiro, Philip A Seymour, Kristian Honnens de Lichtenberg, Abigail Jackson, Palle Serup and Henrik Semb: *Mechanosignalling via integrins directs fate decisions of pancreatic progenitors.* Nature, 564(7734):114–118, 2018.

[382] Miyoshi, Naoki, Hirotaka Wagatsuma, Shigeharu Wakana, Toshihiko Shiroishi, Masashi Nomura, Kohzoh Aisaka, Takashi Kohda, M. Azim Surani, Tomoko Kaneko-Ishino and Fumitoshi Ishino: *Identification of an imprinted gene, Meg3/Gtl2 and its human homologue MEG3, first mapped on mouse distal chromosome 12 and human chromosome 14q*. Genes to Cells, 5(3):211–220, mar 2000.

[383] You, LiangHui, Ning Wang, DanDan Yin, LinTao Wang, FeiYan Jin, YaNan Zhu, QingXin Yuan and Wei De: *Downregulation of Long Noncoding RNA Meg3 Affects Insulin Synthesis and Secretion in Mouse Pancreatic Beta Cells*. Journal of Cellular Physiology, 231(4):852–862, apr 2016.

[384] Kameswaran, Vasumathi, Nuria C. Bramswig, Lindsay B. McKenna, Melinda Penn, Jonathan Schug, Nicholas J. Hand, Ying Chen, Inchan Choi, Anastassios Vourekas, Kyoung-Jae Won, Chengyang Liu, Kumar Vivek, Ali Naji, Joshua R. Friedman and Klaus H. Kaestner: *Epigenetic Regulation of the DLK1-MEG3 MicroRNA Cluster in Human Type 2 Diabetic Islets*. Cell Metabolism, 19(1):135–145, jan 2014.

[385] Grarup, Niels, Ida Moltke, Mette K. Andersen, Peter Bjerregaard, Christina V. L. Larsen, Inger K. Dahl-Petersen, Emil Jørsboe, Hemant K. Tiwari, Scarlett E. Hopkins, Howard W. Wiener, Bert B. Boyer, Allan Linneberg, Oluf Pedersen, Marit E. Jørgensen, Anders Albrechtsen and Torben Hansen: *Identification of novel high-impact recessively inherited type 2 diabetes risk variants in the Greenlandic population*. Diabetologia, 61(9):2005–2015, sep 2018.

[386] Locke, Jonathan M., Gerald Hysenaj, Andrew R. Wood, Michael N. Weedon and Lorna W. Harries: *Targeted allelic expression profiling in human islets identifies cis-regulatory effects for multiple variants identified by type 2 diabetes genome-wide association studies*. Diabetes, 64(4):1484–1491, 2015.

[387] Quiroga, Ariel D., Lena Li, Martin Trötzmüller, Randy Nelson, Spencer D. Proctor, Harald Köfeler and Richard Lehner: *Deficiency of carboxylesterase 1/esterase-x results in obesity, hepatic steatosis, and hyperlipidemia*. Hepatology, 56(6):2188–2198, dec 2012.

[388] Pan, F. C., M. Brissova, A. C. Powers, S. Pfaff and C. V. E. Wright: *Inactivating the permanent neonatal diabetes gene Mnx1 switches insulin-producing -cells to a -like fate and reveals a facultative proliferative capacity in aged -cells*. Development, 142(21):3637–3648, nov 2015.

[389] Joseph, R, D Dou and W Tsang: *Neuronatin mRNA: alternatively spliced forms of a novel brain-specific mammalian developmental gene*. Brain research, 690(1):92–8, aug 1995.

[390] Suh, Young Ho, Won Ho Kim, Changsuk Moon, Yun Hwa Hong, Su-Yong Eun, Joo Hyun Lim, Joo Sun Choi, Jihyun Song and Myeong Ho Jung: *Ectopic expression of Neuronatin potentiates adipogenesis through enhanced phosphorylation of cAMP-response element-binding protein in 3T3-L1 cells*. Biochemical and Biophysical Research Communications, 337(2):481–489, nov 2005.

[391] Vrang, Niels, David Meyre, Phillippe Froguel, Jacob Jelsing, Mads Tang-Christensen, Vincent Vatin, Jens D. Mikkelsen, Kenneth Thirstrup, Leif K. Larsen, Karina B. Cullberg, Jan Fahrenkrug, Per Jacobson, Lars Sjöström, Lena M.S. Carlsson, Yongjun Liu, Xiaogang Liu, Hong-Wen Deng and Philip J. Larsen: *The Imprinted Gene Neuronatin Is Regulated by Metabolic Status and Associated With Obesity*. Obesity, 18(7):1289–1296, jul 2010.

[392] Millership, Steven J., Gabriela Da Silva Xavier, Agharul I. Choudhury, Sergio Bertazzo, Pauline Chabosseau, Silvia M.A. Pedroni, Elaine E. Irvine, Alex Montoya, Peter Faull, William R. Taylor, Julie Kerr-Conte, Francois Pattou, Jorge Ferrer, Mark Christian, Rosalind M. John, Mathieu Latreille, Ming Liu, Guy A. Rutter, James Scott and Dominic J. Withers: *Neuronatin regulates pancreatic β cell insulin content and secretion*. Journal of Clinical Investigation, 128(8):3369–3381, aug 2018.

[393] Dalgaard, Kevin, Kathrin Landgraf, Steffen Heyne, Adelheid Lempradl, John Longinotto, Klaus Gossens, Marius Ruf, Michael Orthofer, Ruslan Strogantsev, Madhan Selvaraj, Tess Tsai-Hsiu Lu, Eduard Casas, Raffaele Teperino, M. Azim Surani, Ilona Zvetkova, Debra Rimmington, Y.C. Loraine Tung, Brian Lam, Rachel Larder, Giles S.H. Yeo, Stephen O'Rahilly, Tanya Vavouri, Emma Whitelaw, Josef M. Penninger, Thomas Jenuwein, Ching-Lung Cheung, Anne C. Ferguson-Smith, Anthony P. Coll, Antje Körner and J. Andrew Pospisilik: *Trim28 Haploinsufficiency Triggers Bi-stable Epigenetic Obesity*. Cell, 164(3):353–364, jan 2016.

[394] Yan, Wei and Rong Shao: *Transduction of a mesenchyme-specific gene periostin into 293T cells induces cell invasive activity through epithelial-mesenchymal transformation*. Journal of Biological Chemistry, 2006.

[395] Talchai, Chutima, Shouhong Xuan, Tadahiro Kitamura, Ronald A DePinho and Domenico Accili: *Generation of functional insulin-producing cells in the gut by Foxo1 ablation.* Nature Genetics, 44(4):406–412, apr 2012.

[396] Chen, Yi-Ju, Stacy R. Finkbeiner, Daniel Weinblatt, Matthew J. Emmett, Feven Tameire, Maryam Yousefi, Chenghua Yang, Rene Maehr, Qiao Zhou, Ruth Shemer, Yuval Dor, Changhong Li, Jason R. Spence and Ben Z. Stanger: *De Novo Formation of Insulin-Producing "Neo-β Cell Islets" from Intestinal Crypts.* Cell Reports, 6(6):1046–1058, mar 2014.

[397] Lupien, Mathieu, Jérôme Eeckhoute, Clifford a Meyer, Qianben Wang, Yong Zhang, Wei Li, Jason S Carroll, X Shirley Liu and Myles Brown: *FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription.* Cell, 132(6):958–70, mar 2008.

[398] Eeckhoute, J., M. Lupien, C. A. Meyer, M. P. Verzi, R. A. Shivdasani, X. S. Liu and M. Brown: *Cell-type selective chromatin remodeling defines the active subset of FOXA1-bound enhancers.* Genome Research, 19(3):372–380, dec 2008.

[399] Mango, S E, E J Lambie and J Kimble: *The pha-4 gene is required to generate the pharyngeal primordium of Caenorhabditis elegans.* Development, 120(10):3019 LP – 3031, oct 1994.

[400] Gaudet, J and S E Mango: *Regulation of organogenesis by the Caenorhabditis elegans FoxA protein PHA-4.* Science (New York, N.Y.), 295(5556):821–5, feb 2002.

[401] Tuteja, Geetu, Shane T Jensen, Peter White and Klaus H Kaestner: *Cis-regulatory modules in the mammalian liver: composition depends on strength of Foxa2 consensus site.* Nucleic acids research, 36(12):4149–57, jul 2008.

[402] Levitsky, Victor G, Ivan V Kulakovskiy, Nikita I Ershov, Dmitry Oshchepkov, Vsevolod J Makeev, T C Hodgman and Tatyana I Merkulova: *Application of experimentally verified transcription factor binding sites models for computational analysis of ChIP-Seq data.* BMC Genomics, 15(1):80, 2014.

[403] Verzi, Michael P, Hyunjin Shin, H Hansen He, Rita Sulahian, Clifford a Meyer, Robert K Montgomery, James C Fleet, Myles Brown, X Shirley Liu and Ramesh a Shivdasani: *Differentiation-specific histone modifications reveal dynamic chromatin interactions and partners for the intestinal transcription factor CDX2.* Developmental cell, 19(5):713–26, nov 2010.

[404] Verzi, M. P., H. Shin, L.-L. Ho, X. S. Liu and R. A. Shivdasani: *Essential and Redundant Functions of Caudal Family Proteins in Activating Adult Intestinal Genes.* Molecular and Cellular Biology, 31(10):2026–2039, may 2011.

[405] Kumar, Namit, Yu-Hwai Tsai, Lei Chen, Anbo Zhou, Kushal K. Banerjee, Madhurima Saxena, Sha Huang, Natalie H. Toke, Jinchuan Xing, Ramesh A. Shivdasani, Jason R. Spence and Michael P. Verzi: *The lineage-specific transcription factor CDX2 navigates dynamic chromatin to control distinct stages of intestine development.* Development, 146(5):dev172189, mar 2019.

[406] Li, Zhaoyu, Peter White, Geetu Tuteja, Nir Rubins, Sara Sackett and Klaus H. Kaestner: *Foxa1 and Foxa2 regulate bile duct development in mice.* Journal of Clinical Investigation, 119(6):1537–1545, jun 2009.

[407] Li, Zhaoyu, Jonathan Schug, Geetu Tuteja, Peter White and Klaus H Kaestner: *The nucleosome map of the mammalian liver.* Nature Structural & Molecular Biology, 18(6):742–746, jun 2011.

[408] Korostylev, Alexander, Pallavi U. Mahaddalkar, Oliver Keminer, Kamyar Hadian, Kenji Schorpp, Philip Gribbon and Heiko Lickert: *A high-content small molecule screen identifies novel inducers of definitive endoderm.* Molecular Metabolism, 6(7):640–650, jul 2017.

[409] Heinz, Sven, Christopher Benner, Nathanael Spann, Eric Bertolino, Yin C. Lin, Peter Laslo, Jason X. Cheng, Cornelis Murre, Harinder Singh and Christopher K. Glass: *Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities.* Molecular Cell, 38(4):576–589, may 2010.

[410] Subramanian, A., P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander and J. P. Mesirov: *Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles*. Proceedings of the National Academy of Sciences, 102(43):15545–15550, 2005.

[411] Mootha, Vamsi K, Cecilia M Lindgren, Karl-Fredrik Eriksson, Aravind Subramanian, Smita Sihag, Joseph Lehar, Pere Puigserver, Emma Carlsson, Martin Ridderstråle, Esa Laurila, Nicholas Houstis, Mark J Daly, Nick Patterson, Jill P Mesirov, Todd R Golub, Pablo Tamayo, Bruce Spiegelman, Eric S Lander, Joel N Hirschhorn, David Altshuler and Leif C Groop: *PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes*. Nature Genetics, 34(3):267–273, 2003.

[412] Hellemans, Jan, Geert Mortier, Anne De Paepe, Frank Speleman and Jo Vandesompele: *qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data*. Genome biology, 8(2):R19, jan 2007.

[413] Patro, Rob, Geet Duggal, Michael I Love, Rafael A Irizarry and Carl Kingsford: *Salmon provides fast and bias-aware quantification of transcript expression*. Nature Methods, 14(4):417–419, apr 2017.

[414] Ignatiadis, Nikolaos, Bernd Klaus, Judith B. Zaugg and Wolfgang Huber: *Data-driven hypothesis weighting increases detection power in genome-scale multiple testing*. Nature Methods, 2016.

[415] Zhu, Anqi, Joseph G Ibrahim and Michael I Love: *Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences*. Bioinformatics, nov 2018.

[416] Guo, Yuchun, Shaun Mahony and David K. Gifford: *High Resolution Genome Wide Binding Event Finding and Motif Discovery Reveals Transcription Factor Spatial Binding Constraints*. PLoS Computational Biology, 8(8), 2012.

[417] Xing, Haipeng, Yifan Mo, Will Liao and Michael Q. Zhang: *Genome-wide localization of protein-DNA binding and histone modification by a bayesian change-point method with ChIP-seq data*. PLoS Computational Biology, 8(7), 2012.

[418] ENCODE Project Consortium: *An integrated encyclopedia of DNA elements in the human genome*. Nature, 489(7414):57–74, 2012.

[419] Denny, Sarah K., Dian Yang, Chen Hua Chuang, Jennifer J. Brady, Jing Shan S. Lim, Barbara M. Grüner, Shin Heng Chiou, Alicia N. Schep, Jessika Baral, Cécile Hamard, Martine Antoine, Marie Wislez, Christina S. Kong, Andrew J. Connolly, Kwon Sik Park, Julien Sage, William J. Greenleaf and Monte M. Winslow: *Nfib Promotes Metastasis through a Widespread Increase in Chromatin Accessibility*. Cell, 166(2):328–342, 2016.

[420] Stephens, Matthew: *False discovery rates: a new deal*. Biostatistics, page kxw041, oct 2016.

[421] Wolf, F. Alexander, Philipp Angerer and Fabian J. Theis: *SCANPY: large-scale single-cell gene expression data analysis*. Genome Biology, 19(1):15, dec 2018.

[422] Johnson, W. Evan, Cheng Li and Ariel Rabinovic: *Adjusting batch effects in microarray expression data using empirical Bayes methods*. Biostatistics, 8(1):118–127, jan 2007.

[423] McInnes, Leland, John Healy and James Melville: *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. feb 2018.

Contributions

Michael Sterr performed ChIP-seq and ATAC-seq experiments and analyzed imaging, microarray, ChIP-seq, ATAC-seq and scRNA-seq data. Xianming Wang performed stem cell experiments. Alexandra Aliluev performed stainings, lineage tracing experiments and FACS for microarry ans scRNA-seq analysis. Sophie Tritschler performed scRNA-seq experiment and initial scRNA-seq data analysis, including alignment, barcode and UMI counting, preprocessing, clustering, annotation and pseudotime analysis. Anika Böttcher performed scRNA-seq experiment and stainings. Martin Irmler performed microarray experiment.

Acknowledgements

First of all, I would like to thank all the people who were involved in the success of this thesis! Without your help, I could not have accomplished this work!

In particular, I would like to thank my PhD advisor, Heiko Lickert, for giving me the opportunity to prepare my thesis in his lab. Thank you for providing an excellent scientific environment, for being a great supervisor and for all your support, ideas and enthusiasm! I have learned a lot during the past years and I'm very grateful for that!

Moreover, I would like to thank Anika Böttcher for supervising my project. Thank for all your great help and critical input. For the discussions, all the answers to countless questions and a great time in the gut group! It was great to work with you and I have learned a lot from you!

I would like to thank our collaboration partners. In particularly I would like to thank Gunnar Schotta and Filippo Cernilogar for welcoming me in your lab and for teaching me ChIP and ATAC-seq. Thanks for all your support and help during the last years! It was a pleasure working with you! I also want to thank Martin Irmler and Johannes Beckers for performing microarray experiments and great help with the GEO uploads, and for valuable input in the thesis committee! Furthermore, I want to thank Stefan Krebs and Elisabeth Graf for sequencing.

Special thanks go also to Xianming Wang and Mostafa Bakhti for pushing our joint papers and great help to bring them over the finish line!

I would like to thank the gut group, Anika Böttcher, Alexandra Aliluev, Lena Oppenländer and Jürgen Schultheiß for the great teamwork and support, and a great time in the lab and office! The single-cell group, Anika Böttcher, Sophie Tritschler and Ines Kunze, for teaching me scRNA-seq, all the great help and code for data analysis, and for the superb support in performing the experiments and for backing me up in last months! I want specially thank Alexandra Aliluev and Felizitas von Hahn for the great friendship and support, and the good times we had in the last years! I want to thank all the IDR members, and in particular Donna Thomson, Kerstin Diemer, Pallavi Mahaddalkar, Katharina Scheibner and Noel Moya for a great time and support!

Moreover, I want to thank Mostafa Bakhti, Anika Böttcher, Felizitas von Hahn, Andrzej Malinowski, Aimee Bastidas-Ponce, Zofia Sterr and Franz Sterr who helped me to correct this thesis!

At last, I want to thank my family for their support! I want to thank my parents for all their help and support during my entire studies. Thank you for everything! And especially, I want to thank Steffi for the love, patience and support! And thank you Anna for making me so happy!

Publications

Wang, X., **Sterr, M.**, Ansarullah ,Burtscher, I., Böttcher, B., Beckenbauer, J., Siehler, J., Meitinger, T., Häring, H.-U., Staiger, H., Cernilogar, F. M., Schotta, G., Irmler, M., Beckers, J., Wright, C. V. E., Bakhti, M., Lickert, L. (2019). *Point mutations in the PDX1 transactivation domain impair human β-cell development and function.* Mol. Metab. Accepted

Wang, X.*, **Sterr, M.***, Burtscher, I., Chen, S., Hieronimus, A., Machicao, F., Staiger, H., Häring, H.-U., Lederer, G., Meitinger, T., Cernilogar, F. M., Schotta, G., Irmler, M., Beckers, J., Hrabě de Angelis, M., Ray, M., Wright, C. V. E., Bakhti, M., Lickert, L. (2018). *Genome-wide analysis of PDX1 target genes in human pancreatic progenitors.* Mol. Metab. 9, 57–68 ***Contributed equally**

Bastidas-Ponce, A., Roscioni, S.S., Burtscher, I., Bader, E., **Sterr, M.**, Bakhti, M., and Lickert, H. (2017). *Foxa2 and Pdx1 cooperatively regulate postnatal maturation of pancreatic β-cells.* Mol. Metab. 6, 524–534

Wang, X., Chen, S., Burtscher, I., **Sterr, M.**, Hieronimus, A., Machicao, F., Staiger, H., Häring, H.-U., Lederer, G., Meitinger, T., and Lickert, H. (2016). *Generation of a human induced pluripotent stem cell (iPSC) line from a patient carrying a P33T mutation in the PDX1 gene.* Stem Cell Res. 17, 273–276.

Wang, X., Chen, S., Burtscher, I., **Sterr, M.**, Hieronimus, A., Machicao, F., Staiger, H., Häring, H.-U., Lederer, G., Meitinger, T., and Lickert, H. (2016). *Generation of a human induced pluripotent stem cell (iPSC) line from a patient with family history of diabetes carrying a C18R mutation in the PDX1 gene.* Stem Cell Res. 17, 292–295.

Willmann, S.J., Mueller, N.S., Engert, S., **Sterr, M.**, Burtscher, I., Raducanu, A., Irmler, M., Beckers, J., Sass, S., Theis, F.J., and Lickert, H. (2016). *The global gene expression profile of the secondary transition during pancreatic development.* Mech. Dev. 139, 51–64.

Cremer, T., Cremer, M., Hübner, B., Strickfaden, H., Smeets, D., Popken, J., **Sterr, M.**, Markaki, Y., Rippe, K., and Cremer, C. (2015). *The 4D nucleome: Evidence for a dynamic nuclear landscape based on co-aligned active and inactive nuclear compartments.* FEBS Lett. 589, 2931–2943.

Smeets, D., Markaki, Y., Schmid, V.J., Kraus, F., Tattermusch, A., Cerase, A., **Sterr, M.**, Fiedler, S., Demmerle, J., Popken, J., Leonhardt, H., Brockdorff, N., Cremer, T., Schermelleh, L., and Cremer, M. (2014). *Three-dimensional super-resolution microscopy of the inactive X chromosome territory reveals a collapse of its active nuclear compartment harboring distinct Xist RNA foci.* Epigenetics Chromatin 7, 8