REGULATION OF PANCREATIC ENDOCRINE CELL DEVELOPMENT THROUGH LINKED TRANSCRIPTIONAL AND SPLICING PROGRAMS

By

Karrie Danielle Dudek

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

In partial fulfillment of the requirements

For the degree of

DOCTOR OF PHILOSOPHY

In

Cell and Developmental Biology

August 13, 2021

Nashville, Tennessee

Christopher Wright, D. Phil.

Guoqiang Gu, PhD

Michelle Southard-Smith, PhD

Wenbiao Chen, PhD

Mark A. Magnuson, MD

DEDICATION

To my parents Kenny and Susan Croy

To my partner Danni and my daughter Rylee

and

To my nieces and nephews

ACKNOWLEDGMENTS

My graduate studies and training were supported for two years by from the Molecular Endocrinology Training Program (METP) (NIH grant T32 DK07563) and enriched by my use of the Vanderbilt Genome Editing and Resource Center (VGER), the Vanderbilt Technologies for Advanced Genomics (VANTAGE), by the Vanderbilt Cell Imaging Shared Resource (CISR), and the Vanderbilt Hormone Assay & Analytical Services Core.

A long list of individuals has provided varying degrees of support and assistance throughout my graduate training. I will forever be grateful for your encouragement, inspiration, and mentorship that have helped me reach this point. First, I would like to thank my thesis adviser, Dr. Mark Magnuson, for providing me the space to pursue my scientific curiosities and grow as a scientist over the last 2200^{+} days as a graduate student. Your critical feedback pushed me to sharpen my thinking and brought my work to a higher level. Your patience during my most trying times has been appreciated.

I would also like to thank all of the members of the Magnuson laboratory, as well as the Vanderbilt Genome Editing Resource core. Pamela Uttz, thank you for your constant encouragement, positive attitude, and for all of our talks on life. I am thankful for Dr. Anna Osipovich, who has been a pillar of the Magnuson lab. Your scientific and technical expertise has been invaluable, and I appreciate the time and patience you provided in imparting your knowledge. Thank you to Jennifer Skelton, Linda Gower, and Dr. Leesa Sampson of VGER for your assistance in generating countless mouse models and for all our sporadic chats about dogs, kids, and our daily scientific challenges. I am also thankful to Dr. JP Cartailler for your undeniable bioinformatics skills and patience in providing more profound levels of understanding. Your positivity and creative spirit have been inspiring and motivating. Thank you to fellow past and present graduate students Linh Trinh, Hannah Clayton, and Jennifer Stancil. It has been fun to see everyone succeed and grow individually. I would like to thank the undergraduates, masters' students and medical students that have worked directly with me to help push my research forward. Jacob Coeur, Austin Chapman, and David Meehan, you have all looked to me at one time for mentorship and guidance, but I was the one to gain so much from each of you.

I am incredibly grateful to my committee members Drs. Chris Wright, Guoqiang Gu, Michelle Southard-Smith, and Wenbiao Chen. You have pushed me in ways that have let me grow as a scientist. Your insights, encouragement, and guidance have allowed me to get where I am today. I would especially like to thank Dr. Wright

for his endless motivation and mentorship. You have been a rock when things have been most challenging, and your infectious enthusiasm for science incites an excitement that has made this journey enjoyable. Our impromptu chats about our studies have always prompted me to think more deeply and broadly.

I must give a huge thank you to members of my family. My mom, Susan, and dad, Kenny, you have been incredible role models and pillars of support. Even when unbelievable trials have marked this season of our lives, your strength throughout has been inspirational and never-ending. I wish to someday have half of the strength the two of you do. There is no doubt that I would not be here without the love and encouragement, as well as the occasional kick in the ass, which you have given all these years. To my nieces and nephews, you were my first loves. Having the privilege to watch you grow into the wonderful young people you are has brought me immense joy and happiness. I can always depend on you all for laughs, love, and for telling me what is “cool” these days.

Last but certainly not least, I want to thank my partner Danni and my daughter Rylee. This has been a journey for you, just as much as it has been for me. I have always been able to count on you for your support and appreciation, and you have often shouldered more than your share of the responsibilities. I greatly appreciate your constant steadiness and understanding and can’t wait to see where the future takes our family and us. To my daughter Rylee, you are my daily motivation and inspiration behind everything I do. You are the highlight of all of my days. I hope that you never lose your happiness and exuberance and that you choose to chase all of your dreams no matter how big or far they may seem.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AS	Alternative splicing
DRG	Downregulated genes
E	Embryonic day
EMT	Epithelial to mesenchymal transition
FACS	Fluorescence activated cell sorting
Gcg	Glucagon
GFP	Green fluorescent protein
Hes1	Hairy and enhancer of split-1
Hnf	Hepatic nuclear factor
Ins	Insulin
Insm1	Insulinoma 1
Jazf1	JAZF zinc finger 1
MPC	Multipotent progenitor cell
Neurod1	Neuronal differentiation 1
Neurog3	Neurogenin 3 (Ngn3)
Pax4	Paired box 4
Pax6	Paired box 6
PCR	Polymerase chain reaction
Pdx1	Pancreatic and duodenal homeobox 1
PP	Pancreatic Polypeptide
Ptf1a	Pancreas transcription factor 1 subunit alpha
RBP	RNA binding protein
Snap25	Synaptosome associated protein 25
Sox9	SRY (sex determining region)-box 9
Sst	Somatostatin
Syt14	Synaptotagmin 14
URG	Upregulated genes
Zfhx4	Zinc finger homeobox 4
Zfp329	Zing finger protein 329
Zfp800	Zing finger protein 800
Zfp92	Zing finger protein 92

CHAPTER I – INTRODUCTION

Endoderm to pancreas: Early pancreas development, specification, and growth

The pancreas is a complex organ with two primary functions coordinated by two distinct regions within the pancreas. The exocrine compartment makes up ~95% of the pancreas and functions to secrete enzymes (trypsin, amylase, etc.) necessary to digest food to be used as the fuel source for the rest of the body. The exocrine acinar cells secrete these enzymes into a duct system, formed by pancreatic ductal cells, and are deposited into the duodenum (Shih et al., 2013). The duct cells also serve a secretory function by releasing bicarbonate into the duodenum to neutralize the acidic stomach bile (Grapin-Botton, 2005; Reichert and Rustgi, 2011). The endocrine pancreas, which only makes up approximately 2% of the entire organ, comprises five different specialized hormone-secreting islet cells. That is, insulin-expressing β-, glucagon-expressing α-, somatostatin-expressing δ-, pancreatic polypeptide-expressing (Ppy) PP-, and a minor population of ghrelin-expressing ε-cells (Oliver-Krasinski et al., 2009; Pan and Wright, 2011; Shih et al., 2013). Regulation of blood glucose homeostasis is a critical function of the endocrine pancreas, driven by the reciprocal secretion of glucagon (Gcg) and insulin (Ins) from α- and β-cells, respectively, into the bloodstream. Tightly controlled glucose homeostasis is crucial for the proper function of other organs in the body. Outcomes associated with dysfunctional endocrine cells are discussed later.

Pancreatic differentiation and development is highly conserved across vertebrates and driven by highly coordinated processes dependent on the interplay of extrinsic signaling cues and intrinsic transcriptional regulation (Oliver-Krasinski et al., 2009; Pan and Wright, 2011). The underlying genetic programs and intricate regulatory interactions are essential for developing pancreatic multipotent progenitor cells (MPCs) into each of the pancreatic cell types. Pancreas development begins as early as E8.0 in mice and continues through post-natal stages with the maturation of fully differentiated cells (**Fig. 1.1A**). Developmental stages of pancreas organogenesis occur in three transitional phases: the primary transition (~E8.5-E12.5), the secondary transition (~E12.5-E16.5), and the tertiary transition (~E16.5-postnatal) (**Fig. 1.1B**) (Dassaye et al., 2016; Rutter et al., 1968). In general, the primary transition is a period marked by pancreas specification of the endoderm and robust proliferation of pancreatic progenitor cells, followed by a phase of rapid growth, differentiation, and decreased cellular proliferation in the secondary transition.

Finally, the tertiary transition state is defined by the final stages of maturation, cellular organization, and ultimately the shift to the cells' primary function in secreting specific protein as a metabolic response to dietary changes (Dassaye et al., 2016; Rutter et al., 1968).

During the early stages of pancreas specification, the pre-patterning of the definitive endoderm is stimulated by retinoic acid (RA), FGF, and BMP signaling cues from the surrounding mesoderm, notochord, and dorsal aorta (**Fig. 1.2A-C, 1.3A-C**) (Dessimoz and Grapin-Botton, 2006; Jorgensen et al., 2007; Kumar et al., 2003; Rankin et al., 2018). This step begins shortly after gastrulation at ~E8.0-E8.5 and is denoted by the flat sheet of endoderm being converted into the primitive gut tube (**Figs. 1.1 and 1.3A**) (Wells and Melton, 1999). Formation and closure of the gut tube require the expression of many factors, including GATA4, Furin, MMP2, and several BMP signaling molecules (**Fig. 1.3A, D**) (Jorgensen et al., 2007). The primitive gut tube is further stimulated by a combination of extrinsic signaling molecules (e.g., Wnt, RA, Fgf, Notch, etc.) and intrinsic transcriptional changes to initiate anterior-posterior (A-P) patterning, from which the foregut (anterior), midgut, and hindgut (posterior) are established (**Fig. 1.2A-C**) (Jorgensen et al., 2007; Pan and Wright, 2011; Rivera-Perez and Hadjantonakis, 2014). These domains become increasingly distinct through divergent gene expression patterns. The foregut endoderm is specifically marked by the expression of *Hhex*, *Sox2*, and *Foxa2* (Zorn and Wells, 2009).

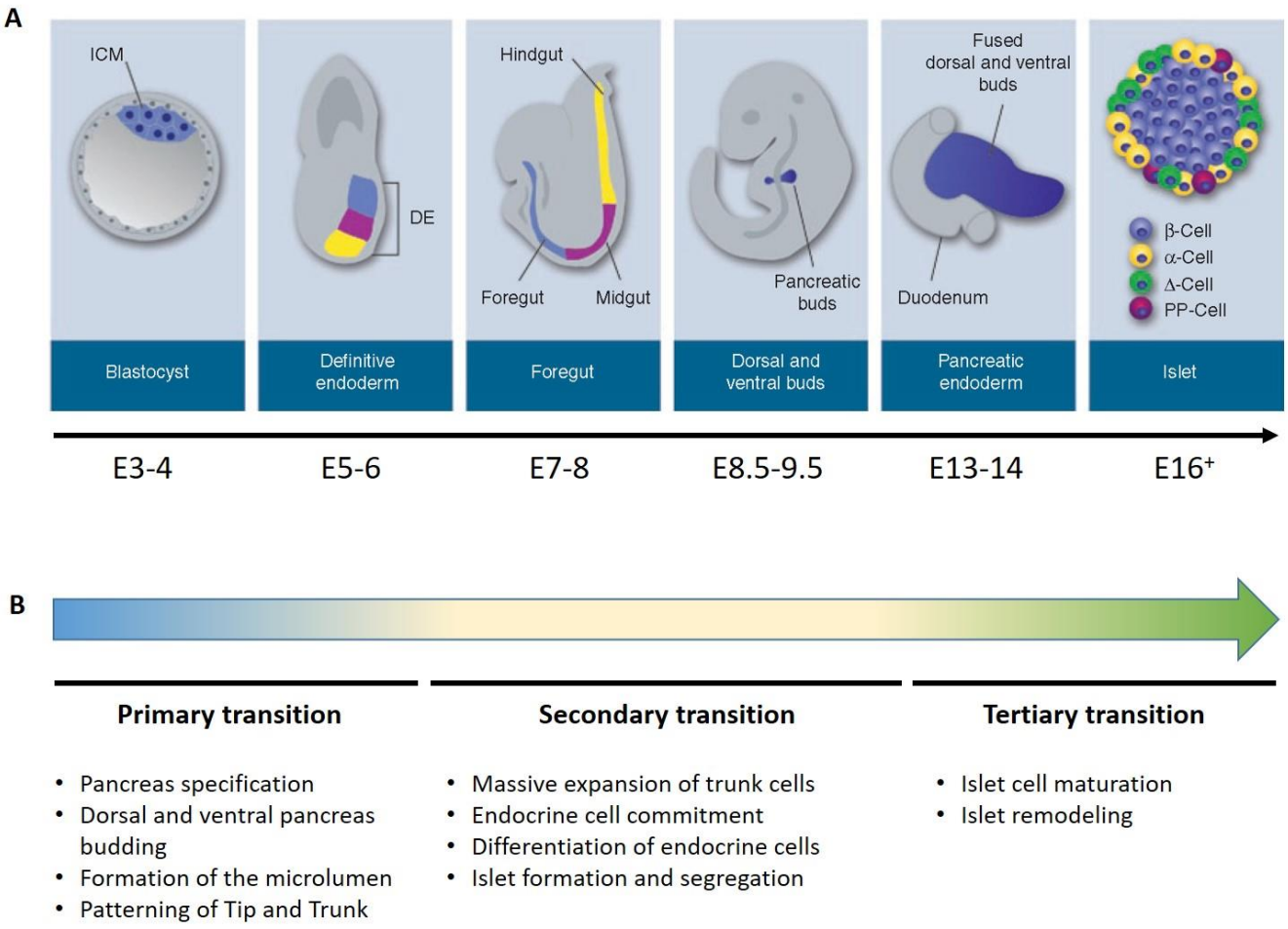


Figure 1.1. Simplified schematic presentation of pancreas development in mice. The inner cell mass (blue) of the blastocyst, sometimes referred as embryoblast, gives rise to the three germ layers in the process of gastrulation. The definitive endoderm is then formed by the recruitment of epiblast cells through the primitive streak via a mesendodermal progenitor with the latter cells of the foregut (blue), midgut (purple), and hindgut (yellow). Morphogenesis of the primitive gut is a result of an invagination movement by which the layered definitive endoderm becomes a tube structure. The pancreas formation begins with the independent budding of the dorsal and ventral buds at the posterior region of the foregut. These two buds grow into the surrounding mesenchyme, branch in a tree-like structure and eventually fuse after rotation of the gut to form the definitive pancreatic endoderm. This predifferentiated epithelium grows in size with distinct endocrine and exocrine differentiation. The endocrine cells are organized in islets which are embedded in exocrine tissue and are composed of four major hormone-secreting cells types. Insulin is secreted by β-cells (blue), glucagon by α-cells (yellow), somatostatin by Δ-cells (green), and pancreatic polypeptide by PP-cells (purple). The timeline plots these key events for mouse. For comparison only, comparable stages of human β-cell development have been mapped on the timeline. Several markers characteristic of each developmental step are listed. DE, definitive endoderm; GLC, glucagon; ICM, inner cell mass; PP, pancreatic polypeptide; SST, somatostatin. This figure was used with permission from Naujok, et al., 2011.

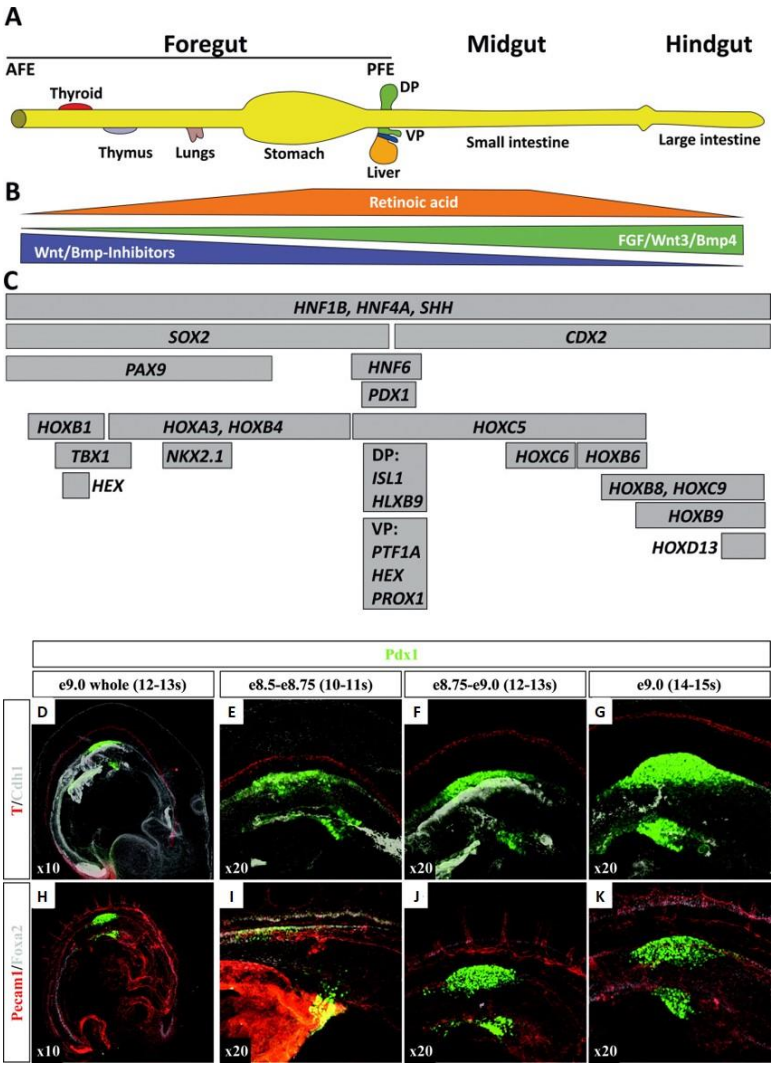


Figure 1.2. Anterior–posterior (A–P) patterning of the endoderm. Top: (**A**): During gastrulation and early somite stages the definitive endoderm is patterned by interactions with the splanchnic mesoderm and other mesodermal derivatives into the gut tube domains foregut, midgut, and hindgut. (**B**): Combined data from several model organisms indicate that differential activation or inhibition of Wnt/β-catenin-, FGF-, retinoic acid-, and BMP-signaling along the A–P axis patterns the endoderm in distinct domains. (**C**): Expression domains of transcription factors in the primitive gut tube of an e9.5 mouse embryo are depicted. (Scheme adapted from Davenport et al., 2016). Abbreviations: AFE, anterior foregut endoderm; DP, dorsal pancreatic bud; PFE, posterior foregut endoderm; VP, ventral pancreatic bud. **Bottom**: Positioning of notochord and dorsal aorta during pancreas specification. Collapsed image stacks of whole mount immunohistochemical detection: **A–D**, Pdx1 in the pancreatic epithelium, T in the notochord, and Cdh1 in epithelial cells in general; and E–H, Pdx1 in the pancreatic epithelium, Pecam1 in endothelial cells, and Foxa2 staining all endoderm and neural tube floor plate in E8.5–E9.0 wild-type mouse embryos. **A** and **E**, Low magnification overview pictures of the E9.0 embryos shown in panels **C** and **G**. **B**, Between E8.5 and E8.75, the notochord lies in close proximity to the dorsal pancreatic epithelium. **C** and **D**, At E9.0, the notochord and the pancreatic epithelium become separated, and the distance between them increases as development proceeds. **F**, Between E8.5 and E8.75, there are two lateral aortas along the sides of the notochord. **G** and **H**, At E9.0, the two aortas fuse between the notochord and the dorsal pancreatic epithelium. In all panels, anterior is to the right, posterior is to the left, dorsal is above, and ventral is below. These images were modified and used with permission from Davenport, et al., 2016 and Jorgensen, et al. 2007.

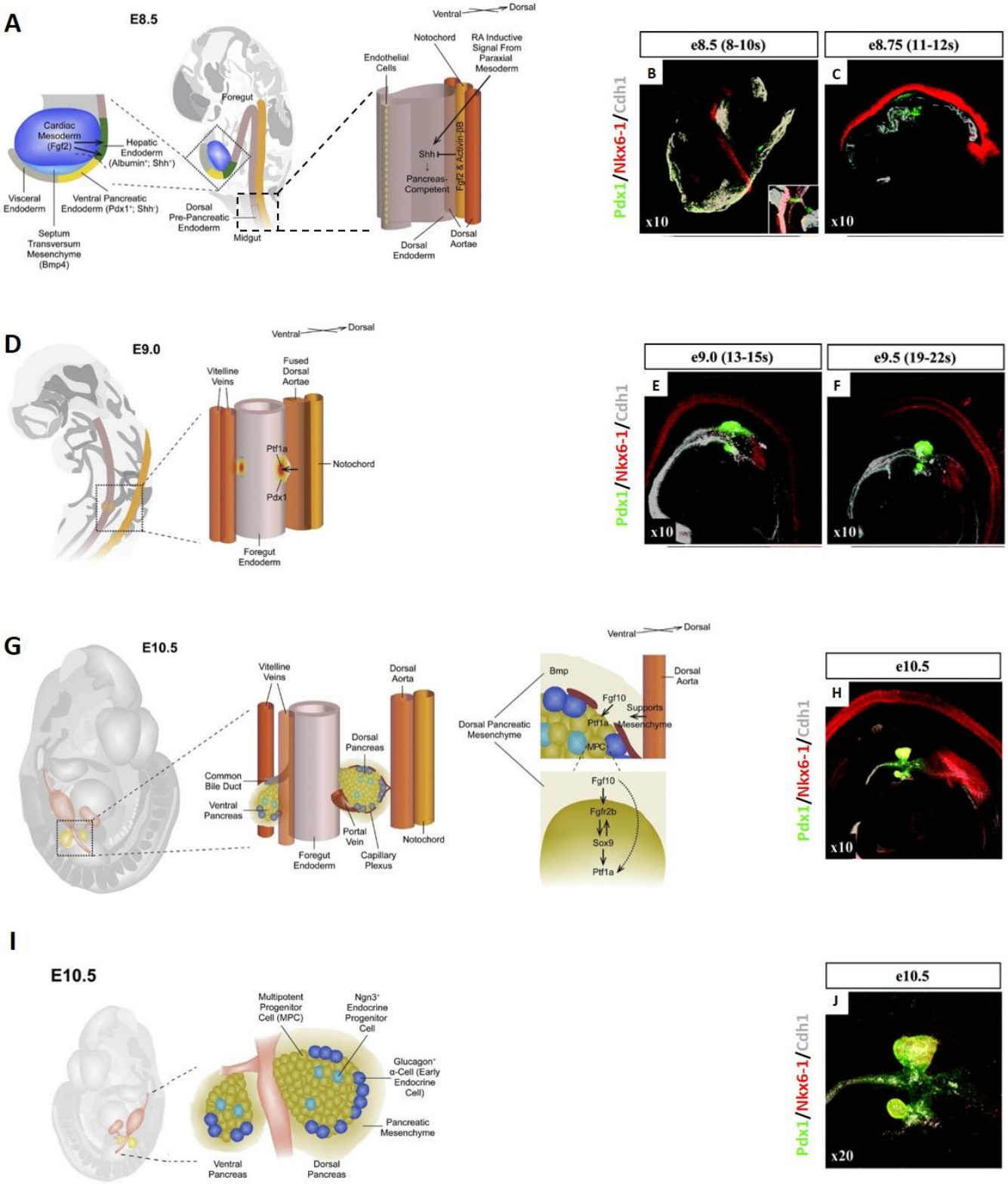

A
E8.5
Ventral
Dorsal
Notochord
RA Inductive Signal From Paraxial Mesoderm
Endothelial Cells
Foregut
Cardiac Mesoderm (Fgf2)
Hepatic Endoderm (Albumin+; Shh+)
Shh
Pancreas-Competent
Fgf2 & Activin-βB
Visceral Endoderm
Ventral Pancreatic Endoderm (Pdx1+; Shh-)
Dorsal Pre-Pancreatic Endoderm
Midgut
Septum Transversum Mesenchyme (Bmp4)
Dorsal Endoderm
Dorsal Aortae
e8.5 (8-10s)
e8.75 (11-12s)
B
C
Pdx1/Nkx6-1/Cdh1
x10
x10
D
E9.0
Ventral
Dorsal
Vitelline Veins
Fused Dorsal Aortae
Ptf1a
Notochord
Pdx1
Foregut Endoderm
e9.0 (13-15s)
e9.5 (19-22s)
E
F
Pdx1/Nkx6-1/Cdh1
x10
x10
G
E10.5
Vitelline Veins
Dorsal Aorta
Dorsal Pancreas
Common Bile Duct
Ventral Pancreas
Portal Vein
Capillary Plexus
Foregut Endoderm
Notochord
Ventral
Dorsal
Bmp
Dorsal Aorta
Fgf10
Supports Mesenchyme
Dorsal Pancreatic Mesenchyme
Ptf1a
MPC
Fgf10
Fgfr2b
Sox9
Ptf1a
e10.5
H
Pdx1/Nkx6-1/Cdh1
x10
I
E10.5
Multipotent Progenitor Cell (MPC)
Ngn3+ Endocrine Progenitor Cell
Glucagon+ α-Cell (Early Endocrine Cell)
Pancreatic Mesenchyme
Ventral Pancreas
Dorsal Pancreas
e10.5
J
Pdx1/Nkx6-1/Cdh1
x20

Figure 1.3. Pancreatic fate is assigned in the gut endoderm via a series of inductive interactions with adjacent mesodermal tissues. (**A**) Prior to E8.0, before midline fusion of the dorsal aortas at E8.5 displaces it, the notochord signals to the adjacent dorsal endoderm, possibly via Fgf2 and activin-βB, to repress (barred line) endodermal Shh expression in the putative dorsal pancreatic domain, rendering it competent for subsequent Pdx1 induction and pancreatic fate commitment. Concurrently, the dorsal pancreas responds to RA emanating from the paraxial mesoderm and ventrally, Hhex expression is required to drive proliferation and movement (arrow) of the leading edge of the ventral foregut endoderm (VFGE) lip to position it beyond the influence of signals from the cardiac mesoderm. (**B**) VFGE exposed to Fgf2 emanating from the cardiac mesoderm (arrows) and Bmp4 from the septum transversum mesenchyme retains Shh expression and undergoes hepatic differentiation, but VFGE beyond the influence of such signaling (denoted by the curved dashed line) loses Shh and induces Pdx1 expression, committing to a pancreatic fate. (**C**) By E9.0, turning and gut closure have brought the presumptive Pdx1+ dorsal and ventral pancreatic endoderm into apposition where they are positioned adjacent to the fused dorsal aorta and the vitelline veins, respectively. Signaling from the adjacent dorsal aorta (arrow) induces Ptf1a expression within the broader dorsal Pdx1+ domain. (**D**) By E10.5, mesenchyme has enveloped the dorsal and ventral buds and a capillary plexus develops within the mesenchyme overlying the epithelium and the portal vein develops in proximity to the proximal dorsal pancreas. Signals from the dorsal aorta are required to maintain the juxtaposed dorsal pancreatic mesenchyme while Fgf10 is required to promote both the expansion and pancreatic commitment of the pancreatic epithelium. Mesenchymal Fgf10 signaling is transduced via epithelial Fgfr2b to maintain expression of Sox9. Fgf10 also maintains epithelial Ptf1a expression, either directly, or indirectly via Sox9 which directly activates MPC Ptf1a expression. These images were modified and used with permission Seymour and Serup, 2019 and Jorgensen, et al., 2007.

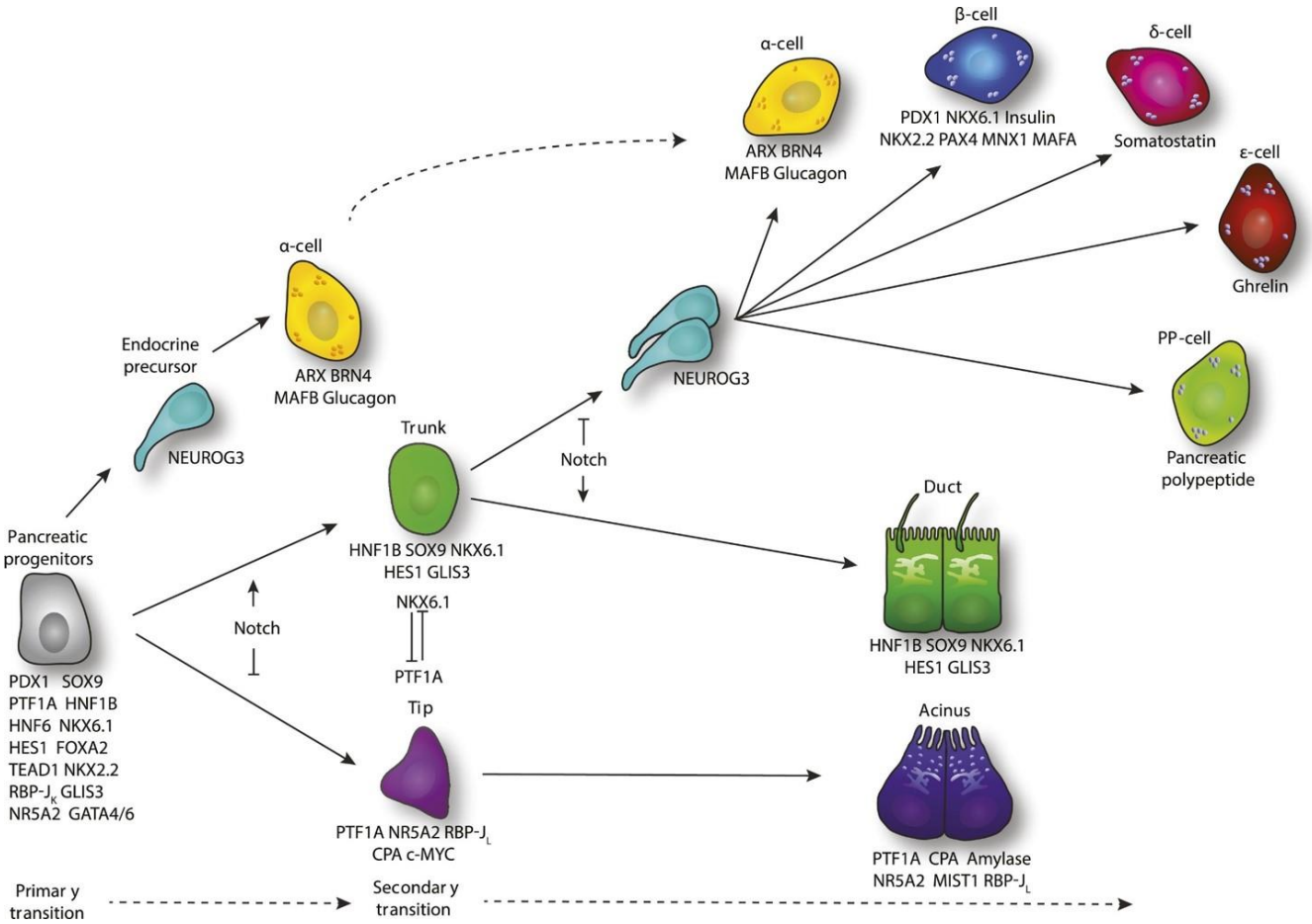


Figure 1.4. Lineage hierarchy during pancreas organogenesis. Following specification, the multipotent pancreatic progenitors co-express a range of pancreas-associated transcription factors facilitating establishment of the gene regulatory network mediating multipotency at the early stages of pancreas development. Priming towards the endocrine lineage occurs in scattered multipotent progenitors via transient expression of Neurog3, subsequently leading to the emergence of primary transition-derived α-cells. Notch signaling and reconfiguration of the transcriptional connectivity between Ptf1a and Nkx6.1 next initiate tip-trunk segregation, leading to the segregation of the acinar and ducto-endocrine lineages. Expression of Rbp-JL facilitates the generation of the Ptf1a-Rbp-JL transcriptional complex mediating maturation of acinar progenitors into mature endocrine cells along with acinar specific transcription factors such as Nr5a2 and Mist1. In the ducto-endocrine bipotent population, scattered progenitors are primed towards the endocrine lineage via NEUROG3 expression facilitated by low levels of Notch signaling. The population of secondary transition-derived endocrine precursors primarily gives rise to β-cells but also to the other four endocrine subtypes. Remaining ductal progenitors eventually mature into ductal cells displaying primary cilia and hydrogen bicarbonate production. These images were modified and used with permission Larsen and Grapin-Botton, 2017.

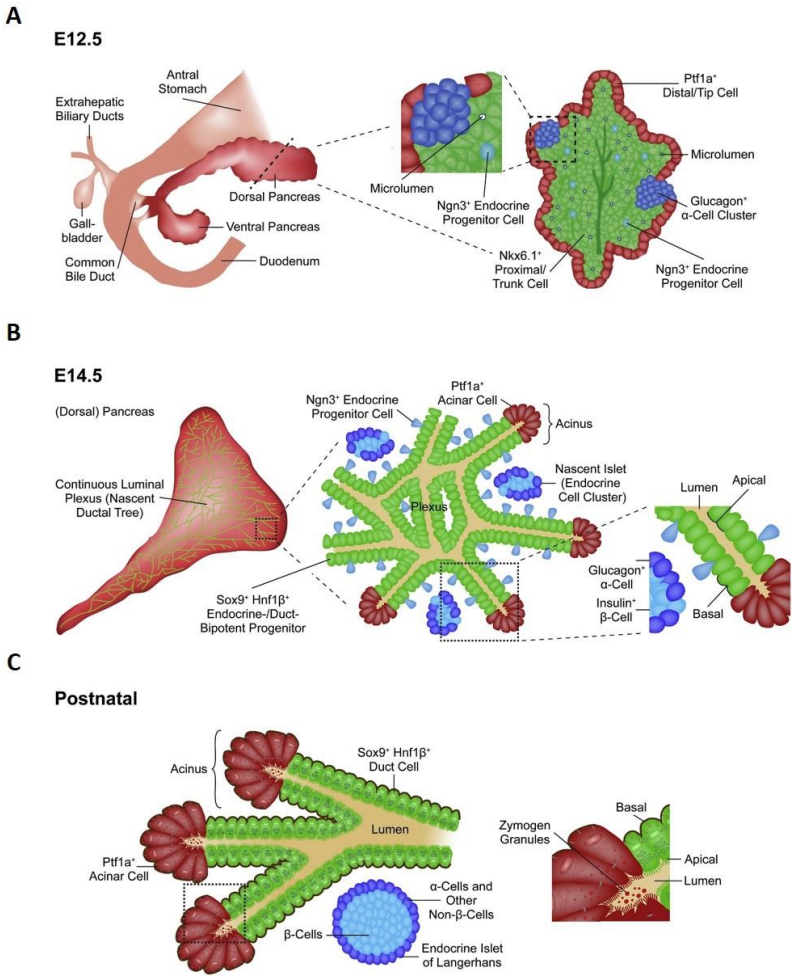


Figure 1.5. Overview of pancreas development during the secondary transition and later stages. (A) By E10.5, the dorsal and ventral pancreatic "buds" are composed of multipotent progenitor cells (MPCs) with isolated Ngn3+ endocrine progenitors, which at this stage give rise to α-cells. Between E10.5 and E12.5, the pancreatic buds grow parallel to the presumptive duodenum and stomach and gut rotation results in their fusion by E12.5. (**B-C**) From E11.5 the pancreatic epithelium organizes into a contiguous luminal plexus via Cdc42-dependent microlumen expansion and luminal coalescence and this plexus remodels between E13.5-E15.5 into the ductal tree. Concurrently, MPCs become segregated via proximodistal patterning into proximal, Sox9+Hnf1β+Nkx6.1+ bipotent trunk progenitors of the plexus and distal, Ptf1a+ acinar-committed tip progenitors. Ngn3+ endocrine progenitors begin to give rise to insulin+ cells at this stage. (**C**) By the mid-point of the secondary transition at E14.5, the plexus, still composed of bipotent progenitors marked by Sox9, Nkx6-1 and Hnf1β, has begun to resolve into the branched tree-like structure in the periphery while the plexus is maintained in the central part. Ngn3Hi cells originating from the plexus commit to an endocrine fate, delaminate and coalesce into endocrine clusters, the nascent islets of Langerhans. Trunk cells which fail to initiate endocrine differentiation adopt a ductal identity. The distal ends of the duct progenitors are capped by tip-derived acinar cells which continue to express Ptf1a. (**D**) Following terminal maturation, Ptf1a+ acinar cells produce and secrete digestive enzymes in zymogen granules into the gut via the ductal tree, a subpopulation of which continues to be marked by Sox9 and Hnf1β. Differentiation of later-arising endocrine (such as δ-) cells completes the ontogeny of the islets of Langerhans with their characteristic architecture with a core of β-cells and a mantle of α- and other non-β-cell types. This image was modified and used with permission Seymour and Serup, 2019.

Final patterning of the foregut endoderm leads to organ-specific sub-domains that will later yield the thyroid, thymus, lungs, liver, stomach, and pancreas (**Fig. 1.2A**) (Davenport et al., 2016; Lewis and Tam, 2006). The patterning of the dorsal and ventral pancreas begins at ~E8.0, marked by the induction of *Ptf1a* (*pancreas associated transcription factor 1a*). This is followed quickly by the induction of *Pdx1* (*pancreatic and duodenal homeobox 1*) expression (~E8.5-E9.0) via signals from the notochord (**Fig. 1.3A-C**). Pancreas organogenesis becomes apparent by E9.5, during the primary transition, when each distinct domain begins to thicken and protrude from the endoderm and into the neighboring mesenchyme (**Fig. 1.3D-F**) (Pan and Wright, 2011). At this stage, the pre-pancreatic epithelium is comprised primarily of multipotent progenitor cells (MPCs) capable of making all pancreatic cell types (Herrera, 2000; Pan and Wright, 2011). The dorsal and ventral pancreatic buds begin to express additional TFs critical for the continued progression of pancreas development and the differentiation of pancreatic acinar, ductal, and endocrine cells (Pan and Wright, 2011). It is important to note that these two regions of the pre-pancreas undergo independent developmental processes relying on similar but slightly different TF cascades prior to fusion (Jensen, 2004; Murtaugh, 2007; Zorn and Wells, 2009). Key factors involved in these TF cascades include *Mnx1* (*Motor Neuron and Pancreas Homeobox 1*), *Ptf1a* (*Pancreas Associated Transcription Factor 1a*), *Sox9* (*SRY-Box Transcription Factor 9*), *Nkx6.1* (*NK6 Homeobox 1*), *Foxa2* (*Forkhead Box A2*), *Gata4* (*GATA Binding Protein 4*), and *Gata6* (*GATA Binding Protein 6*). Each are important for the progression of primary to secondary multipotent progenitor cells (MPCs) (**Fig. 1.4**) (Pan and Wright, 2011).

As previously mentioned, pancreas specification relies on the signaling factors from the surrounding mesoderm-derived tissues, but an important shift in proximity to these also takes place (Seymour and Serup, 2019; Stemple, 2005; Wells and Melton, 1999). Signals from the notochord, a transient embryonic midline structure responsible for providing patterning signals to surrounding tissues, induce expression of *Pdx1* at E8.5 in the dorsal pancreas, and shortly after (~8.75-9.0), the notochord moves away from the dorsal pancreas region to be replaced by the dorsal aorta and mesenchyme (**Fig. 1.2D-K**) (Seymour and Serup, 2019). This shift in localization is important for both the hyper-vascularization of the pancreas by the dorsal aorta and for the associated signaling cues these endothelial cells provide to stimulate continued development of the pancreas (Seymour and Serup, 2019; Zorn and Wells, 2009). Expression of *Pdx1* in the ventral pancreas occurs soon after and is an early marker and key driver of pancreatic tissues. Mice lacking *Pdx1* exhibit pancreas agenesis, and sub-optimal levels result in early-onset hypoplasia of the dorsal bud and a complete lack of the ventral pancreatic bud (Ahlgren et al., 1998; Fujitani et al.,

2006; Holland et al., 2002; Jonsson et al., 1994; Offield et al., 1996). Thus, these spatiotemporally regulated niche signaling cues and localization are essential.

The secondary transition (E11.5-E16.5) is a phase of pancreas development marked by extensive growth, cellular differentiation, remodeling, and formation of the microlumen (**Fig. 1.5A-C**) (Pan and Wright, 2011). At a gross morphological level, the gut tube rotates at ~E11.5, bringing the ventral pancreatic bud in the proximity of the dorsal bud, allowing them to fuse and become a single organ pancreas (Seymour and Serup, 2019; Slack, 1995). During this phase of remodeling, the microlumen formation is observed at E11.0, with microlumen fusion beginning at E11.5 (Pan and Wright, 2011). Shortly after, the pancreatic epithelium is further segregated into the tip and trunk regions of the developing pancreas (**Figs. 1.5A and 1.6**). The tip region resides at the tip of the microlumenal spaces and consists of progenitors that give rise to acinar cells. Conversely, the trunk region contains bipotent progenitor cells that become fated for the ductal and endocrine cell lineages (Pan and Wright, 2011; Zhou et al., 2007). Many endocrine cells are simultaneously specified and cued for endocrine commitment by transient expression of *Neurog3* (*Neurogenin 3*). This triggers the expression of downstream TFs critical for establishing the endocrine-specific regulatory program and drive differentiation and delamination (**Figs. 1.4 and 1.6**) (Bechard et al., 2016; Pan and Wright, 2011). As the delaminated cells continue to differentiate, they cluster together away from the pancreatic ducts to form the islets of Langerhans (**Fig. 1.6**) (Pan and Wright, 2011).

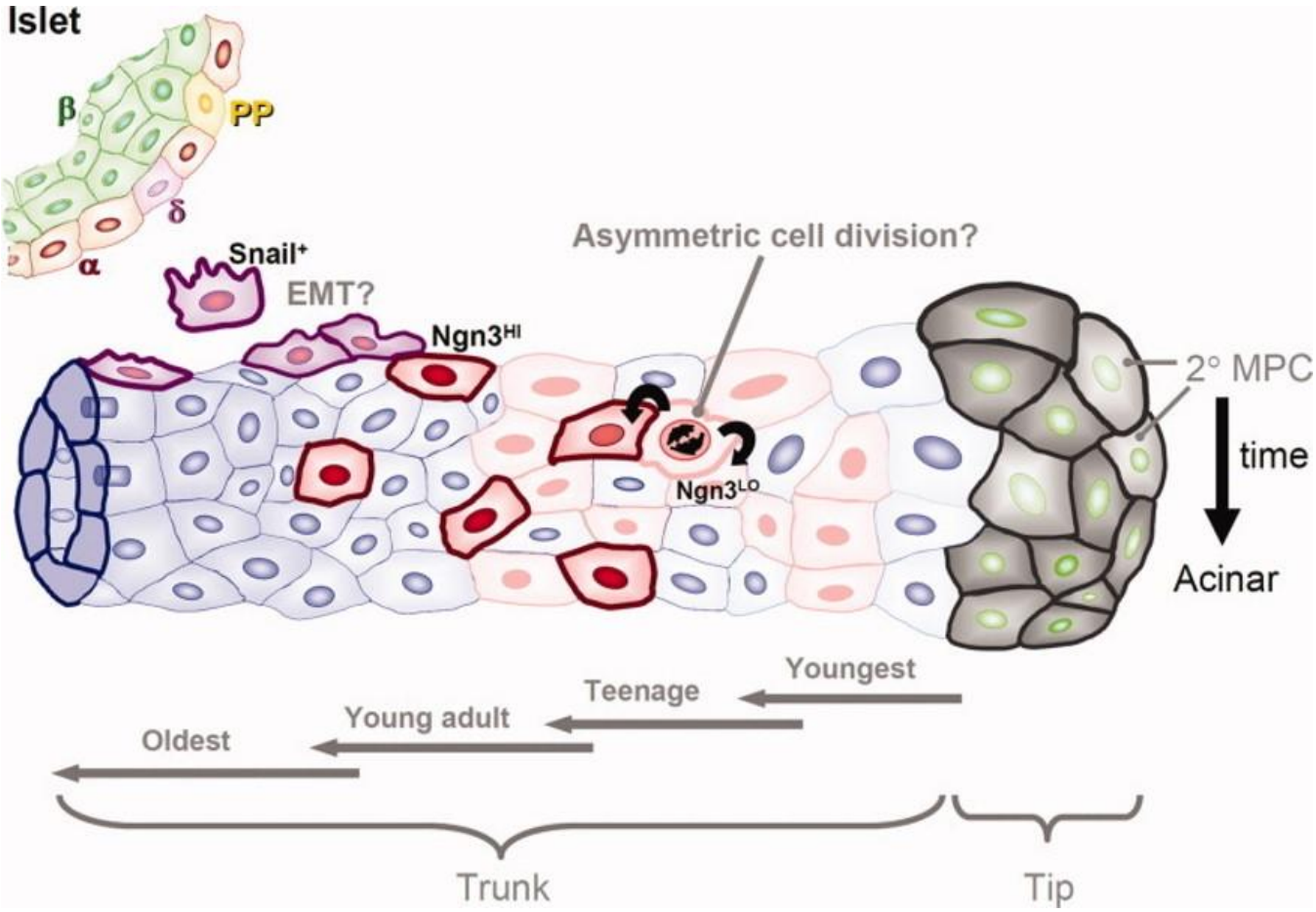


Figure 1.6. Model of epithelial organization during the secondary transition. A single piece of tip region epithelium from the secondary transition (E13.5) is depicted, consisting at this stage of two major domains: an MPC-containing tip domain (grey), and a trunk region harboring the endocrine/ duct bipotential progenitor pool (duct progenitors are shaded blue). The trunk domain is subdivided with respect to their age after birth from tip MPC. Trunk cells nearest the tip are youngest with older cells moving back down the trunk (teenage through oldest). Scattered cells within the trunk epithelium activate *Ngn3* expression, with $Ngn3^{LO}$ cells (light pink) representing a putative metastable, relative plastic, uncommitted but endocrine-biased mitotic state. Ngn3LO asymmetric cell division leads to one daughter having higher Ngn3 expression ($Ngn3^{HI}$: darker nucleus, red border) that becomes endocrine-committed, leaving a $Ngn3^{LO}$ progenitor available for more rounds of endocrine birth via production of additional $Ngn3^{HI}$ daughters. Committed $Ngn3^{HI}$ endocrine precursors rapidly activate Snail2 (purple cells) and escape the trunk epithelium, probably via epithelial–mesenchymal transition (EMT), before clustering to form the endocrine islets of Langerhans. This image was modified and used with permission Pan and Wright, 2011.

Endocrine specification and maturation

A key factor for driving the processes of endocrine cell specification, expansion, and differentiation during the secondary transition is *Neurog3*. *Neurog3* functions by activating the lineage-specific transcriptional program and is the master regulator of endocrine cell commitment (Rukstalis, et al., 2009). The spatial and temporal regulation of its expression is required for proper specification of the different islet cell types, as each cell type has specific windows of differentiation during development (Apelqvist et al., 1999; Schwitzgebel et al., 2000). Johannsson et al. further demonstrated this using an inducible *Pdx1*-driven *Neurog3* "addback" transgenic mouse model maintained in a *Neurog3*-null background. This approach allowed the authors to temporally control *Neurog3* expression during development. They found that each of the endocrine subtypes is generated in distinct competence windows, such that α-cells are produced first at ~E8.5, followed by β- and PP cells (~E10.5-12.5), with β-cells continuing to be prominently produced through E14.5 (Johansson et al., 2007). The differentiation of δ-cells falls into a later complacency window, with most of those cells being formed at E14.5 and later (Johansson et al., 2007). Similarly, clonal analyses have also suggested that *Neurog3*-positive progenitors are unipotent in nature (Desgraz and Herrera, 2009). Subsequent delamination of pro-endocrine cells likely requires an EMT transition, although the delaminating cells never fully lose their epithelial characteristics (**Fig. 1.6**) (Rukstalis and Habener, 2009). Ultimately, these cells transition away from the ductal epithelium of the trunk and form tightly organized islets (Miller et al., 2009).

Neurog3 expression has long been considered a marker of irreversible endocrine commitment. However, recent studies suggest that the cellular dynamics are much more nuanced. Bechard et al. have shown that the transition of *Neurog3*$^{+}$ cells from a mitotically active progenitor state into a fully committed pro-endocrine cell is *Neurog3* dosage-dependent. Exit from the progenitor state requires high levels of Neurog3 protein (Bechard et al., 2016; Bechard et al., 2017; Bechard and Wright, 2017). This phenomenon is discussed in greater detail in the following section. Given its prominent role in endocrine lineage specification, the mechanisms and factors that regulate *Neurog3* expression, and those regulated by *Neurog3*, have been heavily studied. Neurog3, in combination with actions of its downstream transcriptional targets (*Insm1, Neurod1, Pax4, Pax6, Rfx6, Nkx2.2*, and others), promotes the differentiation and maturation of different endocrine cell types (**Fig. 1.4**).

TFs expressed by Neurog3 are often cell-type specific and responsible for regulating a narrower sub-network of genes and more specific developmental processes. In addition to having functions important for promoting the differentiation into a certain cell type, many are also important for the repression of genes that correspond to other cell types. For example, the transcription factors *Paired box 4* (*Pax4*) and *Aristaless related homeobox* (*Arx*) have a mutually repressive relationship (Courtney et al., 2013). Expression of *Pax4* is necessary for proper differentiation of β-cells, and *Arx* is critical for α-cells. They both inhibit the expression of the other to prevent the activation of genes needed for establishing the cellular identity of the opposing cell type. Mis-expression of either of these factors in the opposing endocrine cell lineage (i.e., α- or β-cells) results in defects of α- or β-cell identity (Collombat et al., 2005; Collombat and Mansouri, 2009; Collombat et al., 2009; Courtney et al., 2013).

The transcription factors *MafA* and *MafB* function similarly and are also important for establishing and maintaining α- or β-cell identities, respectively. *MafB* expression is observed in both cell lineages during development; however, it becomes restricted to α- cells in adult mice, although low levels of expression are maintained in β-cells of adult humans (Artner et al., 2010; Conrad et al., 2016; Nishimura et al., 2006). In contrast, MafA is considered a marker of mature β-cells and functions primarily to regulate the terminal differentiation of β-cells, not α-cells (Artner et al., 2008; Nishimura et al., 2006). MafA also plays a critical role in β-cell function, regulating the expression of *Ins* and glucose-stimulated insulin secretion (GSIS) (Artner et al., 2008; Matsuoka et al., 2004; Matsuoka et al., 2007; Nishimura et al., 2006).

Furthermore, and much like *MafA*, endocrine-specific TFs do not always function solely to promote or inhibit other TFs within the network. These factors are often also important for the cellular function of mature cells, as well as maintaining their cellular identity. This includes a cell's ability to sense circulating glucose levels and to respond appropriately through the secretion of the necessary hormones (*e.g.*, insulin or glucagon). In the case of β-cells, this mechanism is known as glucose-stimulated insulin secretion (GSIS) and is a key feature of functionally mature cells (**Fig. 1.7**). For example, *Pdx1*, *Hnf4a*, and *Neurod1* have also been shown to directly bind and regulate *Ins* expression (Artner et al., 2008; Gu et al., 2010b; Iype et al., 2005; Schaffer et al., 2013; Sharma et al., 1999; Wang et al., 2000; Zhang et al., 2005). They work to govern the cellular processes necessary for glucose sensing and insulin secretion of β-cells. Others, such as *Pax6 Isl1*, and *Brn4,* regulate the expression of α-cell specific genes, including *Gcg*. The expression of functional genes important for establishing alternative cell types is also governed

by many of these factors. *Pdx1*, for instance, has been shown to maintain beta cell identity and function through the active repression of α-cell specific genes and activation of β-cell genes. β-cell specific loss of *Pdx1* results in the upregulation of α-cell genes like *MafB* and *Gcg*, in addition to the downregulation of the β-cell specific genes *Glut2*, *Ins1/2*, and *G6pc2* (Gao et al., 2014). These transcriptional changes are paralleled by shifts towards a more α-cell like ultrastructure and physiological responses (Gao et al., 2014).

These are just a few examples in a long line of evidence indicating that the regulation of a cell's transcriptional profile is critical for maintaining these functional features. The genetic relationships described here also highlight the idea that negative regulatory interactions are just as important as positive interactions in establishing the GRN of endocrine cells. In turn, they are critical for the proper differentiation of each cell type, as well as the function and maintenance of cellular identity.

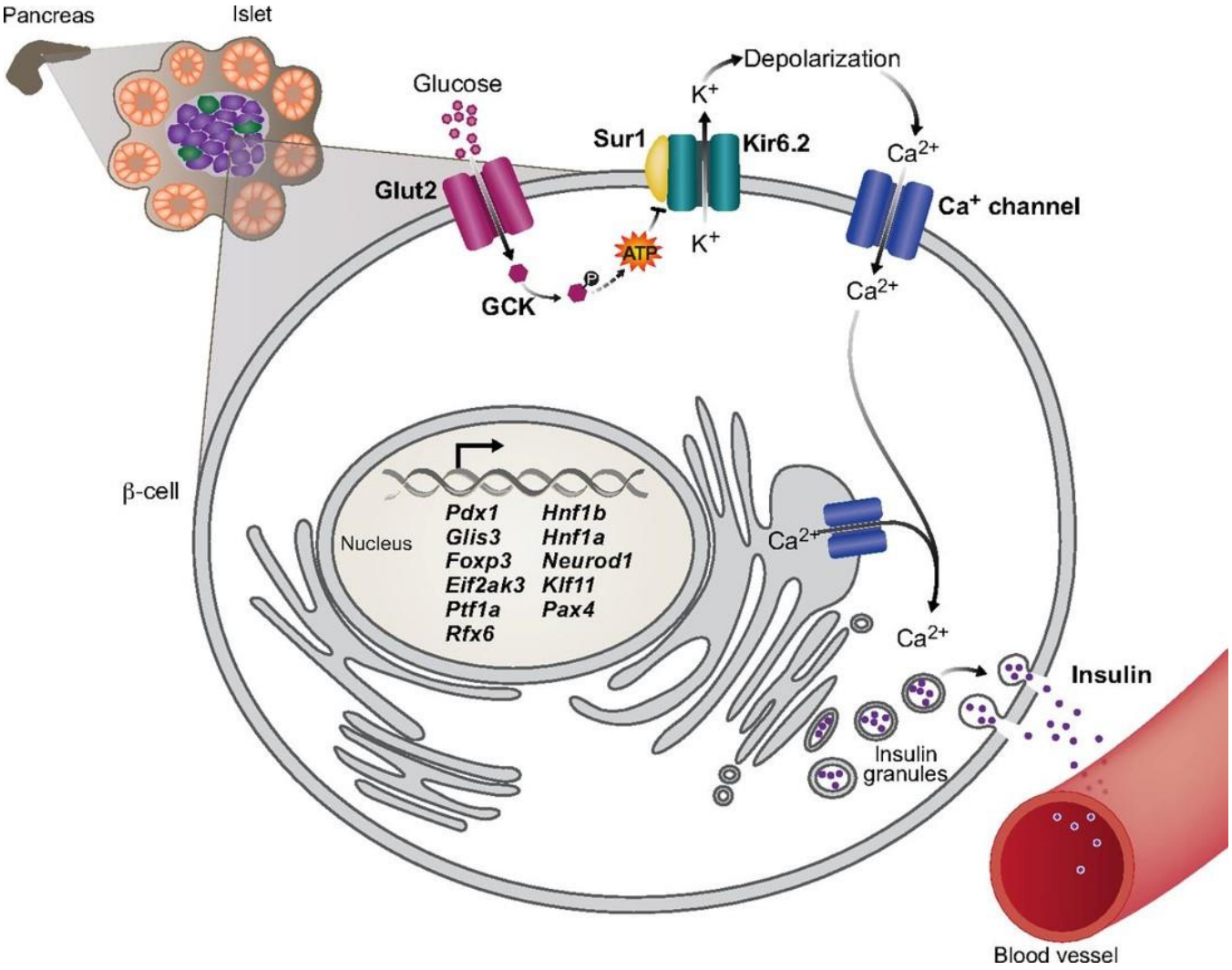
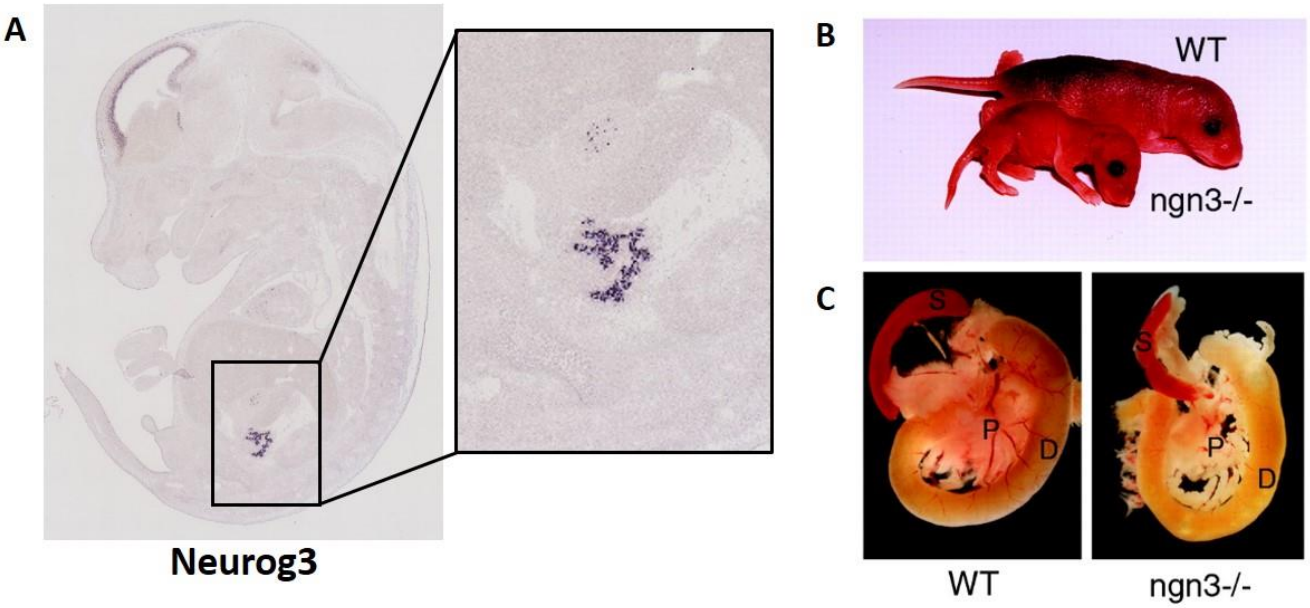


Figure 1.7. Functional β-cells respond to increasing glucose levels by increasing insulin secretion. In glucose-stimulated insulin secretion (GSIS), glucose is transported into the cell via glucose transporters [e.g. Glut1 (Slc2a1) or Glut2 (Slc2a2), pink], where it is phosphorylated by glucokinase (GCK) and converted into ATP by subsequent metabolic reactions. Rising ATP levels (e.g. rising ATP:ADP ratios) trigger the closure of potassium channels [Sur1 (Abcc8) and Kir6.2 (Kcnj11) subunits], membrane depolarization, and the opening of calcium channels (blue). The resultant rise in intracellular calcium levels triggers the exocytosis of insulin-containing granules and hence leads to increased insulin levels in adjacent blood vessels. Human genetic studies of maturity onset diabetes of the young (MODY) patients have identified a number of mutations that trigger diabetes, including those in genes encoding transcription factors (depicted in the nucleus) and components of the GSIS pathway indicated in this figure. Figure was reprinted with permission from Paglia and Melton, 2013.

Neurog3 is a master regulator of pancreatic endocrine cell development.

Neurog3 is a bHLH transcription factor first identified as a differentiation factor for neurogenic lineages within the peripheral and central nervous system (Sommer et al., 1996). In neurons, its expression is generally initiated at E9 and becomes downregulated as the neuronal cells become more differentiated (Sommer et al., 1996). Similarly, in the developing pancreas, *Neurog3* becomes lowly expressed as early as ~E9.5 and highly expressed during the secondary transition beginning at ~E12.0, peaking at E15.5. Its expression becomes diminished by E17.5, with little to no expression in adult cells (**Fig. 1.8A, D-O**) (Apelqvist et al., 1999; Gradwohl et al., 2000; Gu et al., 2003). However, more recent studies have indicated that at the level of individual cells, *Neurog3* can be lowly expressed in later stages, including adults, and is also important for the maturation and maintenance of endocrine cell identity (Wang et al., 2009). Several TFs that are expressed during earlier pancreas specification, such as Pdx1 (Burlison et al., 2008), Sox9 (Lynn et al., 2007), Nkx2.2 (Anderson et al., 2009; Papizan et al., 2011; Sussel et al., 1998), Hnf6 (Jacquemin et al., 2000; Lee et al., 2001), Hnf1a and Hnf3b (Lee et al., 2001), as well as members of the Notch signaling pathway like *Hes-1* (Apelqvist et al., 1999; Georgia et al., 2006; Jensen et al., 2000a; Jensen et al., 2000b; Lee et al., 2001; Li et al., 2015; Qu et al., 2013), have been shown to govern the expression of *Neurog3*. These factors are important for the necessarily tight regulation of the levels and timing of *Neurog3* expression.

Neurog3 is the master regulator of and required for endocrine cell specification (Gradwohl et al., 2000; Pan and Wright, 2011; Schwitzgebel et al., 2000). *Neurog3*-deficient mice are stunted, devoid of all endocrine cells, develop diabetes, and die within three days of birth (**Fig. 1.7B-C**) (Gradwohl et al., 2000). Gu et al. used lineage tracing studies to demonstrate that all endocrine cells are derived specifically from *Neurog3*$^{+}$ progenitor cells (Gu et al., 2002). Furthermore, over-expression of *Neurog3* using the *Pdx1* promoter leads to premature differentiation of endocrine progenitor cells, resulting in underdeveloped pancreatic tissues comprised mostly of endocrine cells (Apelqvist et al., 1999; Schwitzgebel et al., 2000). However, the rapid increase in differentiation from ectopic expression of *Neurog3* comes at the expense of cellular proliferation and expansion. The progenitor pool becomes quickly exhausted, and an overall reduction in the number of exocrine cells is observed (Apelqvist et al., 1999; Petri et al., 2006; Schwitzgebel et al., 2000). Thus, a balance in high and low dosage levels of *Neurog3* necessary for producing an adequate number of endocrine cells.


A
Neurog3
B
WT
ngn3-/-
C
S
P
D
WT
ngn3-/-

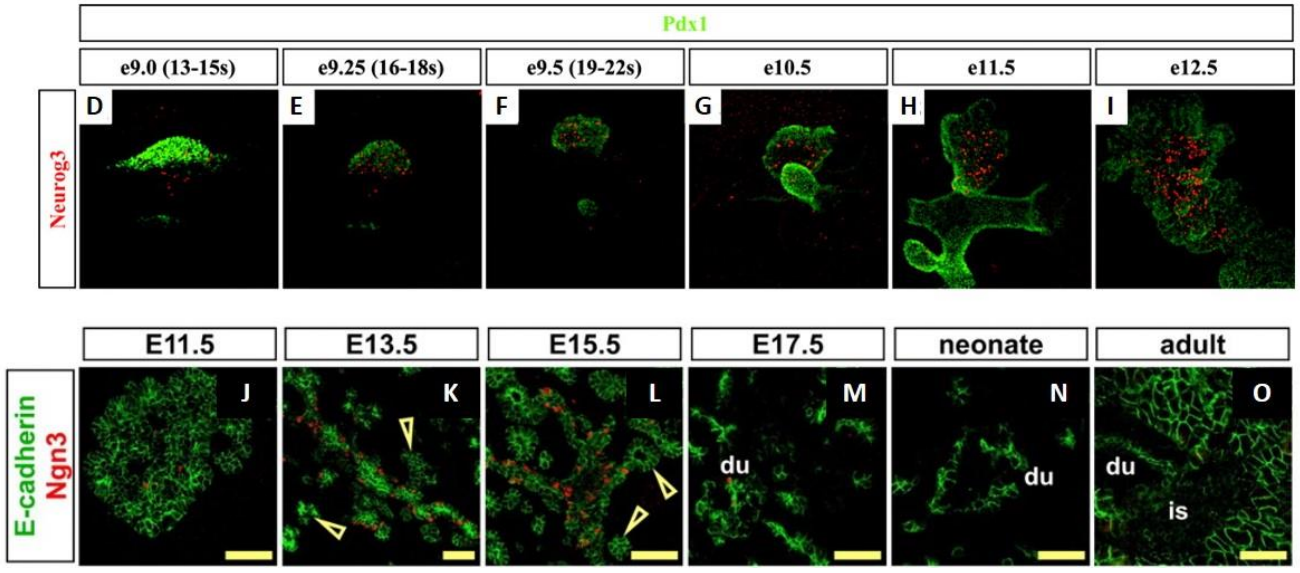

Pdx1
e9.0 (13-15s)
e9.25 (16-18s)
e9.5 (19-22s)
e10.5
e11.5
e12.5
Neurog3
D
E
F
G
H
I
E11.5
E13.5
E15.5
E17.5
neonate
adult
E-cadherin
Ngn3
J
K
L
M
N
O
du
du
du
is

Figure 1.8. Neurog3 expression in developing pancreatic endocrine cells. **Top**: (**A**) In situ hybridization map of Neurog3 expression at E14.5 in wild type C57BL/6J animals. Data were obtained from GenePaint digital atlas (https://gp3.mpg.de/viewer/setInfo/DA126/12). (**B**) This image of two 3-day-old littermates, one wild-type (upper) and one ngn3-deficient (lower), shows the reduced size of the mutant animal. (**C**) Pancreas (p), spleen (s), and duodenum (d) of postnatal day 2 wild-type and ngn3 mutant mice. The mutant pancreas does not have any obvious morphological abnormality. **Bottom**: (**D-I**) Collapsed image stacks of wild-type whole embryos (E9.0–E10.5) or dissected pancreata (E11.5 and E12.5). To visualize the pancreatic epithelium, all embryos were stained for Pdx1 and only the dorsal pancreata are shown at E12.5. In all panels, anterior is to the right, posterior is to the left, dorsal is above, and ventral is below. (**D-E**) The proendocrine gene Neurog3 can be detected robustly at E9.0. Most of the Neurog3-positive cells reside within the *Pdx1* expression domain, but a number of cells can also be observed in the endoderm between the two pancreas domains. This is also true at E9.25. (**F-I**) From E9.5 to E12.5, *Neurog3* becomes restricted to the pancreas endoderm. (At later stages not described here, *Neurog3* will be expressed in the duodenum.) (**F**) The first ventral Neurog3 expression can be detected at E9.5. (**I**) At E12.5, Neurog3 is expressed in a scattered pattern in the central part of the epithelium. (**J-K**) Confocal immunofluorescence photomicrographs at equivalent stages, for the pan-epithelial marker E-cadherin (green) and the islet precursor marker Ngn3 (red). *Ngn3* expression is rare at E11.5, dramatically peaks during the secondary transition (E13.5-E15.5) and declines again at E17.5, becoming undetectable in neonatal and adult pancreas. Arrowheads indicate proto-acinar clusters at the periphery of the branched epithelium, from which Ngn3expression is consistently excluded. These images were modified and reprinted with permission from Gradwohl, et al., 2000, Jorgensen, et al., 2007, and Murtaugh, et al., 2007.

The importance of tightly regulating the *Neurog3* expression point has recently been further elucidated by recent works. Bechard et al. demonstrated the effects of varying expression levels within developing endocrine progenitor cells using a sensitive *Neurog3* BAC transgenic reporter in combination with real-time analysis of mitotic events (**Fig. 1.9**). The authors found that progenitor cells expressing low levels of *Neurog3* (*Ngn3^{LO}*) remained mitotically active but were biased towards the endocrine lineage. Conversely, cells with high levels of expression (*Ngn3^{HI}*) undergo an irreversible commitment toward the endocrine lineage, consistent with previous reports (Bechard et al., 2016; Wang et al., 2010). In short, *Ngn3^{LO}* cells maintain a proliferative state to expand the progenitor cell population and generate an adequate number of endocrine cells, while *Ngn3^{HI}* cells are instructed to undergo cell cycle exit and terminal endocrine cell differentiation necessary for generating fully functional mature islet cells (**Figs. 1.6 and 1.9**) (Bechard et al., 2016). Subsequent studies determined that Neurog3 stability is regulated by cell cycle-dependent protein phosphorylation, with high levels of Neurog3 phosphorylation being associated with rapid protein degradation. These same studies found that the dephosphorylation of Neurog3 leads to an increase in activity (due to an increase in stability), the enhancement of cell cycle exit, and promotion of differentiation (**Fig. 1.10**) (Azzarelli et al., 2017; Krentz et al., 2017).

Ngn3^{HI} endocrine committed cells to undergo further differentiation into their terminal islet cell types. This process is triggered by the *Neurog3*-dependent activation of numerous downstream TFs. Genetic perturbation, biochemical assays, and sequencing studies have allowed for the identification and validation of many direct target genes, including *Insm1* (Breslin et al., 2002; Goto et al., 1992; Mellitzer et al., 2006), *Neurod1* (Gradwohl et al., 2000; Huang et al., 2000), *Pax4* (Smith et al., 2003; Sosa-Pineda et al., 1997), *Pax6* (Gradwohl et al., 2000; Heremans et al., 2002), *Rfx6* (Soyer et al., 2010), and *Nkx2.2* (Churchill et al., 2017). Neurog3 has also been found to regulate its own expression (Ejarque et al., 2013). The function of many direct target genes has been studied to varying degrees, but gaps remain in our understanding of their individual roles within the transcriptional network. My studies have been focused on further elucidating those functions of *Insm1*, *Neurod1*, and *Pax6,* three *Neurog3*-dependent TFs.

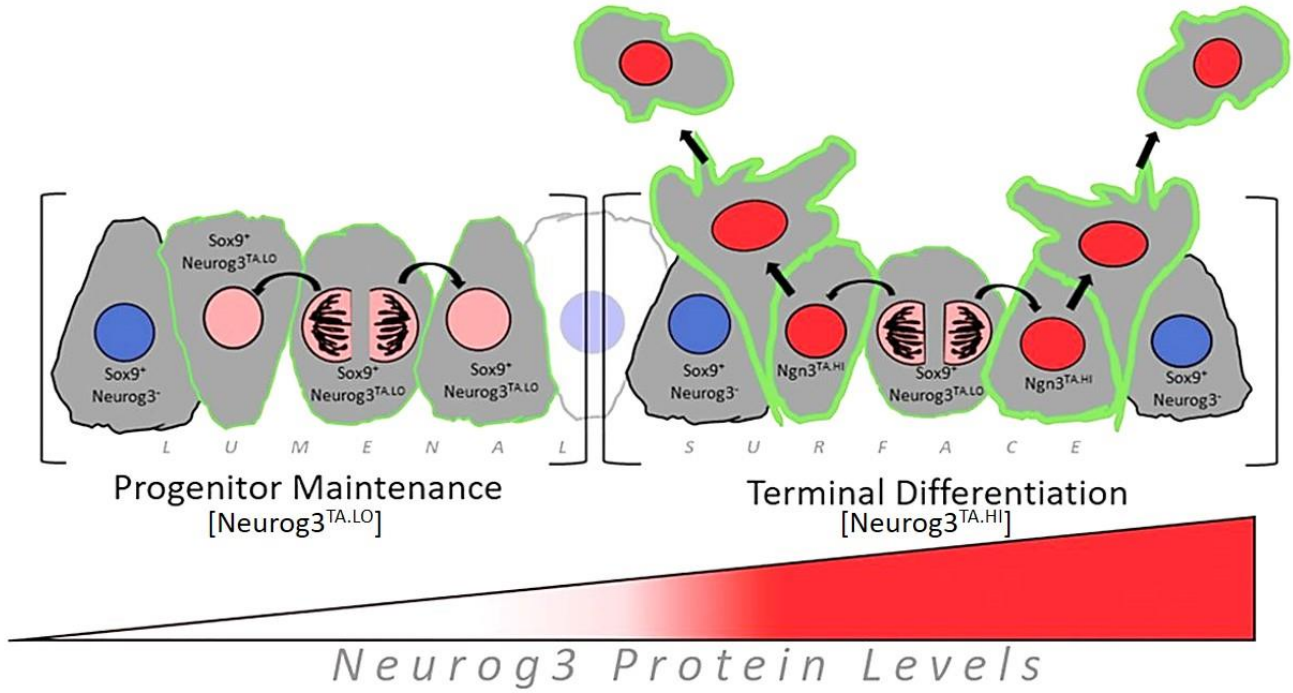


Figure 1.9. Model depicting the behavior of Neurog3TA.LO cells in the pancreatic epithelium. The mitotic index, and other differentiation behaviors, of progenitor cells is effected by the level of Neurog3 protein during pancreas development. Putative endocrine progenitor cells expressing low levels of Neurog3 ($Ngn3^{LO}$) remain in a proliferative state at the luminal surface. Cells expressing high levels of Neurog3 ($Ngn3^{HI}$) become fully committed to the endocrine lineage and begin to denominate away from the Trunk epithelium. This figure modified and reprinted with permission from Bechard et al., 2016.

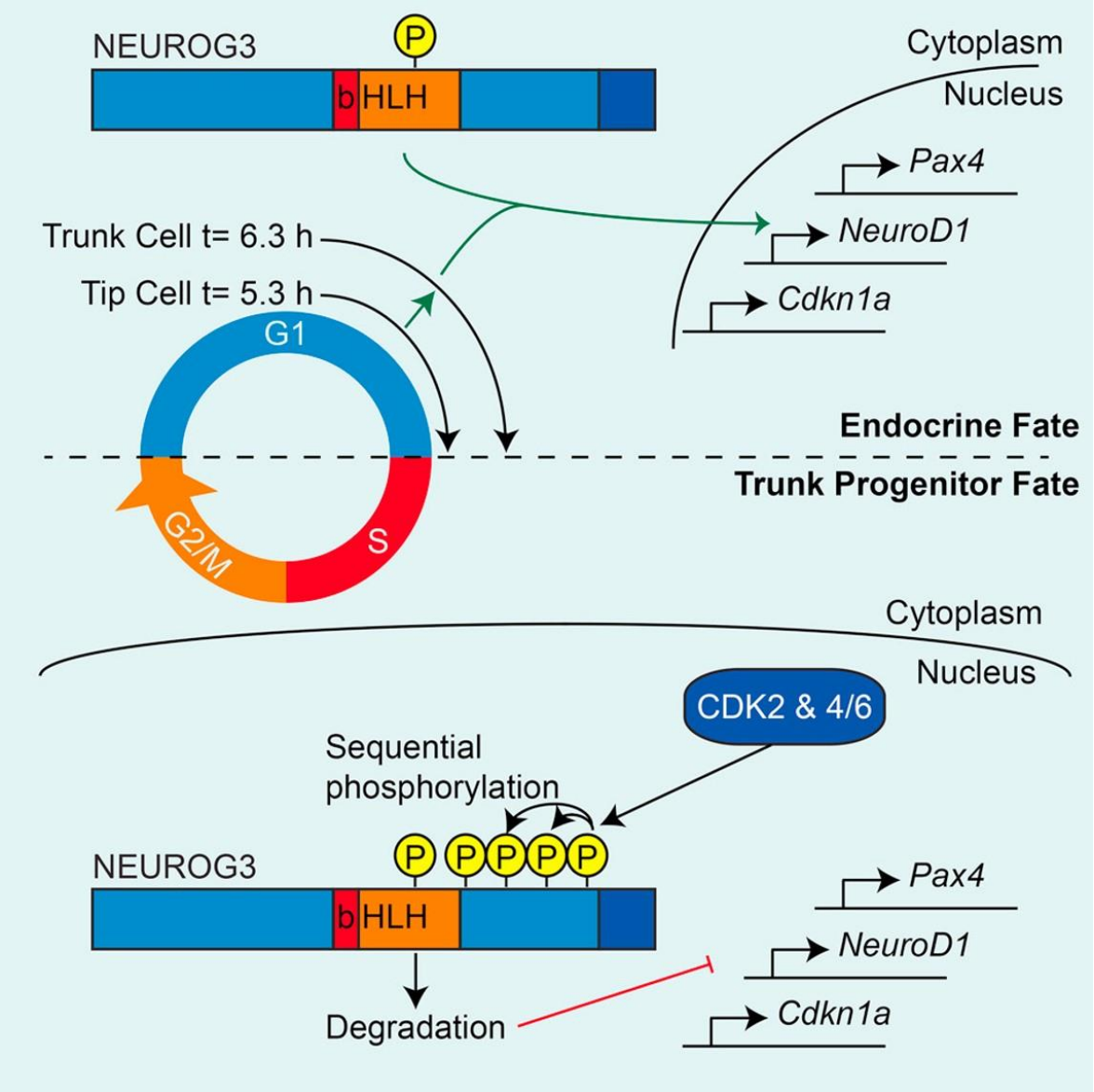


Figure 1.10. Neurog3 de-phosphorylation enhances cell-cycle exit. In rapidly dividing pancreatic progenitor cells, there is high activity of Cdks. These kinases phosphorylate Neurog3, resulting in its degradation. During development, there is a lengthening of G1-phase from 4.5 hours at E11.5 to 7.2 hours at E13.5. This change in cell cycle length would result in less activity of Cdks, reduced phosphorylation of Neurog3, and differentiation to the endocrine cell fate. This figure modified and reprinted with permission from Krentz et al., 2017.

Insm1 regulation of endocrine cell development and mature islet cell function.

Insulinoma-associated 1 (*Insm1*) is an intronless gene that encodes for a transcription factor. The highly evolutionarily conserved protein contains a zinc finger DNA-binding domain at the C-terminus consisting of five Cys_2-His_2 (C_2H_2) zinc finger motifs and a putative pro-hormone domain (Lan and Breslin, 2009; Lan et al., 1994). *Insm1* (previously *IA-1*) is expressed robustly in embryonic stages of the developing central nervous system, neuroendocrine, and pancreatic tissues (Lan and Breslin, 2009). Specifically, *Insm1* is expressed in the spinal cord, cerebellum, forebrain, midbrain, hindbrain, and olfactory epithelium as indicated through multiple in situ hybridization experiments (**Fig. 1.11A-C**) (Duggan et al., 2008; Mellitzer et al., 2006). *Insm1* expression is highest in proliferative neuronal progenitor cells and in nascent neurons and is completely absent in fully mature cells (Duggan et al., 2008). Additionally, it is also known to be a specific biomarker of tumors originating from neuroendocrine tissues, where it was originally identified (Goto et al., 1992; Lan et al., 1994; Mahalakshmi et al., 2020). Its expression has specifically been highly correlated with small-cell lung cancers (Mukhopadhyay et al., 2019; Pedersen et al., 2003).

In the developing endocrine pancreas, *Insm1* expression is reported beginning at ~E10.5 in scattered pro-endocrine cells and peaking at E15.5 (Mellitzer et al., 2006). *In situ* hybridization experiments show that *Insm1* overlaps with *Neurog3* expressing cells and is exclusively found in endocrine cells, excluded from acinar and ductal cell lineages (Mellitzer et al., 2006). Direct regulation of *Insm1* by Neurog3 was determined through reporter assays measuring *Insm1* promoter-driven luciferase activity, showing an increase in activity when cells were co-transfected with a *Neurog3* expressing plasmid (Mellitzer et al., 2006). Furthermore, *Insm1* expression in endocrine cells is directly activated by another Neurog3-dependent TF, Neurod1. Relative to that of Neurog3, positive regulation by Neurod1 is only moderate, and *Insm1* expression is not lost in *Neurod1* knockout (KO) mice as it is in *Neurog3* KO models (Breslin et al., 2003; Mellitzer et al., 2006; Zhu et al., 2002). Reports have suggested that these results indicate a feedforward loop of gene expression, where Neurod1 helps to maintain or increase *Insm1* levels once *Neurog3* levels have diminished (Mellitzer et al., 2006). Interestingly, other studies have used luciferase reporter assays in HEK293 cells to demonstrate that Insm1 may function, at least in part, by directly binding and repressing *Neurod1* expression (Breslin et al., 2002; Liu et al., 2006).

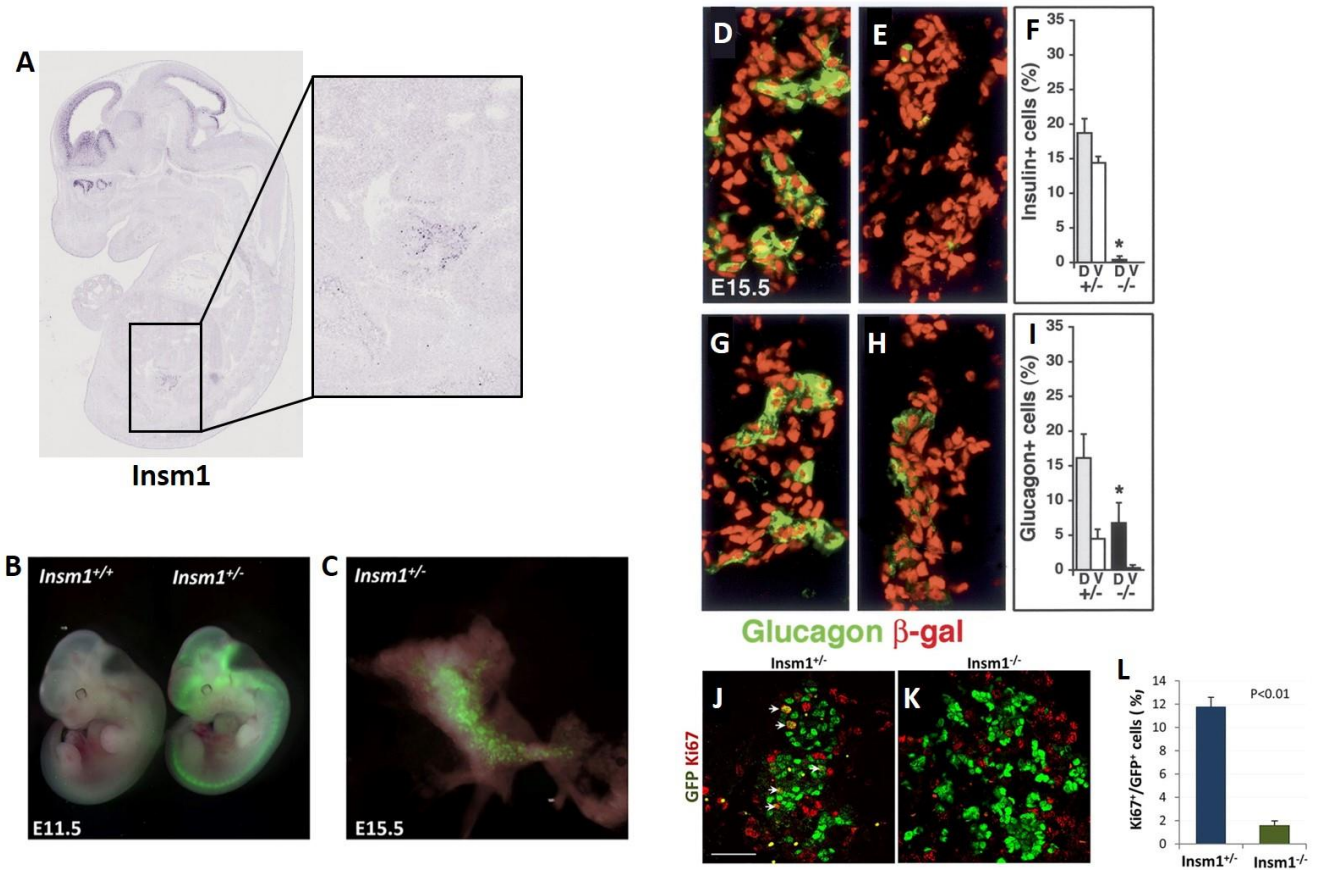


Figure 1.11. Insm1 expression and regulation of endocrine cell development. (**A**) In situ hybridization map of *Insm1* expression at E14.5 in wild type C57BL/6J animals. Data were obtained from GenePaint digital atlas (https://gp3.mpg.de/viewer/setInfo/MH3077/10). (**B**) Green fluorescence in a whole mouse embryo carrying the *Insm1*^{GFPCre} allele at E11.5 broadly marks the neural system. (**C**) Green fluorescence in a pancreas at E15.5 marks pre-endocrine cells. Fluorescence images were overlaid with images taken with white light. Immunohistological analysis of the developing dorsal pancreas of (**D,J**) *Insm1*$^{lacZ/+}$ and (**E,K**) *Insm1*$^{lacZ/lacZ}$ mice at E15.5; the genotypes are indicated by +/− and −/−, respectively. The antibodies used were directed against β-galactosidase (red) and insulin (green) (**D,E**), and β-galactosidase (red) and glucagon (green) (**J,K**). (**F,L**) Proportions of β-galactosidase^{+} cells that express insulin (**F**) or glucagon (**L**) in the dorsal (D) and ventral (V) pancreas of *Insm1*$^{lacZ/+}$ and *Insm1*$^{lacZ/lacZ}$ mice at E15.5. Asterisks indicate p-values of <0.01. Bar, 20 μm. This figure modified and reprinted with permission from Gierl, et. al., 2006 and Osipovich et. al., 2014.

Murine *Insm1* knockout models are embryonically lethal and display defects in the development of the central and peripheral nervous system, as well as pancreas development (Farkas et al., 2008; Gierl et al., 2006; Wildner et al., 2008). In the developing nervous system, its acts as a regulator of neurogenic progenitor cell differentiation and of the formation and proliferation of basal progenitor cells (Farkas et al., 2008; Rosenbaum et al., 2011; Wildner et al., 2008). Duggan et al. concluded that during neuronal development, *Insm1* functions in part to regulate the termination of proliferation during terminal divisions based on the decreased expression of *Insm1* later stages of non-proliferative cells (Duggan et al., 2008). Subsequent studies provided support for this hypothesis by showing that Insm1 can function beyond transcriptional regulation by binding directly to Cyclin D1, interrupting the cyclin D1/CDK4 interaction. This leads to cell cycle arrest and a switch from cellular proliferation to cellular differentiation (Liu et al., 2006; Zhang et al., 2009). Additionally, studies in *C. elegans* lacking *egl-46* (the *Insm1* homolog) have demonstrated impaired axonal outgrowth and migration, suggesting an additional role for *Insm1* in regulating cell migration of nascent neurons (Desai et al., 1988; Desai and Horvitz, 1989).

In the pancreas, mice lacking Insm1 have disrupted endocrine cell development, marked by impaired islet formation, and altered hormone expression at E18.5. There is a specific loss of α- and β-cells and an increase in PP cells (**Fig. 1.11 D-I**) (Gierl et al., 2006; Osipovich et al., 2014). The developmental program of endocrine cell differentiation in mutant mice is delayed, and terminal differentiation is impaired, as highlighted by the altered ratio of hormone expressing cells. This is particularly true of the β-cell lineage, as only a small subset of pro-endocrine cells is fully converted to insulin-expressing β-cells (Gierl et al., 2006). *Insm1* KO endocrine cells also have reduced proliferation rates and fail to move away from the pancreatic ductal cells during development (**Fig. 1.11 J-L**) (Osipovich et al., 2014). These morphological phenotypes observed in KO animals are paralleled by robust changes in the gene expression profiles. For example, genes previously shown to be dysregulated in *Insm1* KO embryos include pancreatic hormones (*Ins1/2*), other TFs (*Bhlhe22, Sox6, Ripply3, MafA, MafB*), and various signaling factors (*Hes1, Notch1, Fgfr1, Bmp1*) (Osipovich et al., 2014).

Unlike some neuronal lineages, *Insm1* expression is maintained throughout development and persists in mature pancreatic endocrine cells (Gierl et al., 2006; Mellitzer et al., 2006). Conditional ablation of *Insm1* in adult β-cells results in reduced endocrine cell size and defects in islet morphology, with α-cells localized throughout the islets rather than the periphery (Jia et al., 2015). Adult animals display impaired cellular function and GSIS, in

addition to a shift to a more immature gene expression profile (Jia et al., 2015). This includes the dysregulation of *Glut2*, *Pfkfb3*, *Prckb*, *Ak5*, and *Glp1r*, which all have functions in glucose-dependent insulin secretion, consistent with disrupted secretory functions (Jia et al., 2015). Tao et al. recently reported that haplo-insufficiency of *Insm1* delays the cell-cycle and disrupts early expansion of β-cells during development, followed by decreases in β-cell mass and defects in glucose tolerance (Tao et al., 2018). These results suggest that *Insm1* has concentration-dependent functions in mature β-cells.

Although much has been learned about the functions of *Insm1* in pancreatic endocrine cells over the last few decades, the precise nature of its role within the regulatory network has not been fully elucidated. Additional experiments focused on tissue-specific functions and genetic interactions are necessary to better understand the molecular mechanisms underlying the morphological phenotypes observed.

Significance of Neurod1 in endocrine cell differentiation and maturation.

Neurod1, or *Neuronal Differentiation 1*, is a member of the basic helix-loop-helix (bHLH) TF family (Naya et al., 1995). TFs of the bHLH family have long been known to play an important role in the development and specification programs of several tissue types (Jan and Jan, 1993; Weintraub, 1993). These proteins contain a basic (b) domain that is associated with DNA binding and an HLH domain containing two alpha-helices linked by a looping region critical for protein dimerization (Bertrand et al., 2002; Dennis et al., 2019; Jones, 2004). Proteins in this class can be further subdivided based on their expression patterns, were ubiquitously expressed bHLH genes are of the Class I subtype, and those with tissue-specific expression fall under the Class II subtype; *Neurod1* is a Class II bHLH (Dennis et al., 2019). These TFs bind specific regions of DNA of target genes, known as E-boxes (enhancer box), by forming homo- and hetero-dimer complexes with other bHLH members to facilitate transcription (Jones, 2004; Weintraub, 1993).

Neurod1 was first discovered via a yeast two-hybrid screen, which utilized a hamster insulinoma cell cDNA library to identify β-cell specific bHLH TFs that bound the *insulin* (*Ins*) gene E-box. Direct interaction between Neurod1 and the *Ins* E-box, as well as transcriptional activation, was later confirmed via a DNA electrophoretic mobility shift assay (EMSA) and biochemical reporter assays (Naya et al., 1995). While Neurod1's N-terminal is less well conserved between mouse and frog (58%), the critical bHLH C-terminal end is highly conserved between species (98%), indicating its functional role is also likely conserved (Lee et al., 1995). Much like *Insm1*, *Neurod1* is

expressed at high levels during embryogenesis in the developing nervous system and endocrine pancreas, particularity in α- and β-cells **(Fig. 1.12A**) (Naya et al., 1995). Early Northern Blot experiments revealed that *Neurod1* expression is restricted to the endocrine cell lineage of the pancreas and otherwise only found in the brain tissues and weak expression in the intestine (Naya et al., 1995). Its expression in pancreatic endocrine progenitor cells can be observed as early as E9.5 and persists in all endocrine cell types throughout development and adulthood (Naya et al., 1997). Similar expression patterns are also observed in human islet cells during development (**Fig. 1.12B**) (Lyttle et al., 2008).

Two independent studies used gene targeting methods to replace the native *Neurod1* allele with a LacZ reporter and create null mouse models that allow for identification of *Neurod1*-expressing cells (Miyata et al., 1999; Naya et al., 1997). Both reports replicated the expression patterns described in Naya et al., 1995, and demonstrated that animals lacking *Neurod1* had deficits in neuronal and pancreatic development. In the developing central nervous system, defects were most apparent in the granule cells of the dentate gyrus (DG) in the hippocampus and in the cerebellum. These cells failed to properly regulate necessary ion channels, were unable to properly differentiate, and underwent premature cell death (Liu et al., 2000; Miyata et al., 1999). More recent studies have observed similar phenotypes in developing DG and vestibulocochlear ganglion (VCG) neurons, in addition to the failure of VGC neurons to migrate successfully (Kim, 2013). Conversely, when *Neurod1* is over-expressed in developing neurons of *Xenopus*, neural precursors undergo premature differentiation. Ectopic expression of *Neurod1* in embryonic non-neuronal ectoderm or neural crest cells has also been shown to drive their conversion to a neuronal cell fate (Lee et al., 1995). Collectively, these data suggest that *Neurod1* plays a role in the terminal differentiation of neurons at a specific stage of development. *Neurod1* has also been shown to be a marker of small cell lung cancer (SCLC). Ikematsu et al. showed that SCLC tumors with higher *Neurod1* expression were in an extensive disease state and had increased migratory activity (Ikematsu et al., 2020). The authors hypothesized that Neurod1 might regulate the migration and epithelial-mesenchymal transition (EMT) activity in these cells by inducing the expression of ZEB2 (zinc finger E-box–binding homeobox 2) (Ikematsu et al., 2020). This would support the argument that Neurod1 plays a role in the metastasis of SCLC, as well as general migratory functions in developing neuronal cells (Ikematsu et al., 2020).

Newborn KO animals display severe diabetes, marked by hyperglycemia and an inability to respond to insulin, and die within 3-5 days after birth as a result (Miyata et al., 1999; Naya et al., 1997). Null animals also have defects in islet architecture, where KO islets are only able to form small, disorganized clusters with a disproportionate ratio of hormone expressing cells, observable by E14.5 (Naya et al., 1997). The inability of cells to properly migrate towards each other and tightly cluster into islets parallels Neurod1's regulatory functions of migration activity in developing neurons and SCLC tumor cells (Ikematsu et al., 2020; Kim, 2013). Furthermore, the development of endocrine cells was arrested at a stage just after the segregation of α-, β-, and δ-cell lineages, as was their cellular expansion. Between E17.5 and P0, the overall number of endocrine cells is reduced, and by birth, KO animals contain 60% fewer endocrine cells than control litter mates (Naya et al., 1997). Increases in apoptosis have been implicated in the decrease in endocrine cell number, although a more recent study has suggested that the loss of proliferation is likely to be the primary cause (**Fig. 1.12C**) (Naya et al., 1997; Romer et al., 2019). Whether due to increased cell death or decreased proliferation, the number of β-cells was reduced, but *Ins* expression was still observed. This indicates that while Neurod1 is a positive regulator of *Ins*, it is not necessary, and this function is at least partially compensated for by other TFs.

Neurod1 maintains expression in adult β- and α-cells and is necessary for mature cellular function (Mastracci et al., 2013; Naya et al., 1997). Conditional ablation of *Neurod1* in the adult β-cell results in immature β-cells, and animals are hyperglycemic (Gu et al., 2010a). These animals are glucose intolerant due to a blunted glucose-stimulated insulin secretion response (Gu et al., 2010a). While studies have demonstrated a clear role for Neurod1 in both the development and function of endocrine cells, our understanding of its specific regulatory roles and interactions remains incomplete.

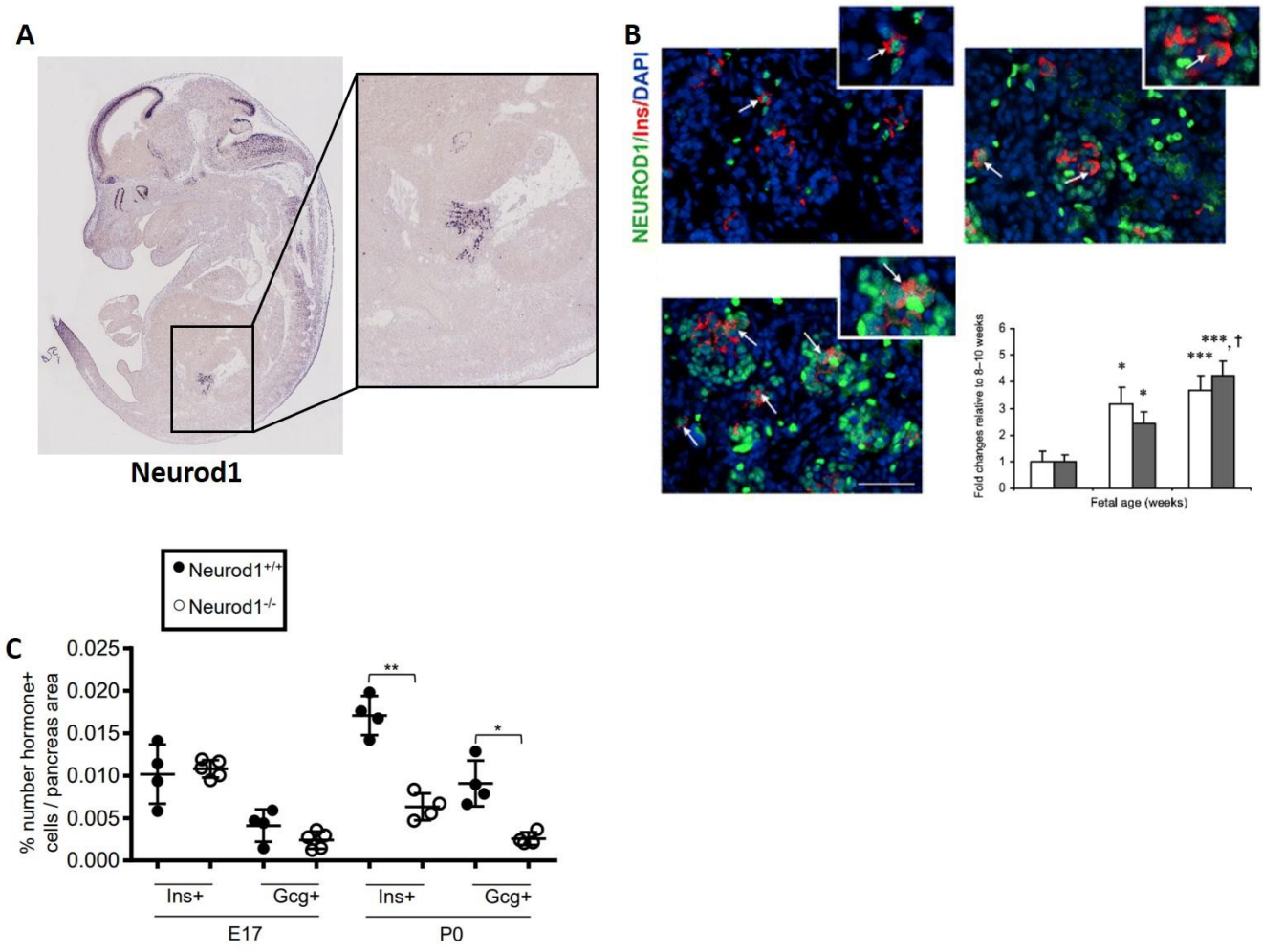


Figure 1.12. Neurod1 during endocrine cell development and differentiation. (**A**) In situ hybridization map of *Insm1* expression at E14.5 in wild type C57BL/6J animals. Data were obtained from GenePaint digital atlas (https://gp3.mpg.de/viewer/setInfo/DA125/16). (**B**) Insulin and glucagon staining at P0 of Neurod1 heterozygous and knockout mice. (**C**) Quantification of hormone expressing cells of *Neurod1*$^{+/+}$ and *Neurod1*$^{-/-}$ samples at E17 and P0. Normalized to area of pancreas with n=4-5 per genotype. Scale Bar = 200um. Insulin, glucagon, CPA1, and DAPI staining of Neurod1 heterozygous and knockout mice. This figure modified and reprinted with permission from Lyttle, et al., 2008 and Romer, et. al., 2019.

Pax6 regulation of endocrine cell fate.

Another downstream target of Neurog3 involved in pancreas development is *Pax6* (*Paired box 6*). The *Pax6* gene (alternatively known as *Small eye* / *Sey*) encodes a protein that contains a paired box domain at its N-terminal and a homeobox domain at its C-terminal, both of which are able to bind DNA (Hanson, 2003; Karthikeyan et al., 2017). Homeobox domain-containing genes are historically known to be major players in developmental genetic programs, regulating developmental processes such as cellular proliferation, differentiation, and migration (Blake and Ziman, 2014; Duverger and Morasso, 2009; Karthikeyan et al., 2017). This family of genes can be subdivided into the HOX and PAX genes. The HOX subfamily are found clustered together in the genome and are prominently involved in the segmentation and patterning of organisms (Karthikeyan et al., 2017; Nusslein-Volhard et al., 1980; Nusslein-Volhard and Wieschaus, 1980). Conversely, the PAX subfamily of homeobox genes, which includes *Pax1-Pax9*, are scattered individually around the genome (Stapleton et al., 1993; Walther et al., 1991). The *Pax* genes have a similar protein sequence and structure and are highly conserved across vertebrate species (Blake and Ziman, 2014; Walther et al., 1991).

A Pax6 mutation in mice was originally identified by chance when researchers identified pups having eyes that were “readily recognized” as being smaller in size compared to their litter mates (Roberts, 1967). Based on the observed phenotype, the mutation, and later its gene, was termed *Small eyes* (*Sey*) until the gene becoming more commonly called by its current name, *Pax6*. This change in terminology was based on its protein structure and classification. For the purposes of this document and to mitigate any confusion, *Pax6* will be used when referring to this gene in the remainder of the text. *Pax6* functions as a TF and has similar expression patterns to that of *Insm1* and *Neurod1*. It is specifically expressed in the developing nervous system and pancreatic endocrine cells, as well as the retina (Hogan et al., 1988; Hogan et al., 1986).

Initial characterization of mice with a non-functional *Pax6* mutation revealed that they were homozygous lethal, and that the specific mutation observed showed incomplete penetrance, with some animals having a more severe eye phenotype than others (Roberts, 1967). Given this information and through decades of research that followed, it was clear that *Pax6* plays a critical role in eye development. In fact, subsequent studies demonstrated that in addition to having small eyes, by E10.5, homozygous mutant embryos fail to undergo lens induction. The optic epithelium and ectoderm are unable to make contact due to separation by mesodermal cells, disrupting

necessary inter-cellular signaling, and confirming perturbations in early development (Hogan et al., 1988; Hogan et al., 1986). The human homolog was later discovered by researchers using positional cloning to identify the gene(s) located in a specific genomic region, known as the Aniridia (AN) locus. Because of the similarities in phenotypes, linkages between the human *Aniridia* locus and the mouse *Sey* had been made (Glaser et al., 1990; van der Meer-de Jong et al., 1990). Deletions and translocations within this locus in humans is associated with the congenital disorder, Aniridia (AN), which presents as the partial or complete absence of the iris, as well as the deterioration of other regions of the eye like the optic nerve and fovea (Nelson et al., 1984; Ton et al., 1991).

Pax6 also plays a critical role in pancreas development and function. It is expressed as early as E9.0 in the pancreatic buds but becomes quickly restricted to the endocrine lineage and maintains expression in all islet cell types of adult animals (**Fig. 1.13A, F-K**) (Ashery-Padan et al., 2004; Sander et al., 1997; St-Onge et al., 1997). Hormone expressing endocrine cells arise from *Pax6*$^{+}$ cells, and it has also been shown to activate expression of pre-proglucagon, insulin, and somatostatin (Jensen et al., 2000a; Sander et al., 1997). Furthermore, previous studies have demonstrated that mice lacking *Pax6* display a diabetic phenotype marked by hyperglycemia and hypoinsulinemia (Ashery-Padan et al., 2004; Sander et al., 1997; St-Onge et al., 1997). These physiological phenotypes are accompanied by disruptions in islet architecture and a reduction in all endocrine cell types, thought to be caused by a failure to expand the endocrine cell population during the secondary transition (**Fig. 1.13B-E**) (Sander et al., 1997; St-Onge et al., 1997). *Pax6* has a particularly critical role in α-cell differentiation, as KO mice completely lack *Gcg*-expressing cells (Sander et al., 1997; St-Onge et al., 1997). In addition to its developmental functions, its role in establishing and maintaining both α- and β-cell identity is further exemplified by the conditional KO of *Pax6* in adult α- and β-cells. These experiments result in a conversion of both cell types towards a *ghrelin*-positive cell type, suggesting that *Pax6* also normally functions to prevent the expression of ghrelin in pancreatic endocrine cells (Ahmad et al., 2015; Heller et al., 2004).

Research has also shown that alterations in expression levels of *Pax6* lead to disruptions in the regulation of glucose, in addition to developmental defects (Kuroda et al., 2004; Rabiee et al., 2020; Tian et al., 2021; Wen et al., 2009; Yasuda et al., 2002). Similarly, perturbations in *Pax6* expression levels, protein structure, and even mutations in target binding sites have been linked to diabetes (Kulzer et al., 2014; Panneerselvam et al., 2019; Tian

et al., 2021). Thus, having a complete understanding of the role of *Pax6* is critical for expanding our knowledge of its molecular mechanisms by which it operates normally and how it may contribute to the diabetic disease state.

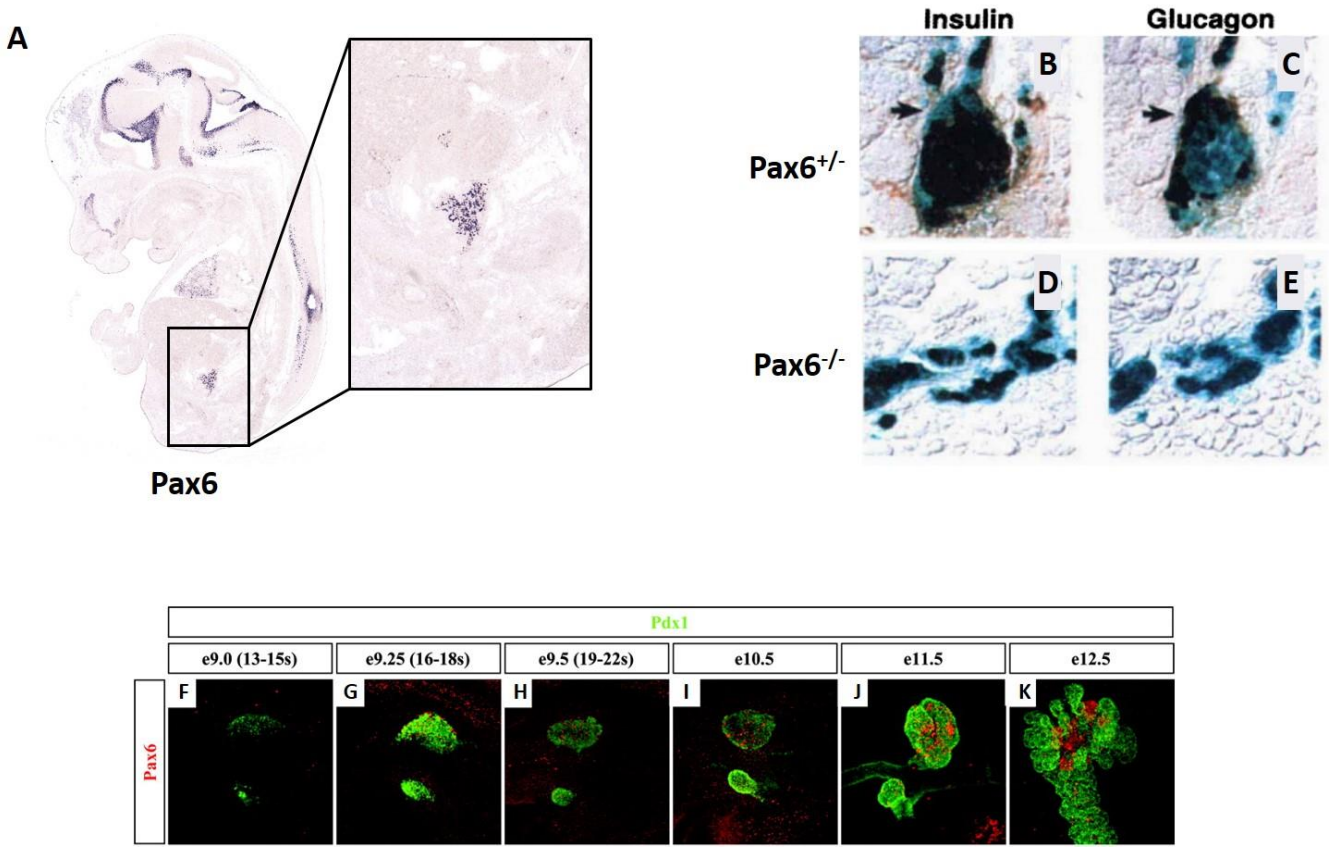


Figure 1.13. Pax6 regulates pancreatic endocrine cell differentiation. (**A**) In situ hybridization map of Insm1 expression at E14.5 in wild type C57BL/6J animals. Data were obtained from GenePaint digital atlas (https://gp3.mpg.de/viewer/setInfo/MH454/10). (**B-E**) Analysis of newborn pancreas of Pax6 heterozygous and Pax6 deficient mice. Tissue sections are stained for B-galactosidase, insulin, glucagon, PP, and amylase expression. B-galactosidase staining is restricted to islet of Langerhans in heterozygous mice and blue stained cells correspond to the hormone indicated. (**F-K**) Collapsed image stacks of wild-type whole embryos (E9.0–E10.5) or dissected pancreata (E11.5 and E12.5). To visualize the pancreatic epithelium, all embryos were stained for Pdx1 and only the dorsal pancreata are shown at E12.5. In all panels, anterior is to the right, posterior is to the left, dorsal is above, and ventral is below. The expression of Pax6 is slightly delayed compared with Neurog3 and Nkx2-2 and can first be detected at E9.25 in the dorsal bud (**G**) and at E10.5 in the ventral bud (**I**) and has not been observed outside the pancreas domains. (**J-K**) At E11.5 and E12.5, many Pax6-expressing cells appear in clusters as observed for the hormones. This image was modified and reprinted with permission from St-Onge, et. al., 1997 and Jorgensen, et. al., 2007.

The role of gene regulatory networks during development

Any given cell is the product of a series of complex developmental processes. Successful development requires adequate proliferation to generate enough cells, tissue specification, cellular differentiation, and organization, and ultimately fully functional mature cells (**Fig. 1.14**). Central to these processes are concerted changes in gene expression patterns. Thus, it is essential that the necessary genes are expressed at the right time, in the right place, and at the right levels for proper development to occur. Precise control of gene expression is dependent on sophisticated circuits of regulatory interactions within the genetic programs (Macneil and Walhout, 2011). It should be appreciated that the gene expression patterns which are observed during development demarcate specific populations of cells and are regulated by multiple tiers of TFs and other regulators (**Fig. 1.15**).

Gene expression is regulated by modifications like methylation and ubiquitylation (Cramer, 2019; Kouzarides, 2007). Once transcribed, the pre-mRNA is regulated by alternative splicing to provide multiple mRNA products, mRNA modifications, stability, and localization (Di Liegro et al., 2014; Moore and von Lindern, 2018; Zhao et al., 2017). The resulting protein products and their function can also be affected by translation events, stability, and localization, as well as through post-translational modifications like acetylation and phosphorylation (Gebauer and Hentze, 2004; Spoel, 2018). Having multiple layers of regulation of gene expression and gene products within the network creates an exponential number of possible interactions and makes studying them difficult. It becomes more complicated still when considering the role of non-coding RNAs in each of these pathways.

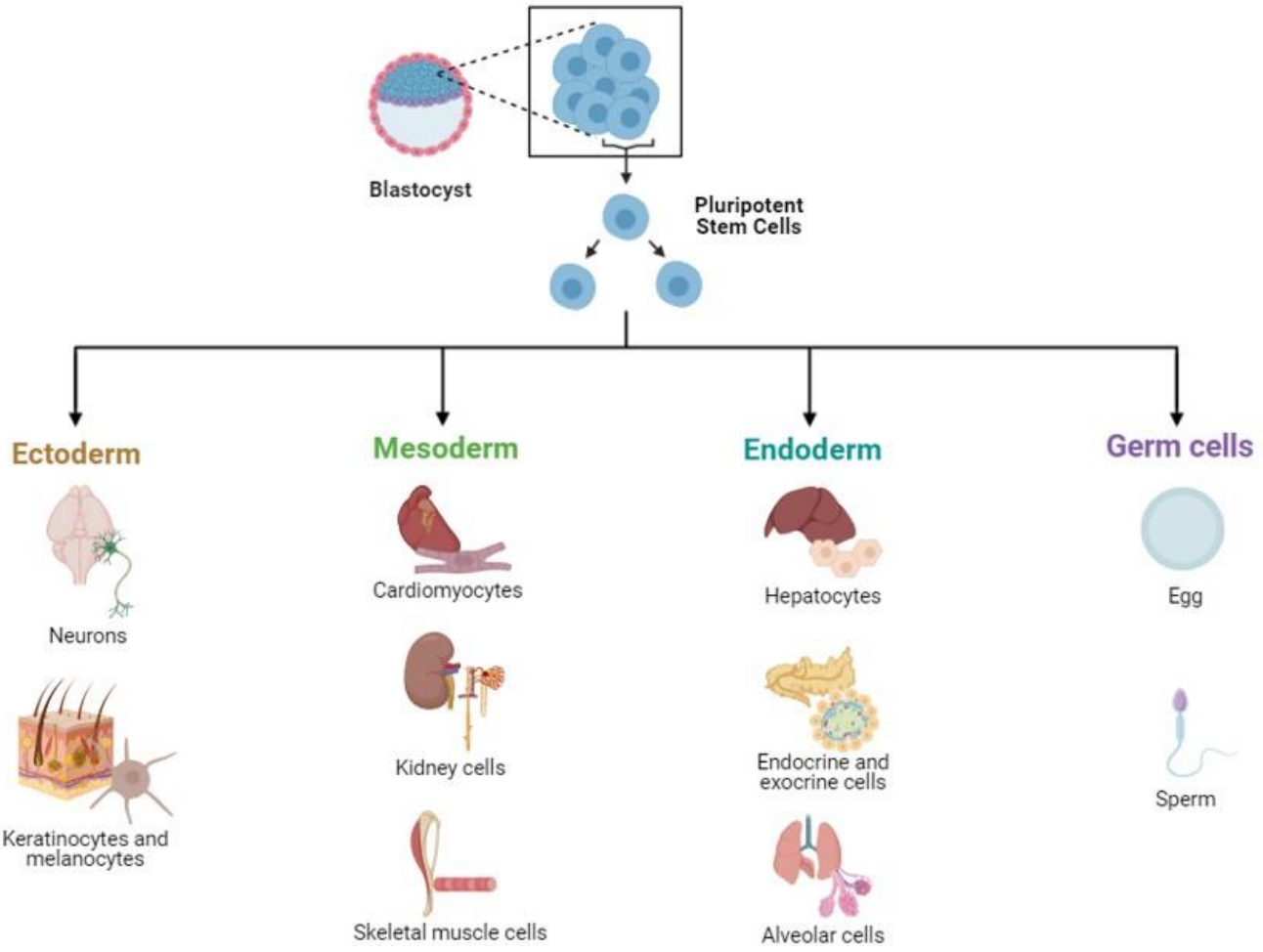


Figure 1.14. Developmental lineage specification of stem cells is driven by an underlying gene regulatory network. Model depicting the differentiation of embryonic pluripotent stem cells into various terminal specialized cell types. The transition from stem cell to somatic cell relies on the integration of signaling cues and underlying gene regulatory networks (GRN). These GRNs are responsible for regulating stable gene expression programs specific to each cell type and provides the instructions necessary for determining cell fate. Original work generated using BioRender.com.

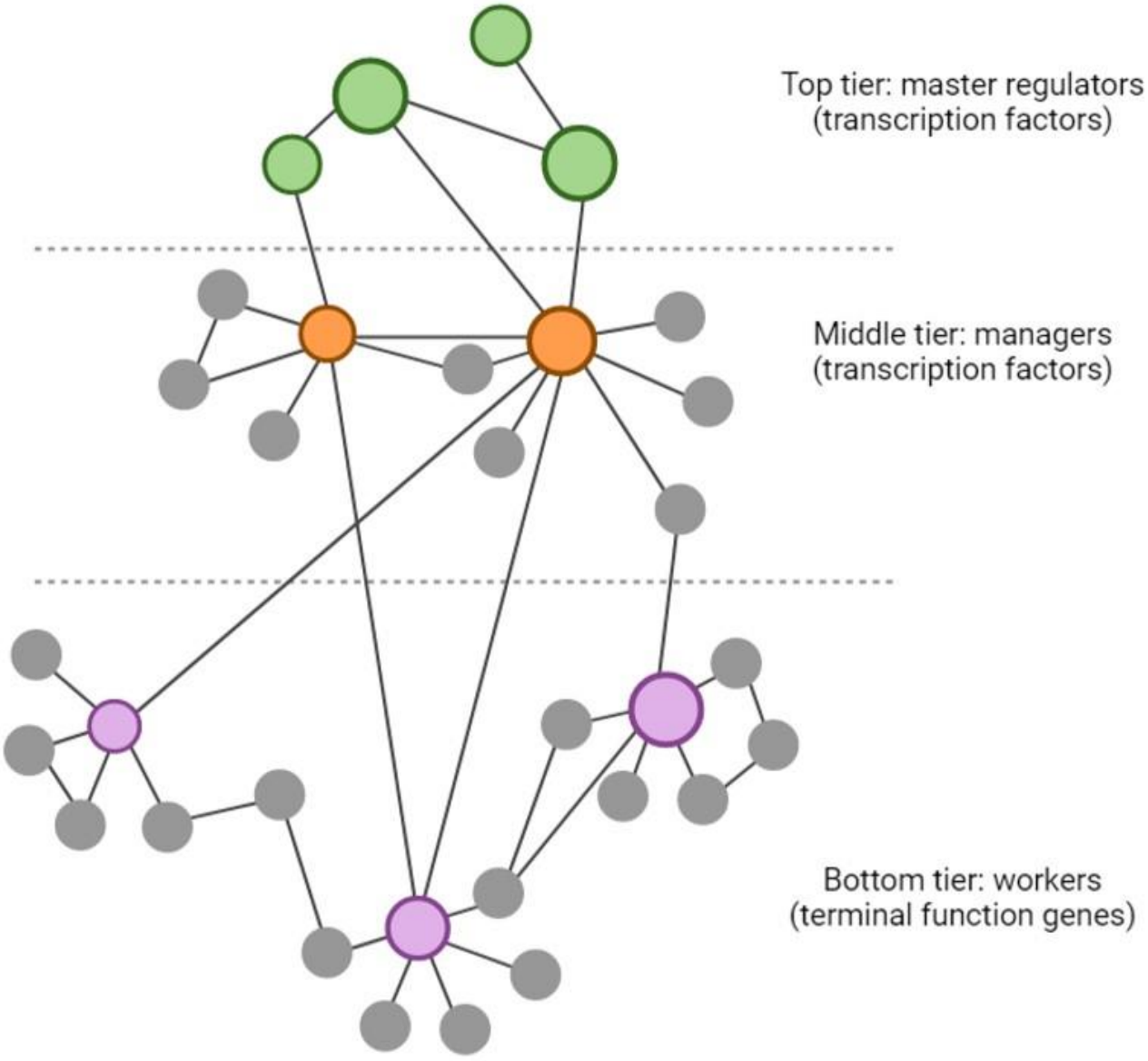


Figure 1.15. Model of hierarchical regulatory interactions within a developmental gene regulatory network. Nodes within the network represent transcription factors (TFs) (larger, colored nodes) and their target genes (smaller, grey nodes). Nodes are connected by edges, depicted as lines, which represent a regulatory interaction between a given pair of nodes. The layout of this gene regulatory network (GRN) schematic is used to emphasize the hierarchical nature of the regulatory interactions within a network. The top tier of TFs comprises those considered to be master regulators, or genes required for lineage specification and initiate a transcriptional cascade necessary for proper differentiation of a given cell type. Factors that fall in the middle tier, are typically “managers” and while they are not required for the initiation, they act broadly and are critical for cells to become fully differentiated. Lastly, genes residing in the bottom tier are “workers” and are generally involved in more specific functions during cellular differentiation. This figure was generated using BioRender.com. The concepts around hierarchical regulation within GRNs has been described and previously reviewed (Jothi, 2009; Nowick and Stubbs, 2010; MacNeil and Walhout, 2011; Rebeiz, 2015).

The biological inputs and outputs of the regulatory genes in the genetic programs can be mapped as gene regulatory networks (GRNs) to provide information about the interactions between the genes expressed in a given cellular context (Levine and Davidson, 2005). GRNs are composed of regulators, such as TFs and signaling factors, and the target genes that they control. The complex interactions that occur within the network are responsible for organ specification, cellular differentiation, and cellular function of adult cells (Sheaffer and Kaestner, 2012). During development, these networks are fluid, and their regulatory states transient due to the rapid changes in gene expression associated with the morphological and functional outputs of cells. Visualizations of GRNs typically represent regulators and targets as "nodes," and the connections between these nodes are depicted as "edges," or lines, indicating a regulatory interaction (**Fig. 1.15**). Edges are directional in nature based on the repressive or activating relationship between the regulator and its target. Due to the expansive number of possible molecular interactions that can impact a GRN, as previously mentioned, many researchers have focused on transcriptional regulatory networks primarily involving only TFs and signaling factors (Levine and Davidson, 2005).

Since the foundations of GRNs are built on the interactions between TFs and their targets, building a network that is accurate and which provides robust predictive power requires vast amounts of data. Data used in creating networks need to provide information about how various factors interact and about the nature of the network inputs and outputs. Current technologies have allowed researchers to gather huge amounts of data in a single experiment but often require a high level of computational power and expertise. RNA-Seq data allows us to infer the interactions between one specific gene (the mutated gene in the model) and literally thousands of other gene products. Likewise, ChIP-Seq data allows us to determine specific protein-DNA interactions by providing sequencing data for DNA regions that are directly bound by a TF of interest. Other experiments like DNA electrophoretic mobility shift assay (EMSA) and DNA pull-down assays can also detect direct protein-DNA interactions. Combined with RNA-Seq, identified protein-DNA interactions correlated by altered gene expression provide evidence for a direct regulatory mechanism, adding confidence to that connection in the network. Other experimental techniques, such as FAIRE-Seq, ATAC-Seq, and DNaseI hypersensitivity assays, can also provide informative data about the broader genomic landscape and accessibility. It should be noted, however, that other biological molecules and events, such as alternatively spliced isoforms, can also contribute to a GRN (**Fig. 1.16**).

These events are becoming easier to study and to incorporate but add a layer of complexity that can make interpretation more difficult.

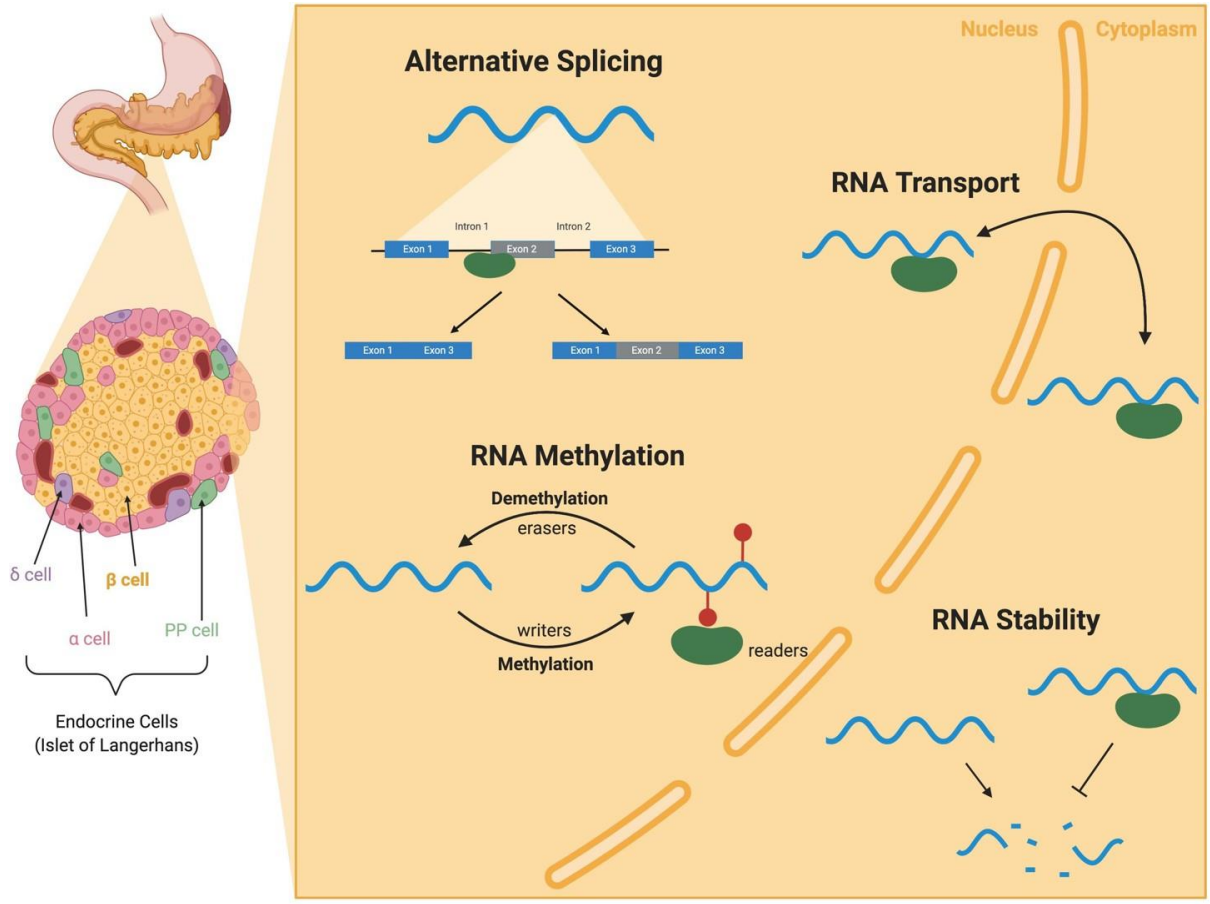


Figure 1.16. RNA-binding protein mediated RNA regulation in the pancreatic β cell. Insulin-secreting β cells reside in the islets of the pancreas along with several other endocrine cell types (α, δ, and PP cells). RNA binding proteins (RBPs) (green) are present in both the nucleus and the cytoplasm of cells and bind to RNA (blue) to perform a variety of functions. RBPs binding to introns and exons of pre-mRNAs contribute to alternative splicing. RBPs can also write, read, and erase methylation modification on mRNAs in RNA methylation. RBPs can also facilitate the transport of RNAs between the nucleus and cytoplasm and throughout the cell. RBP binding to the UTRs can alter mRNA stability and translation. Figure was reprinted with permission from Moss and Sussel, 2020.

While powerful, these techniques in isolation can only provide so much biologically relevant information. There is still a need for genetic manipulation in the form of knockdowns, knockouts, over-expression, or more specific mutations within a gene that may alter its function in some way without completely abolishing its expression. Ideally, these mutations would occur in a spatiotemporally controlled manner to determine the effects within a specific cell type and to not introduce additional variables based on the added effects from other tissues in the organism. With the introduction of CRISPR, scientists can now generate these types of models in a more controlled manner and in a much more rapid time frame. Such genetically modified models followed by the previously mentioned experimental approaches allow us to ask more specific questions and obtain more detailed answers about gene interactions (Thompson et al., 2015). Indeed, many researchers have been following such designs to determine the functional role of genes and to add to the construction of a regulatory network. Combined with the assays mentioned above that can measure the gene expressions changes due to the designed impairments, measurable network outputs can be obtained. In this way, the network can be linked to specific expression outputs and correlated with functional changes observed (Thompson et al., 2015).

In a given network, certain regulators may have a more robust impact if removed, while others display a more subtle or inconsequential role (**Fig. 1.17**). This is largely due to how many connections a node has with other members in the network. For example, a node that is connected to most of the other nodes in a developmental network will greatly disrupt the other interactions and feedback loops potentially causing significant functional phenotypes. However, a different node in a network of a mature cell that has only one or two connections will likely have a limited effect on the remaining interactions between the other regulators, and slightly, if at all, cause defects in the cellular function.

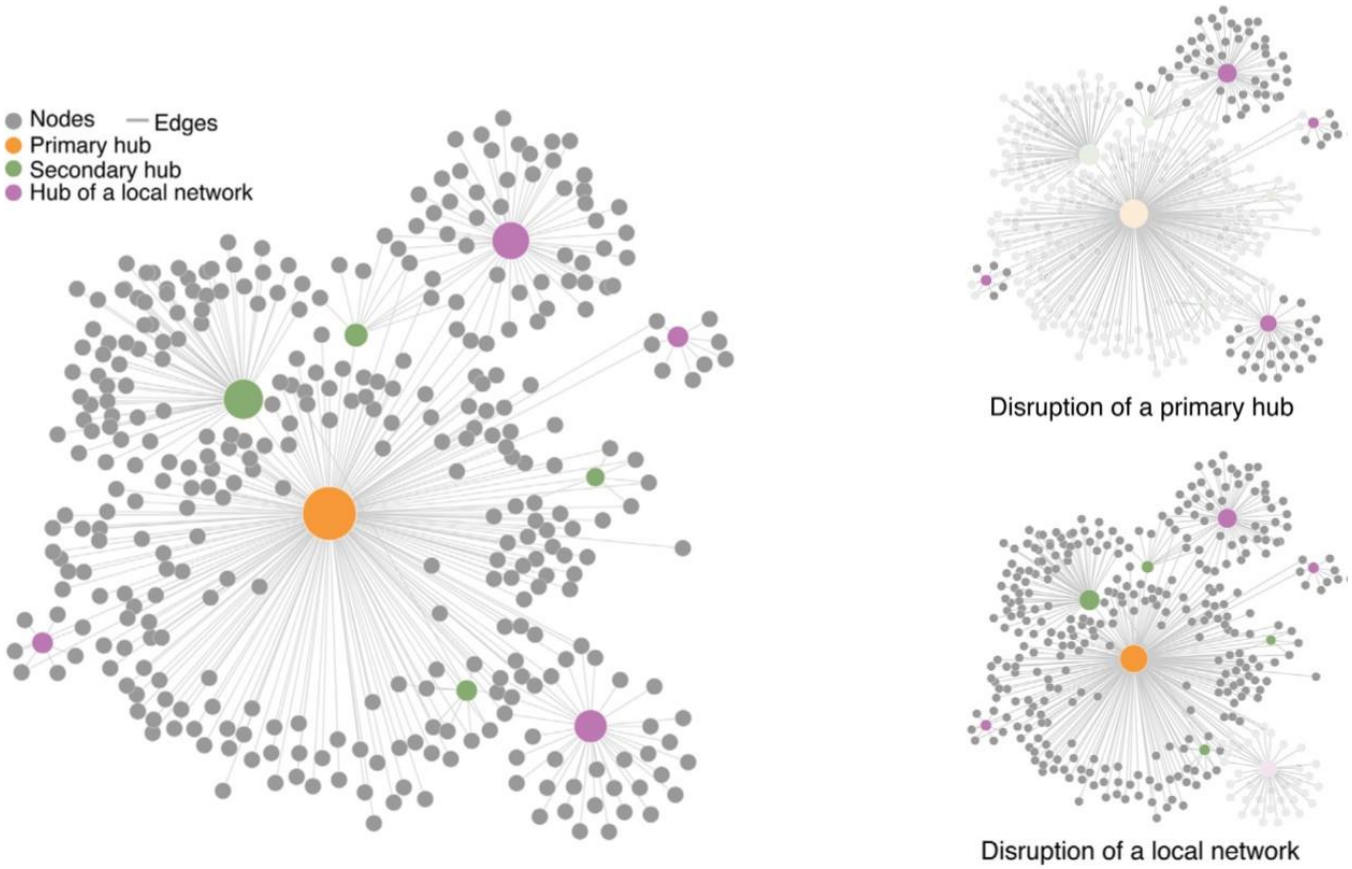


Figure 1.17. Model of hub genes within multi-dimensional gene regulatory networks. Networks consist of nodes and edges, representing cellular components (genes, transcripts, proteins, etc.) and their interactions, respectively. The nodes with the greatest number of interactions constitute primary hubs. Secondary hubs themselves are the interacting partners of primary hubs and simultaneously connect with many other nodes, thereby partially mediating the function of the primary hubs. Local sub-networks are also prevalent. Perturbation of primary or secondary hubs of the network likely poses detrimental impact to the system, while perturbation of a more locally confined network or a node with few interacting partners may not exert substantial effect on the system integrity. Identification of commonly affected networks in diverse retinal diseases will allow the better drug design. Figure was modified and reprinted with permission from Yang, et al. 2015.

Using evidence-based information, networks enable us to apply what is already known about a system to make predictions about the effects of perturbations within that network. For example, consider a simple network in which a given node “A” is shown to be the only activating connection to node “B” and the only inhibiting connection to node “C.” One would predict that if the network were perturbed by the elimination of “A,” that there would be a decrease in the levels of “B” and an increase in the levels of “C.” An obvious way to find the answer to that prediction would be to actually perturb node “A” in a model system and measure the results; however, that can be difficult. Setting up the proper experimental conditions can consume a large amount of time and resources, and it may still not actually have the predicted effect. Alternatively, computational approaches can be used to mathematically model the network under the perturbed state to test the prediction prior to performing any experiments requiring animal models.

Gene regulatory networks in endocrine pancreas development.

Through decades of research focused on lineage tracing, gene expression analysis (microarray, RNA-Seq), gene knockout experiments, and direct TF-DNA interactions to determine the effects of regulator genes, building and studying the regulatory network for the developing pancreas has become possible. Arda et al., reported an account of a curated network based on published data in the field. This report highlighted the numerous TFs and signaling factors known to be important for endocrine cell development and differentiation. The network described included the key regulator Neurog3 and many of its downstream targets such as *Neurod1*, *Pax4*, *Pax6*, and *Myt1* (**Fig. 1.18**) (Arda et al., 2013). While this account of the known interactions is powerful, much work has been performed since then that needs to be accounted for. For example, data characterizing new, or expanding on existing, gene knockout models (e.g., *Insm1, Neurog3, Sox4, Nkx2.2, Myt1)* have since been generated, providing important information about their target genes and the nature of the regulatory relationships (i.e., activating, or repressive) (Churchill et al., 2017; Hu et al., 2020; Osipovich et al., 2014; Sheets et al., 2018; Xu et al., 2015). Incorporating these interactions will improve the accuracy of the GRN architecture and its predictive capabilities.

Another pancreas-specific GRN involving the regulation of T2D associated genes by NFATC4 was also recently published. This study utilized GWAS data, eQTL analysis, and a number of *in vitro* silencing experiments to identify the targets of NFATC4 in adult islets (Sharma et al., 2018). By constructing an expanded regulatory network based on gene expression data from diabetic and non-diabetic cadaver islet samples, the authors were

able to apply a “high control centrality” approach, a parameter that is based on the integration of other centrality measures, such as degree, closeness, eigenvector, betweenness, and PageRank centrality. This approach enabled them to identify pathways that had high centrality measures and were thought to be key drivers of the disease. They then determined if the key pathways identified contained targets of NFATC4 and found that it regulates the expression of many downstream T2D candidate genes. This included the target gene *Tcf7l2*, a gene that has SNPs tightly associated with T2D (Sharma et al., 2018).

Models such as these provide both a visual and a static reference point for generating hypotheses based on known regulatory interactions. However, these models also demonstrate the predictive potential that underlies GRNs in identifying important factors and disease linkages. Such features make them useful for testing hypotheses *in silico* and for aiding in experimental designs.

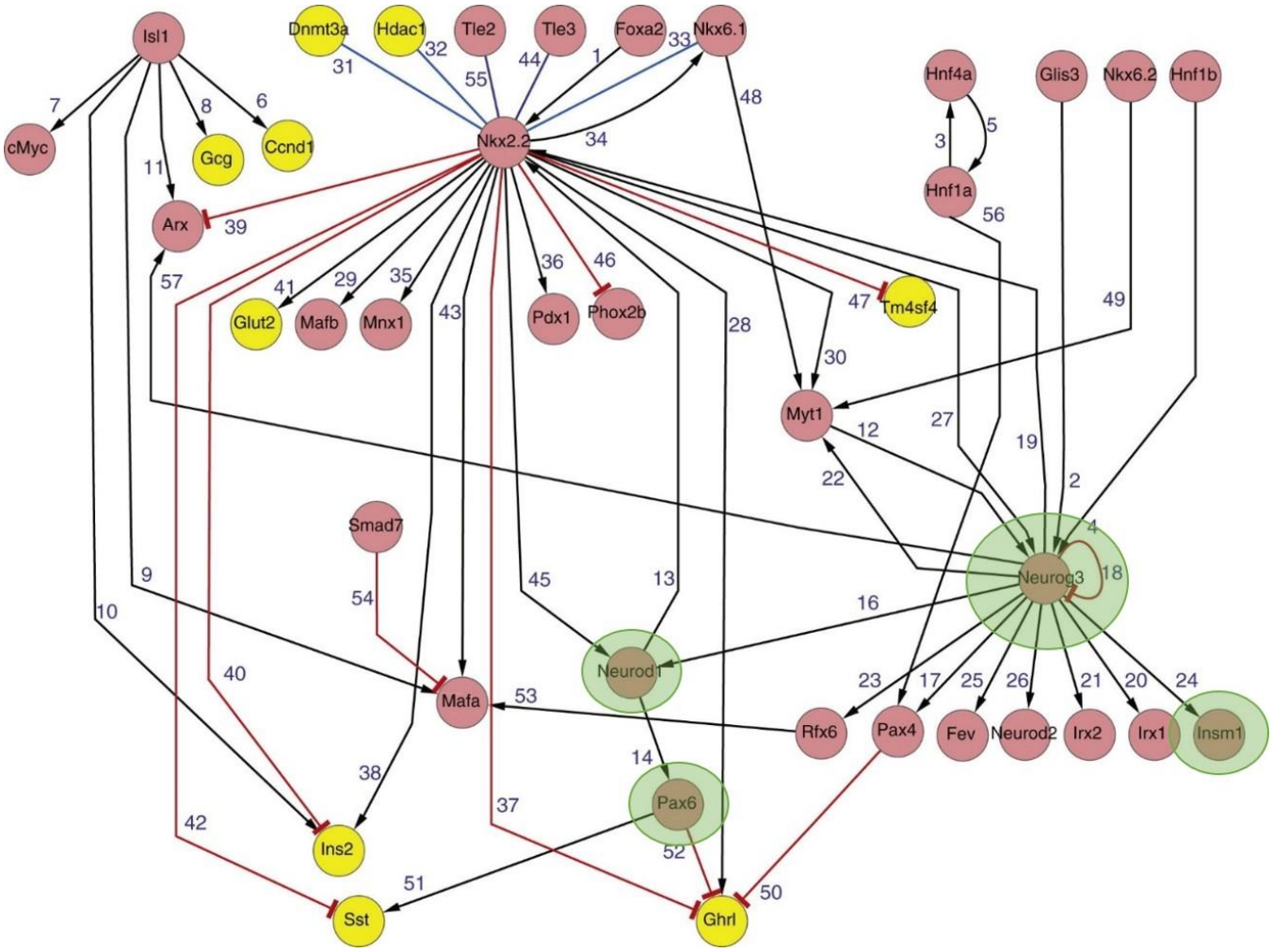


Figure 1.18. Neurog3 is a master regulator of pancreatic endocrine cell development. A schematic representation of an endocrine progenitor gene regulatory network. This network describes regulatory interactions between key transcription factors within the endocrine cell GRN based on experimental evidence. Of note, Neurog3 is a highly connected node in the network, regulating the activation of other critical genes encoding for TFs such as *Insm1*, *Neurod1*, *Pax6*, and *Rfx6*. This figure was modified and reprinted with permission from Arda, et al. 2013.

Diabetes Mellitus – Loss of β-cell function and identity

Diabetes mellitus is a collection of multiple subtype conditions, all ultimately defined by the inability of an individual to maintain blood glucose homeostasis. As of 2019, approximately 48 million people have been diagnosed with diabetes in North America, with over 463 million affected globally (2019)(International Diabetes Federation Diabetes Atlas, 9th edition. 2019.). The primary forms of diabetes include type I diabetes (T1D) and type 2 diabetes (T2D), as well as the less common forms of maturity-onset diabetes of the young (MODY) and neonatal diabetes mellitus (NDM) (Flannick et al., 2016).

T1D accounts for approximately 10% of cases in the US and is most frequently diagnosed in children but can be found in adults. It is an autoimmune disease in which immune cells destroy pancreatic β-cells, and auto-antibodies targeted against insulin, or other β-cell factors, are generated (Sachdeva and Stoffers, 2009). Research suggests that T1D is initiated when a genetically susceptible individual is exposed to an infection, causing the destruction of some β-cells and release of islet-derived pro-inflammatory factors, and leading to an aberrant immune response (Moin and Butler, 2019). The subsequent loss of β-cells leaves the pancreas unable to regulate blood glucose levels leaves individuals dependent on exogenous insulin treatments, as no cure is currently available.

T2D constitutes ~90% of diabetes cases and is marked by chronic hyperglycemia. While this is a heritable form of diabetes, it is complicated by the fact that it is polygenetic and lacks a clear pattern of inheritance (Flannick et al., 2016). Unlike T1D, T2D occurs most frequently in adults, and environmental factors play a larger part in disease susceptibility. Age and obesity are strong risk factors that often lead to insulin resistance, the impaired ability for peripheral tissues to respond to insulin and uptake glucose and place additional stress on pancreatic β-cells (Flannick et al., 2016). With the presence of other stressors such as increased metabolic demands (e.g., high fat and high sugar diets), pancreatic β-cells progressively lose their ability to secrete insulin (**Fig. 1.19**). Combined with a loss of β-cell mass, individuals are unable to meet increased demands and succumb to the T2D disease state (**Fig. 1.20**). Despite T2D associations with other co-morbidities such as cardiovascular disease and stroke, treatment remains limited. A small number of drugs exist that lower blood glucose levels but are unable to prevent further disease progression or prevent the onset of co-morbidities and are not curative (Sachdeva and Stoffers, 2009).

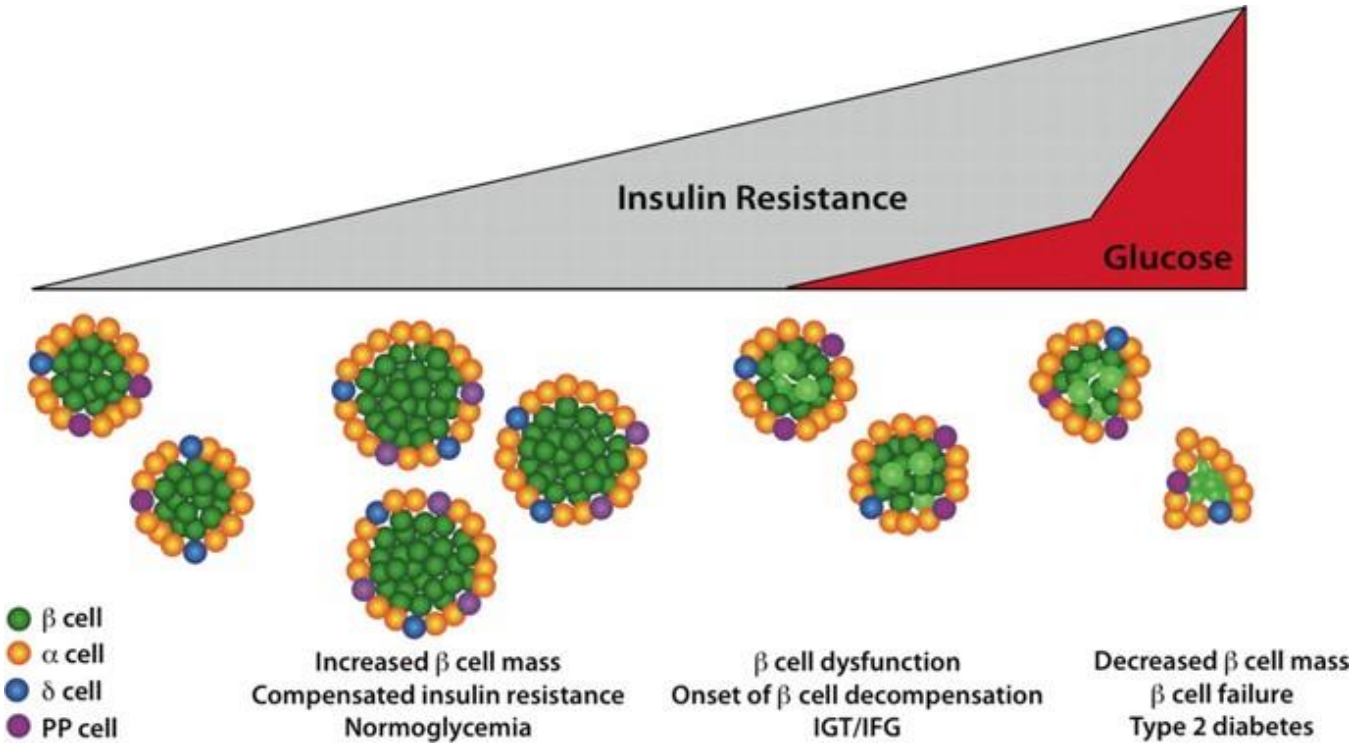


Figure 1.19. Dynamics of β-cell mass during the progression of insulin resistance to diabetes. In the setting of insulin resistance, the pancreatic islet insulin-secreting β-cells respond, in part, via a compensatory increase in β-cell mass, elevating plasma insulin levels to maintain normoglycemia. Changes in both β-cell proliferation and survival play important roles in this adaptive expansion of β-cell mass and in the reduction in mass that is associated with progressive β-cell dysfunction, eventually leading to type 2 diabetes. These cellular processes are regulated by extracellular signals from a number of tissues, as described in detail in the text. IGT, Impaired glucose tolerance; IFG, impaired fasting glucose; PP, pancreatic polypeptide. Figure was reprinted with permission from Sachdeva and Stoffers, 2009.

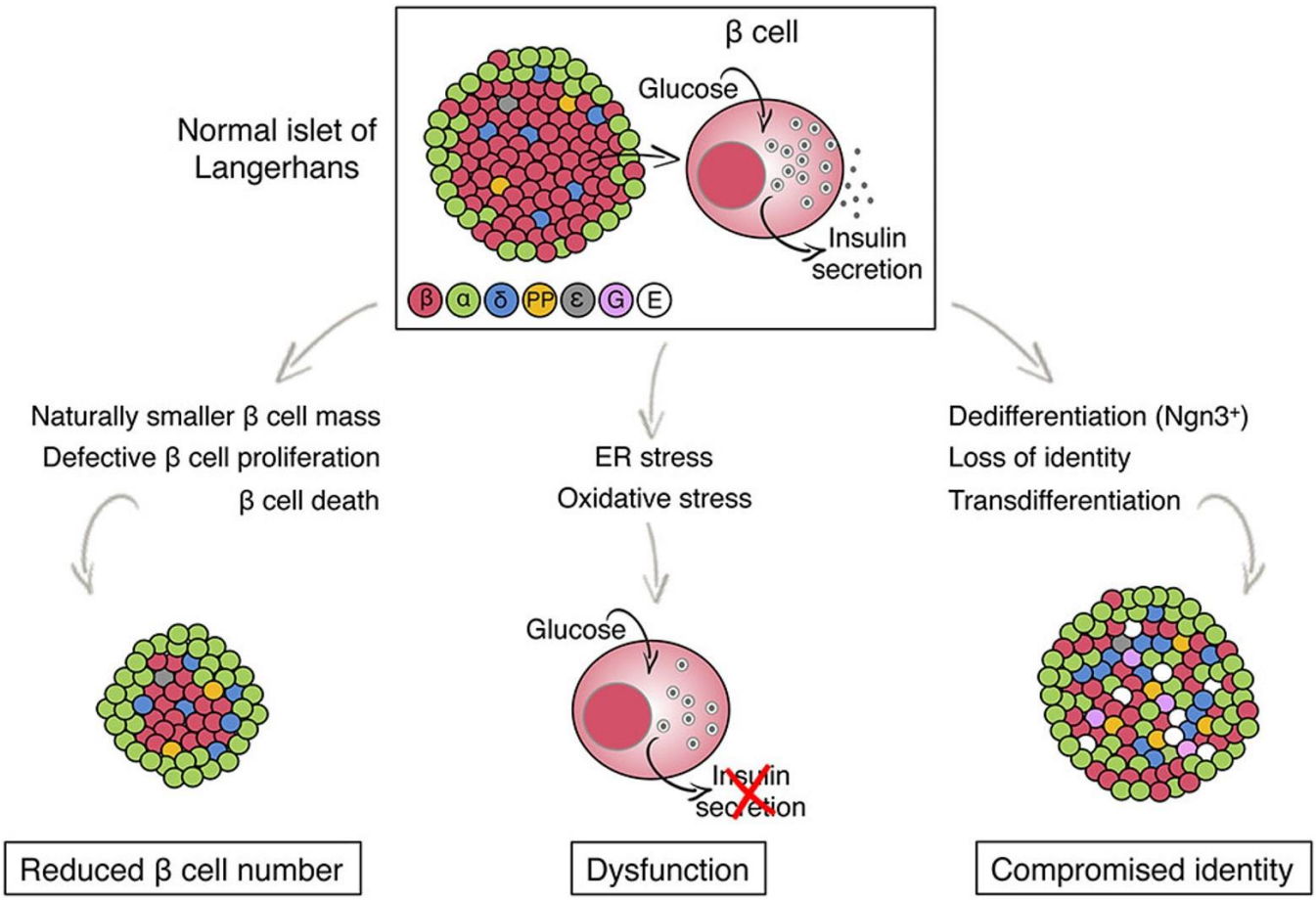


Figure 1.20. Models for β-cell failure in Type 2 Diabetes. The illustration shows the architecture and endocrine cell composition of a normal pancreatic islet of Langerhans (top) and potential changes, distinguished by β-cell fate, that lead to β-cell failure in T2D (bottom). Different colors indicate different islet cell type. ε, ghrelin; G, gastrin, E, endocrine cell with empty granules (no hormone is produced). Figure was reprinted with permission from Wysham and Shubrook, 2017.

Monogenic forms of diabetes differ from the prominent T1D and T2D classes in that they are driven by causal mutations within a single gene. These subclasses include NDM, MODY, and mitochondrial diabetes. The age of onset and degree of severity can vary greatly across these forms of diabetes. For example, while both are characterized by β-cell failure, NDM typically presents at <6 months of age while the onset of MODY generally occurs anywhere from 6 months to 35 years of age, depending on the driving mutation (**Fig. 1.21**) (Flannick et al., 2016). Genes containing driver mutations of these subtypes include *Gata6*, *Neurog3*, *Pax6*, *Gck*, *Neurod1*, *Pdx1*, *Hnf1a*, and *Abcc8,* among others. Many of these same driver genes are found within the same regulatory networks found to be impaired in, and have GWAS associations with, T2D or glucose dysregulation (Flannick et al., 2016).

It has previously been accepted that the loss of β-cell mass was due to increases in β-cell death. Recent studies, however, have suggested that the loss of β-cell identity could be a bigger factor (Cinti et al., 2016; Talchai et al., 2012; Wang et al., 2014). Consistent with this theory is the fact that β-cells that have not fully matured or that have an immature molecular identity are not able to properly regulate insulin secretion in response to glucose. It has also been suggested that the dedifferentiation of β-cells is an adaptive response that may be necessary to increase proliferation (Dor and Glaser, 2013; Weinberg et al., 2007). However, the ability of β-cells to proliferate varies greatly across populations, and conflicting bodies of evidence in this area make this a controversial topic as there have been mixed findings on the efficiency this approach can increase overall β-cell mass (Bader et al., 2016). β-cell dedifferentiation appears to be a stepwise process in which a coordinated loss of regulatory TFs is observed in parallel to the loss of cellular identity and function. Importantly, the sequence of TFs lost is in the reverse order of that seen during development, highlighting the importance of understanding the developmental processes and the translational potential to diabetes.

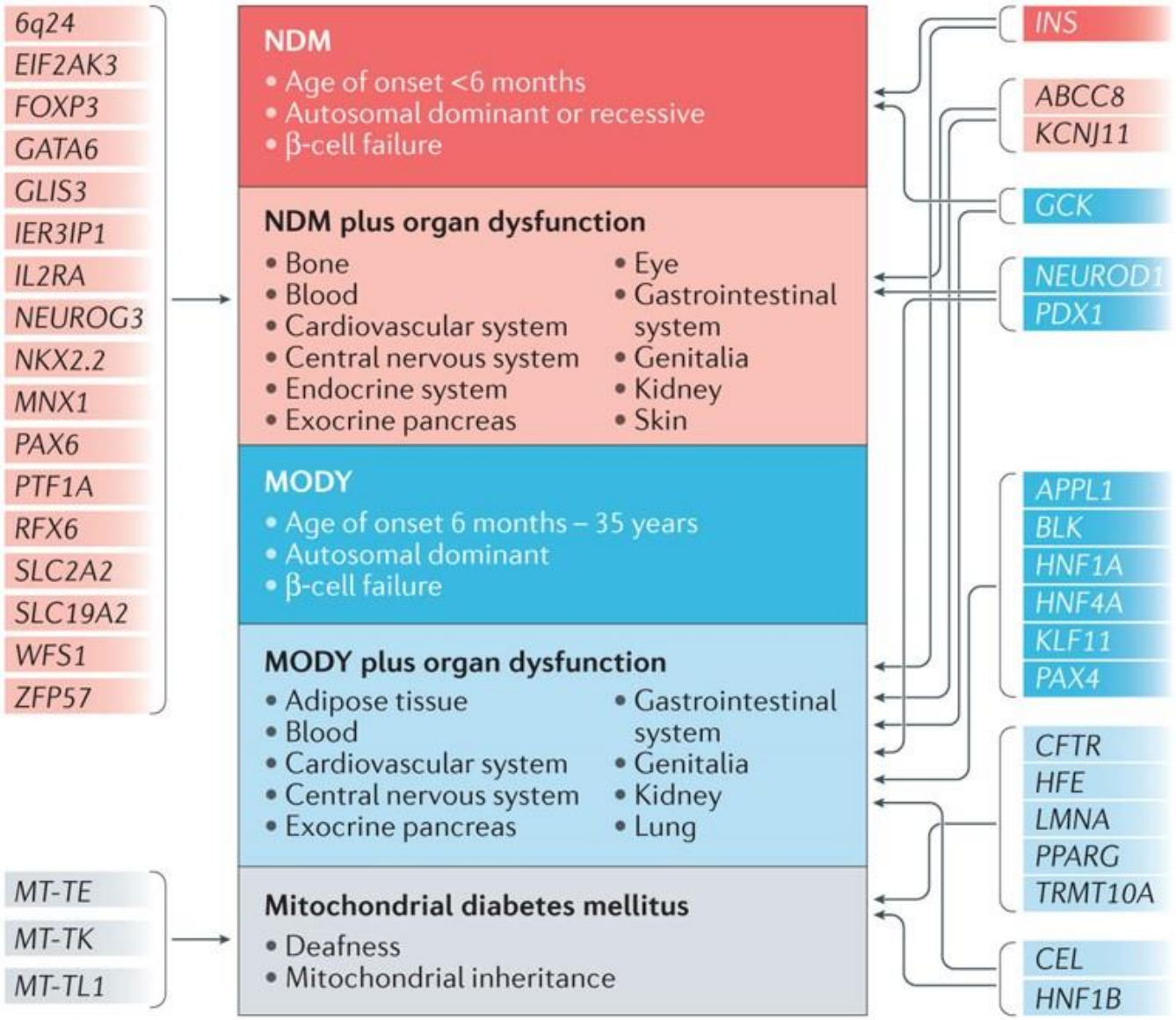


Figure 1.21. Monogenic forms of diabetes mellitus. Shown are genes implicated in a restricted set of monogenic diabetes mellitus subtypes (discussed in the article). Each class of disorder is shown as a box in the central column: neonatal diabetes mellitus (NDM, red); maturity-onset diabetes of the young (MODY, blue); and mitochondrial diabetes mellitus (grey), while syndromic forms are divided based on closer similarity to NDM (pink) or MODY (light blue). The major clinical features of each disorder are described in each box. NDM and MODY with organ dysfunction can have one or more organs affected. Genes are shown as boxes in the outer two columns, with arrows connecting to the disorders for which the genes are implicated. Genes implicated in more than one disorder have arrows to each. CNS, central nervous system. This figure was modified and reprinted with permission from Flannick, et al., 2016.

The precise mechanisms driving dedifferentiation of β-cells have not yet been elucidated, but proposed models include the infiltration of the islet micro-environment by pro-inflammatory factors (inflammation-induced), chronic glucotoxicity, and the up-tick in endoplasmic reticulum (ER) stress and associated impairment of the unfolded protein response (UPR) (Moin and Butler, 2019). No matter which mechanism, or a combination thereof, the loss of β-cell mass remains a critical feature contributing to diabetes phenotypes and an area of interest for researchers. Relevant studies have pointed towards several mechanisms in which cellular mass can be replenished. For instance, functional β-cell mass can be increased through cellular proliferation. Although basal β-cell proliferation in adults is low (~<0.5% in humans), this rate can increase to a degree under stressed conditions (Mosser et al., 2015). The neogenesis of β-cell mass from progenitor cells is another option for making new cells, but this approach remains controversial due to the level of efficiency (Bonner-Weir et al., 2008; Kopp et al., 2011). Collectively, it seems that these mechanisms are unlikely to overcome the inherent apoptosis and dedifferentiation processes which reduce functional β-cell mass in the diabetic disease state (**Fig. 1.9**). This makes efforts based on regenerative approaches enticing and of great interest.

Significance of understanding endocrine cell development for therapeutic gains

Diabetes, defined by the body's inability to maintain glucose homeostasis, remains a global problem despite decades of dedicated research efforts. The trend in the number of cases worldwide has also only been growing for the last several years, with future projections rising at steady rates (2019)(IDF diabetes atlas). To date, insulin treatments remain the most common form of therapy. In the last decade, researchers have made great strides in establishing regenerative medicine approaches to create a more enduring treatment for diabetes (Pagliuca et al., 2014; Rezania et al., 2014). Many approaches have been explored, including guided differentiation of induced pluripotent stem cells (iPSC) or embryonic stem cells (ESCs) and direct reprogramming of one cell type to another, but bypassing the transient pluripotent stage. These approaches have been performed in many different variations, both *in vitro* and *in vivo*, with great progress being made in both areas. However, the knowledge enabling our advances in this endeavor has been driven by basic research focused on the innate factors involved in determining endocrine cell lineage allocation and the cellular function of mature cells. This means that through continuing to

study the fundamentals of endocrine cell development, the field will be able to apply those concepts to optimizing reprogramming protocols.

Furthermore, while efforts have been made to increase our understanding of pancreatic endocrine cell development and to answer many lingering questions, gaps remain. Although I have discussed many of these in the preceding text, I have outlined a select few below. It should be noted that this is a truncated list that has been compile through many personal discussions as well as questions currently posed within the community's literature and is by no means comprehensive. More generally, how are the development programs (e.g., proliferation, differentiation, morphogenesis) coordinated by the underlying transcriptional program? How are pancreatic progenitor cells maintained and regulated in such a way that they can give rise to the right number of endocrine cells at the right time in order to generate a properly functioning organ? What molecular underpinnings are responsible for the cell fate choices of pro-endocrine cells, and what are the cues necessary to trigger those interactions? How do cells choose their final hormone cell type? Can we use the information we have about native *in vivo* developmental programs and cellular decision-making to control cellular choices (and ultimately outcomes) to make progress in therapeutic approaches?

With respect to *Insm1*, *Neurod1*, and *Pax6*, how are these factors connected within the GRN? When, and to what degree, do these factors contribute to the coordination of the developmental programs critical for proper endocrine cell differentiation and maturation? Do these factors regulate a small subset of specific genes and pathways in endocrine cell development, or do they act more broadly? Do they share regulatory targets and have synergistic behaviors, or do they act on separate pathways independent of each other? Do these TFs govern the cycling properties of pro-endocrine progenitors to regulate the number of endocrine cells established? Are there domain-specific functions of each TF that are important for development?

Other emerging areas of interest also inspire unanswered questions. Where does alternative splicing fit into the intricate network that drives developmental processes? Are some RBPs more enriched or important than others in regulating pancreatic endocrine cell development? Do known disease-associated SNPs interrupt mRNA splicing events, causing functional downstream consequences? Does cellular stress lead to alterations in the splicing program and contribute to cellular dysfunction in the disease state? Can we identify alternative splicing events or mechanisms that can be targeted for therapeutic applications? In the case of zinc finger proteins, how many, and

which, ZFPs are involved in endocrine cell development and function? Do these factors function as transcription factors, or do they have other functionalities that contribute to the developmental program? What other TFs do they interact with or, perhaps, regulate?

As research efforts move forward, it is important to reflect on these lingering questions. Doing so will help guide hypotheses and open opportunities to dissect the molecular complexities of pancreatic endocrine cell development and function more precisely. Having high-level questions in mind has helped me to find creative ways to better understand how members of the GRN, specifically *Insm1*, *Neurod1*, and *Pax6*, work to establish fully functional endocrine cells.

Overview of thesis

Despite countless fruitful studies that have shed light on the functional role of key transcription factors in pancreas development, our understanding of their individual roles and their roles within a larger developmental gene regulatory network remains incomplete. Many gaps in our understanding of the precise role of key transcription factors have yet to be answered. With the aim of contributing to this collection of knowledge, I have further investigated the role of three important TFs that are Neurog3-dependent. *Insm1*, *Neurod1*, and *Pax6* are all Neurog3-dependent TFs and have been independently studied. However, analyses directly comparing their functions have not been performed even though they share similar phenotypes in defective islet development when ablated in mouse embryos. My studies have been aimed at further elucidating the individual roles of each of these TFs on the regulatory network of developing endocrine cells and to identify processes that are uniquely and commonly regulated by each. This was achieved by performing immunohistochemical, bulk RNA-Seq, and differential splicing analysis of *Insm1*, *Neurod1*, and *Pax6*-null mouse embryos during critical developmental periods of endocrine cell expansion and differentiation. First, I generated knockout mice of *Insm1*, *Neurod1*, and *Pax6* to determine the effects on the pancreas. Second, I quantified the morphological consequences on islet development by the loss of each TF during embryonic stages. Third, to determine the genes whose expression is regulated by each factor, I performed bulk RNA-Seq analysis of FACS purified KO endocrine cells and a reanalysis of Insm1 and Neurod1 DNA binding data in β-cells (Chapter II). Fourth, I performed an alternative splicing analysis using the

RNA-Seq datasets to identify changes in AS patterns caused by the loss of each TF (Chapter III). Lastly, I sought to explore the biological role of a subset of zinc finger proteins that are expressed in murine islets and transcriptionally regulated by one or more TFs (*Neurog3, Insm1, Neurod1,* or *Pax6*) (Chapter IV).

CHAPTER II - REGULATION OF ENDOCRINE PANCREAS DEVELOPMENT BY INSM1, NEUROD1, AND PAX6

INTRODUCTION

During pancreas development, the formation of endocrine cells within nascent islets of Langerhans requires the formation of an endocrine cell-specific gene regulatory network (GRN). Prior studies have elucidated multiple transcription factors (TFs) that contribute to the establishment of a stable GRN (Arda et al., 2013; Osipovich et al., 2021). Neurog3 plays a vital role in initiating the formation of the GRN in pancreatic endocrine cells by activating a downstream cascade of pro-endocrine TFs (Bechard et al., 2016; Ejarque et al., 2013; Gu et al., 2002). Activation of this transcriptional program ignites a series of developmental processes driven by the downstream TFs critical for the stepwise differentiation of endocrine cells (**Fig. 2.1**).

In its absence, the fate of pancreatic pre-endocrine cells is redirected towards the ductal lineage, and no endocrine cells are formed (Gradwohl et al., 2000; Osipovich et al., 2021; Schwitzgebel et al., 2000; Wang et al., 2010). In contrast, when other pro-endocrine TFs downstream of *Neurog3* are eliminated, including *Insm1*, *Neurod1,* and *Pax6*, pancreatic endocrine-like cells are formed that exhibit marked defects in proliferation and/or hormone expression, suggesting that each factor regulates specific subnetworks of genes (Gierl et al., 2006; Heller et al., 2004; Huang et al., 2000; Mellitzer et al., 2006; Naya et al., 1997; Osipovich et al., 2014; Sander et al., 1997; St-Onge et al., 1997).

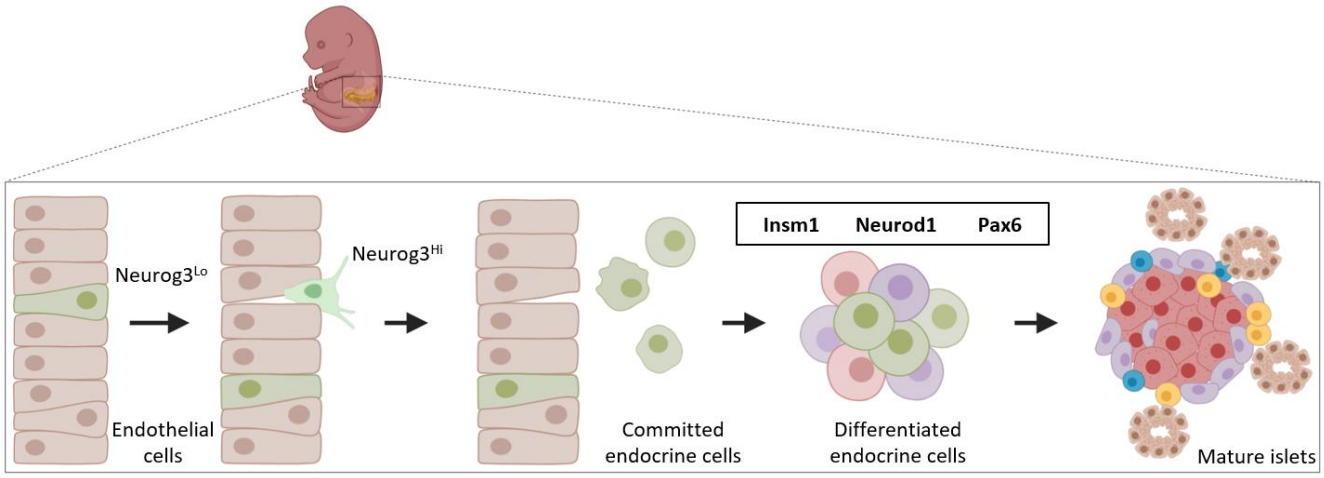


Figure 2.1. Neurog3 marks pro-endocrine committed cells and initiates transcriptional cascade. Simplified schematic of endocrine cell development. Neurog3 drives expression of key transcription factors, which in turn promote cellular differentiation and maturation. This is an original work generated using BioRender.com.

TFs activate target genes by directly binding DNA in promoter regions, and ChIP-Seq represents a useful approach to identify genome-wide TF binding sites. Previous ChIP-Seq studies using cell lines have identified many binding sites for Insm1, Neurod1, and Pax6. Analysis of Insm1 and Neurod1 binding sites in a pancreatic β-cell line derived from adult islets has shown that Insm1, Neurod1, and Foxa2 frequently co-bind to the same genomic regions. However, only 32% of genes in adult β-cells that are regulated by *Insm1* have nearby Insm1 binding (Jia et al., 2015). Similarly, in insulinoma cells, ~40% of Pax6 binding sites overlapped with Neurod1 sites, and both were enriched at islet cell-active enhancers (Lizio et al., 2015). A separate study found that Pax6 had activating and repressive functions in mouse and human β-cell lines and showed a large degree of overlap between Neurod1 and Pax6 binding sites (Swisa et al., 2017).

RNA processing is also an important process that is regulated during development. There are hundreds of different RNA binding proteins (RBPs) that affect the polyadenylation, stabilization, and localization dynamics of mRNA. Some of these proteins also introduce nucleotide modifications and cause differential splicing of introns and exons (Carazo et al., 2019a; Carazo et al., 2019b; Licatalosi and Darnell, 2010; Manning and Cooper, 2017; Wickramasinghe and Venkitaraman, 2016). Alternative RNA splicing significantly enhances transcriptome diversity, and it is important for cell differentiation (Fiszbein and Kornblihtt, 2017). Indeed, Singer *et al.* have recently reported that the alternative splicing of *Pax6*, which requires recruitment of RBPs by the lncRNA *Pauper*, alters both its transcriptional activity and DNA binding specificity (Singer et al., 2019). It has also recently been shown in islets from individuals with type 2 diabetes that many RBPs are dysregulated and that this may impair correct RNA splicing (Jeffery et al., 2019).

In this study, I performed immunohistochemical and bulk transcriptomic analyses to directly compare the effects of eliminating *Insm1*, *Neurod1,* and *Pax6* during endocrine cell development. Our findings indicate that this three TFs individually and coordinately regulate the expression of common and unique sets of genes necessary for the proliferation and function of pancreatic endocrine cells. Additionally, I found that Insm1, Neurod1, and Pax6 differentially affect the splicing of genes, thereby adding complexity to pancreatic endocrine cell proteomes.

MATERIALS AND METHODS

Mice and genotyping. *Insm1*$^{tm1.1Mgn}$ (*Insm1*^{GFPCre}) (MMRRC Stock No: 36986-JAX; MGI: 1859980) (Osipovich et al., 2014) and *Neurod1*^{tm1Jle} (*Neurod1*^{LacZ}) (MGI: 2385826) (Miyata et al., 1999) mice were maintained in a CD-1 background. *Pax6*^{tm2Pgr} (*Pax6*^{flox}) (MGI: 1934348) mice were obtained from Jackson Laboratory and kept in a C57BL6/J background (Ashery-Padan et al., 2000; Sun et al., 2008). All animal experimentation was performed under the Vanderbilt University Institutional Animal Care and Use Committee's oversight. The *Insm1*^{GFPCre} allele was identified in mice as previously described (Osipovich et al., 2014). *Pax6*^{flox} mice were genotyped using the primer pairs CCTAACAGAGCCCCGTATTC (forward) and GCCCAACAGTCCAGAGAAAG (reverse). To detect wild-type *Neurod1* and *Neurod1*^{LacZ,} I used either ACCATGCACTCTGTACGCATT (forward) or GAGAACTGAGACACTCATCTG (forward) in combination with AAACGCCGAGTTAAAGCCATC (reverse), respectively.

Timed matings and tissue collection. Embryos were isolated from timed matings where noon of the day that vaginal plugs were observed was designated E0.5. Embryonic day (E) 15.5 animals were dissected into ice-cold PBS and visually genotyped for the presence of *Insm*$^{GFP.Cre}$ allele based on green fluorescence. For PCR genotyping, embryonic tissues were digested at 55˚C in 100 µl of PCR lysis buffer (1X PCR buffer, 0.1% Triton X-100, 100 µg/ml Proteinase K), inactivated at 90˚C for 15 minutes and used for PCR. The whole pancreas was dissected and either processed for FACS or frozen in OCT for immunostaining analyses.

Immunofluorescence. Immunofluorescence staining of frozen tissue sections was performed as previously described (Burlison et al., 2008). Ten µm tissue sections were stained using anti-GFP (1:500, ThermoFisher, #A10262), insulin (1:1000, Invitrogen, #PA1-26938), glucagon (1:100, Milipore, #AB932), somatostatin (1:1000, Linco), pancreatic polypeptide (1:1000, Linco), and/or Ki67 (1:500, ThermoFisher, #RM-9106-S1) antibodies, and ProLong Gold anti-fade reagent with DAPI (Life Technologies, #P36941) used to mount coverslips. TUNEL (TMR red) assays were performed following the manufacturer's protocol (ROCHE). Respective images were obtained via confocal microscopy using an LSM 510 Meta microscope at 20x magnification. The image processing software, ImageJ (National Institute of Health), was used to manually identify and quantify fluorescently positive cells (Schneider et al., 2012). GFP-positive cells were counted independently first, followed by quantification of hormone-positive islet cells. The ratios of GFP-positive cells, which also express other proteins or hormones, were then

calculated. Three different biological replicates from each genotype and at least 500 cells were analyzed for each measurement.

Fluorescence-Activated Cell Sorting (FACS). Pancreatic cells were dispersed into a single-cell suspension by incubation for 5 minutes at 37°C in Accumax (Sigma) containing 50 µg/mL DNase I. Reactions were quenched by the addition of sorting buffer (FACS staining buffer (R&D Systems), DNAse (1:1000)). Cells were then filtered through a 35 µm nylon mesh into a FACS tube (Corning), washed once with FACS staining buffer, centrifuged at 1100 rpm for 3 minutes, then re-suspended in 500 µl of FACS sorting buffer (Osipovich et al., 2014). 7-aminoactinomycin (7-AAD) was added at a concentration of 1:1000 immediately before sorting using either a Benton Dickenson FACS Aria-II or Aria-III instrument to distinguish between live and dead cells. GFP-positive, 7-AAD negative cells were collected directly into Trizol LS (Invitrogen) containing 40 µg/mL of mussel glycogen (Roche/Sigma). An average of 6,000 cells was obtained per embryo sample.

RNA isolation, library construction, and RNA-Seq. Three replicate samples for each genotype were collected, and total RNA was isolated using Trizol LS then treated with DNase I (Life Technologies). RNA was column-purified using the RNA Clean and Concentrator Kit (Zymo Research) and previously published protocols (Osipovich et al., 2021). RNA integrity was determined using an Agilent 2100 Bioanalyzer (Agilent Technologies, CA), and samples with an RNA integrity number (RIN) of 7.0 or greater were used for RNA-Seq. All samples were amplified using the SMART-Seq Ultra Low Input RNA Kit (TAKARA/Clontech) at 10 cycles, except for the *Neurog3*$^{GFP/+}$ and *Neurog3*$^{GFP/GFP}$ samples, which were prepared using the Ovation RNA-Seq System V2 (NuGEN). cDNA was prepared using the Low Input Library Kit (Clontech) and sequenced using either an Illumina HiSeq3000 genome analyzer to obtain paired-end, 75-bp reads, or Illumina NovaSeq6000 to get paired-end, 100-bp reads. RNA samples were sequenced to an average depth of ~5.0 x 10^7 reads.

*Bioinformatic analysis of RNA-Seq***.** FastQ files were routinely processed then sanitized (e.g., removing adapter contamination and trimming low-quality nucleotides) using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and TrimGalore (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), respectively. Individual reads were aligned to the mouse genome (mm-10) using Spliced Transcripts Alignment to a Reference (STAR) software (Dobin et al., 2013). Pairwise differential gene expression analyses were performed using HTSeq (Anders et al., 2015) to obtain

read counts, and DESeq2 was used (Love et al., 2014) to quantify and display the differences. Only genes that exceeded an adjusted p-value ≤ 0.05 were included in the analysis. Gene ontology (GO) analyses were performed using the online Database for Annotation, Visualization, and Integrated Discovery (DAVID v6.8) (Huang da et al., 2009a, b). *Neurog3*$^{GFP/+}$ and *Neurog3*$^{GFP/GFP}$ samples were separately analyzed as described (Osipovich et al., 2021).

Bioinformatic analysis of ChIP-Seq. ChIP-Seq datasets for INSM1 and NEUROD1 binding sites were obtained from ArrayExpress (E-GEOD-54046). Both datasets were obtained using an insulinoma cell line propagated from RIP1-Tag2 transgenic mice (Jia et al., 2015). Snakemake (5.2.4), a workflow management system, was used to manage the data processing. FastQC (0.11.7) returned a quality score report on each sample's sequences before and after Trimgalore (0.6.5) discarded sequences with a read length shorter than 20 bp were discarded and trimmed read ends that do not meet the threshold of 20. Trimmed sequences were used by the RNA-Seq aligner Bowtie2 (2.3.5) to align against the mouse genome (gencode 17; GRCm38). On average, 95% of reads were mapped to the reference per sample. Samtools (1.9) converted the SAM files into BAM files and filtered out alignments with a lower MAPQ value than 10, and sort alignments by leftmost coordinates and by read name. Macs2 (2.2.6) was used to perform peak calling on the samples using default parameters for bandwidth and effective genome size. The resulting BED files were modified by bedtools (2.26.0) for pile-up visualization via IGV. Unmodified BED files were reported as very high-quality data by phantompeakqualtools (1.2.2).

RESULTS

An integrated analysis of Insm1, Neurod1, and Pax6, three Neurog3-dependent pro-endocrine transcription factors. Comparison of the gene expression profiles of purified *Neurog3*$^{GFP/+}$ and *Neurog3*$^{GFP/GFP}$ (*Neurog3* KO) cells from E15.5 embryos by bulk RNA-Seq (Osipovich et al., 2021) revealed a marked reduction in the expression of pro-endocrine TFs *Insm1, Neurod1,* and *Pax6* (**Fig. 2.2 A**). The marked dysregulation of these genes provides further evidence that each lies downstream of *Neurog3* in a Neurog3-driven endocrine cell-specific gene regulatory network **(Fig. 2.2B)** essential for the formation and function of pancreatic endocrine cells (Gierl et al., 2006; Mellitzer et al., 2006; Naya et al., 1997; St-Onge et al., 1997).

To systematically assess and compare the independent effects of *Insm1, Neurod1,* and *Pax6* in developing endocrine cells, I next intercrossed mice containing *Insm1*^{GFPCre}, *Neurod1*^{LacZ}, and *Pax6*^{flox} alleles. From timed matings, I obtained embryos that were heterozygous for *Insm1* (*Insm1*$^{+/-}$) for use as controls, embryos that were globally deficient in *Insm1* (*Insm1* KO) or *Neurod1* (*Neurod1* KO), and embryos that lacked *Pax6* specifically in *Insm1*-expressing cells (*Pax6* KO) (**Fig. 2.3**). Our strategy utilized green fluorescence from the *Insm1*^{GFPCre} allele to selectively purify endocrine cells by FACS at E15.5 for RNA-Seq analysis and to perform accurate quantification by immunohistochemistry in tissue samples from E18.5 embryos (**Fig. 2.3**). The findings were then comparatively analyzed.

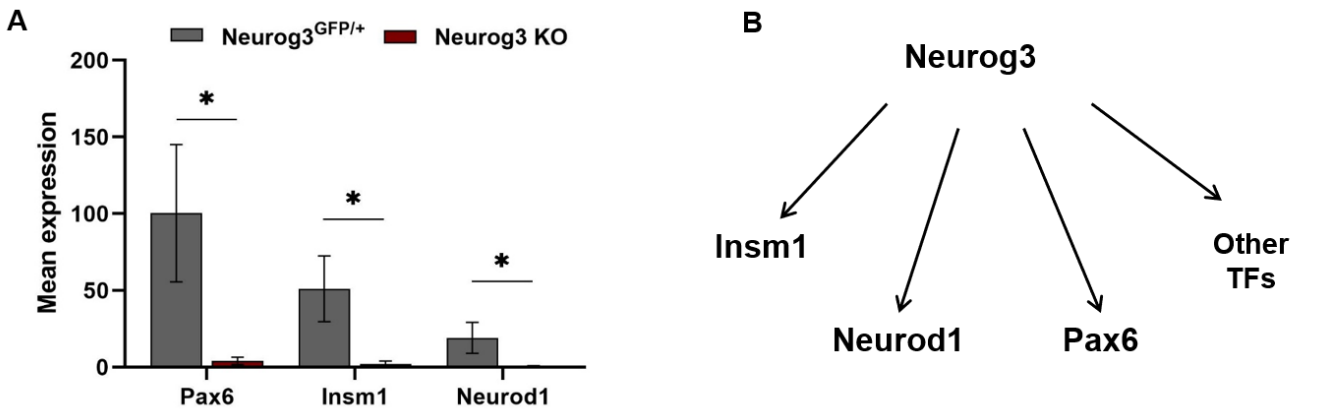


Figure 2.2. Neurog3 drives expression of *Insm1*, *Neurod1*, and *Pax6*. RNA-Sequencing of *Neurog3*$^{+/-}$ and *Neurog3*$^{-/-}$ endocrine cells at E15.5 shows the reduced expression of pro-endocrine TFs (**A**) indicating that *Neurog3* is an upstream regulator of *Insm1*, *Neurod1*, and *Pax6* (**B**). Error bars in (**A**) indicate the standard error of the mean, * p-value < 0.05, Student's unpaired t-test.

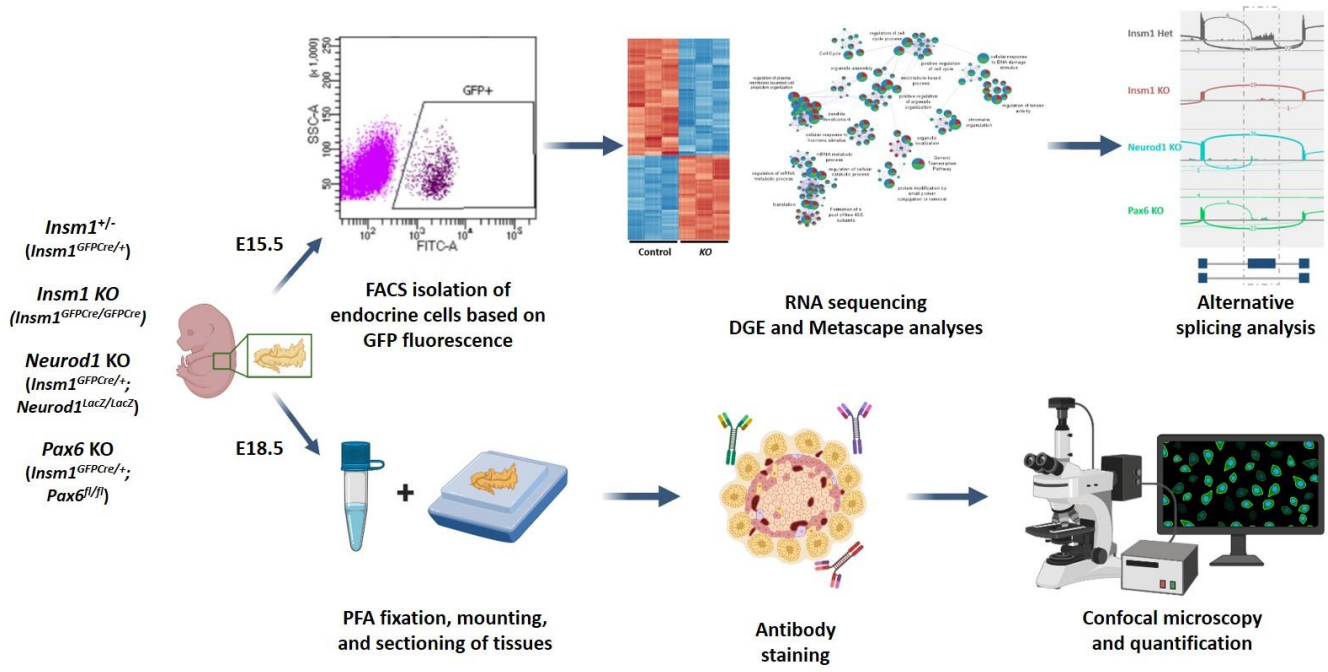


Figure 2.3. An integrated approach for analyzing the roles of Insm1, Neurod1, and Pax6 in pancreatic endocrine cell development.

Control and KO embryos were collected at either E15.5 or E18.5. GFP-positive cells from E15.5 pancreata were FACS sorted and collected for RNA-Seq, followed by differential expression, Metascape, and differential splicing analyses. Pancreata from E18.5 embryos were fixed in PFA, sections stained for nuclear GFP signal, endocrine hormones, Ki67 and TUNEL reporter, and marked cells were subsequently quantified. Schematic was generated with BioRender.com.

Mice lacking Insm1, Neurod1, or Pax6 exhibit defects in endocrine cell differentiation, proliferation, and apoptosis. Immunofluorescence staining and morphometric analysis of pancreatic tissues from *Insm1*$^{+/-}$ (control) and *Insm1* KO, *Neurod1* KO, and *Pax6* KO embryos at E18.5 was done by co-staining and quantifying the percentage of pancreatic hormone-positive cells, cell proliferation marker Ki67-positive, or apoptotic TUNEL-positive cells per total number of *Insm1*-driven GFP-positive endocrine cells.

Similar to prior reports, embryos that lack *Insm1* exhibit both a disruption in endocrine cell organization and a reduced number of endocrine cells (Gierl et al., 2006; Osipovich et al., 2014). *Insm1* KO pancreata have a lower number of insulin-, glucagon-, and somatostatin-expressing endocrine cells and an increase in pancreatic polypeptide-expressing cells, indicating profound defects in endocrine differentiation and lineage specification **(Figs. 2.4B, 2.4F and 2.4I-L)**. Immunostaining with Ki67 revealed reduced proliferation of the *Insm1* KO endocrine cells (3.6% vs. 15% of control cells) **(Figs. 2.5B, 2.5F and 2.5I-J),** whereas the number of apoptotic cells between the control and KO cells was unchanged.

As in previous studies, *Neurod1* KO embryos also exhibited a lower overall number of endocrine cells, disrupted islet organization, and increased percentage of pancreatic polypeptide-expressing endocrine cells increased **(Figs. 2.4C, 2.4G and 2.4I-L)**. The *Neurod1* KO embryos had fewer Ki67-positive endocrine cells (2.8% vs. 15% in the controls) and a marked increase in TUNEL-positive cells (4.5% vs. 0.6% in control animals) **(Figs. 2.5C, 2.5G and 2.5I-J)**, which is also like prior reports (Naya et al., 1997; Romer et al., 2019).

Similarly, analysis of the *Pax6* KO embryos revealed disorganized endocrine islets and a reduced overall number of endocrine cells. However, in these mice, there was a marked decrease in the number of glucagon-positive cells (**Figs. 2.4D and 2.4J**), although the numbers of other hormone-expressing GFP-positive cells were not significantly affected (**Figs. 2.4H, 2.4K and 2.4L**). There was also a considerable reduction in the number of Ki67-positive cells in the *Pax6* KO animals (4% vs. 15% in the controls) (**Figs. 2.5D and 2.5I**) and an increase in the number of TUNEL-positive endocrine cells (**Figs. 2.5H and 2.5J**). Notably, I had difficulty obtaining *Insm1*$^{GFPCre/+}$*, Pax6* KO embryos at E18.5 and only succeeded in obtaining two embryos of the sought-after genotype out of 62 embryos genotyped (p = 4.77E^{-05} by chi-squared test), suggesting that the haplo-insufficiency of *Insm1* may compound the effects of the *Pax6* KO alone (**Fig. 2.5K**).

Together, these analyses indicate that mice lacking *Insm1, Neurod1,* or *Pax6* have abnormal islet morphologies, a reduced number of endocrine cells, and varying defects in differentiation towards hormone-expressing cells. Moreover, all three KOs exhibited decreased endocrine cell proliferation rates, with both the *Neurod1* and *Pax6* KOs also showing increased apoptosis.

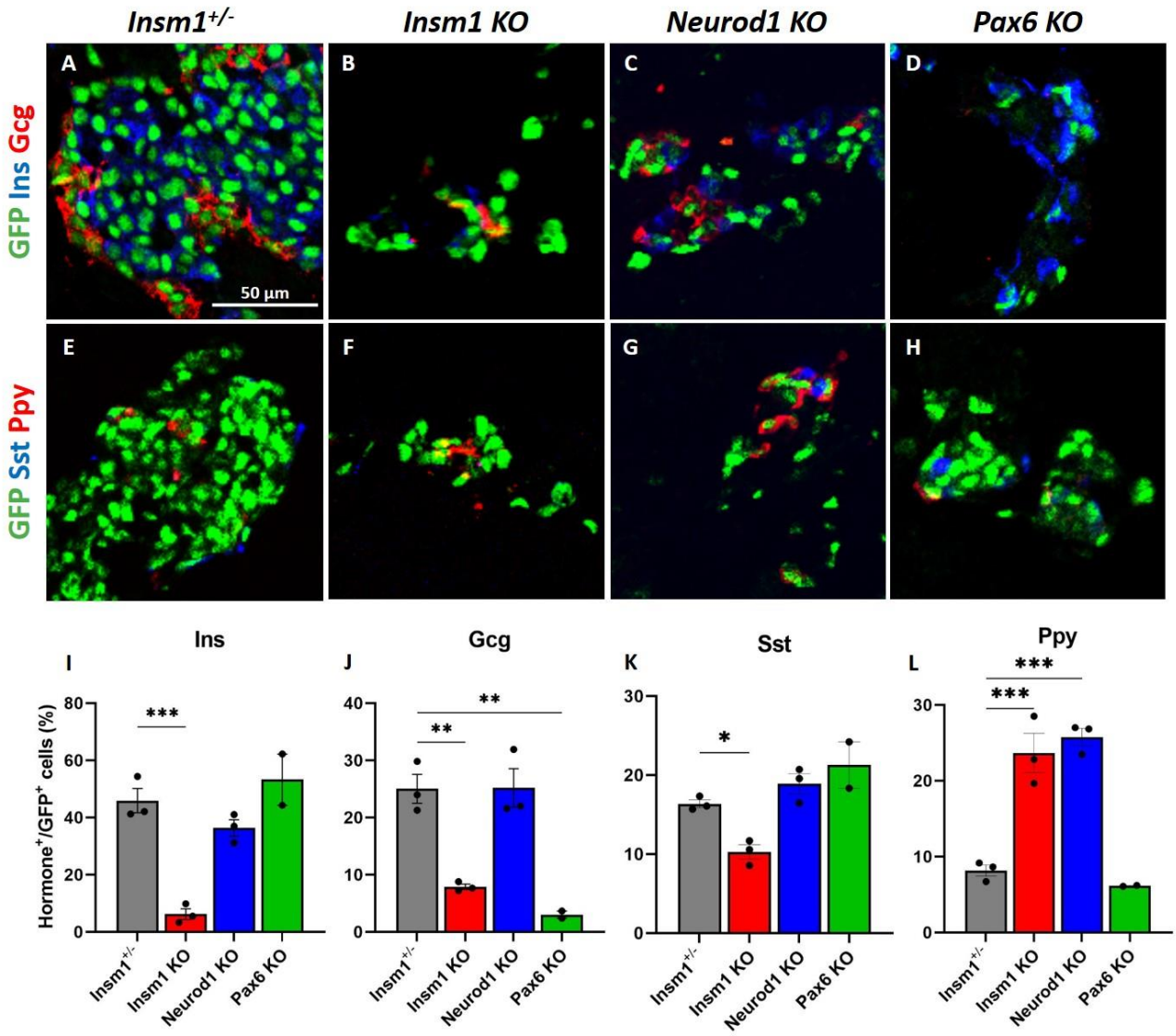
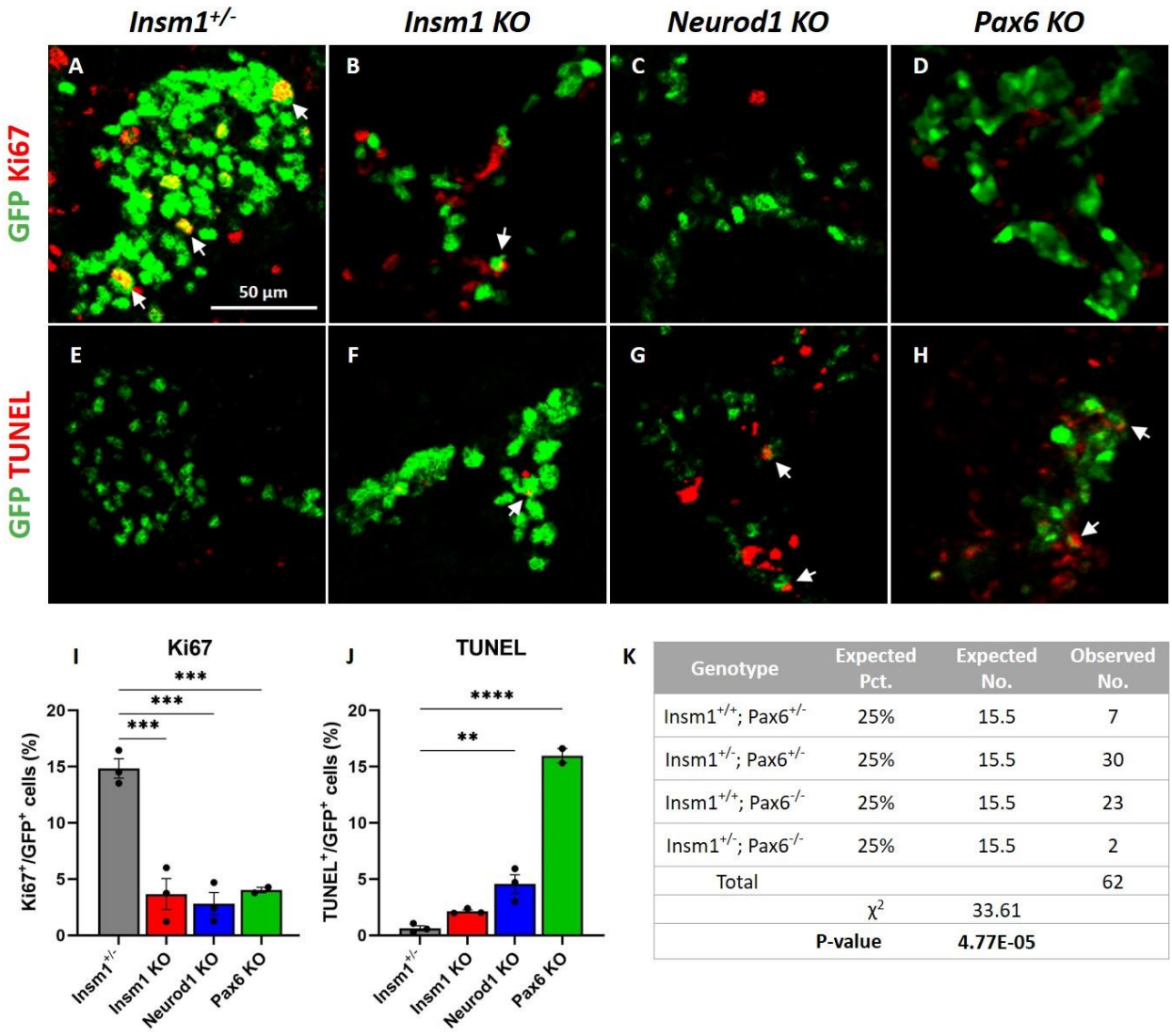


Figure 2.4. Impaired differentiation of pancreatic endocrine cells in *Insm1*, *Neurod1*, and *Pax6* KO embryos. (**A-D**) Immunofluorescence labeling of pancreata from E18.5 *Insm1*^{GFP}-expressing embryos using antibodies against GFP (green) that marks pre-endocrine cells, and pancreatic hormones insulin (Ins, blue), and glucagon (Gcg, red). Compared with *Insm1*$^{+/-}$ mice (**A**), mice lacking *Insm1* (***B***), *Neurod1* (***C***), *and Pax6* (***D***) exhibit a decrease in total number of endocrine cells, altered endocrine cell morphology and numbers of hormone expressing cells. (**E-H**) Immunofluorescence labeling of pancreata from E18.5 embryos using antibodies against GFP (green), pancreatic polypeptide (Ppy, red), and somatostatin (Sst, blue) shows altered numbers of hormone expressing cells in *Insm1* (**F**), *Neurod1* (**G**), and *Pax6* (**H**) KO mice. (**I-L**) Quantification of a percentage of hormone-positive cells among GFP-positive endocrine cells demonstrates defects in differentiation of cells positive for hormones: insulin (Ins) (**I**), glucagon (Gcg) (**J**), somatostatin (Sst) (**K**), and pancreatic polypeptide (Ppy) (**L**) in *Insm1*, *Neurod1*, and *Pax6* KO embryonic pancreata in comparison with *Insm1*$^{+/-}$. Error bars indicate SEM (*n*=3); *P*-values were determined by one-way ANOVA test. Asterisks indicate *p*-values of *<0.05, **<0.01, ***<0.001. Scale bars: 50 µm.



K

Genotype	Expected Pct.	Expected No.	Observed No.
$Insm1^{+/+}$; $Pax6^{+/-}$	25%	15.5	7
$Insm1^{+/-}$; $Pax6^{+/-}$	25%	15.5	30
$Insm1^{+/+}$; $Pax6^{-/-}$	25%	15.5	23
$Insm1^{+/-}$; $Pax6^{-/-}$	25%	15.5	2
Total			62
	χ^2	33.61	
	P-value	**4.77E-05**	

Figure 2.5. Decreased proliferation of endocrine cells in *Insm1*, *Neurod1*, and *Pax6* KOs, and increased apoptosis in *Neurod1* and *Pax6* KO embryos. (**A-D**) Immunofluorescence labeling of pancreatic tissues from E18.5 $Insm1^{GFP}$-expressing embryos using antibodies against GFP (green) that marks all pre-endocrine cells, and antibodies against cell proliferation marker Ki67 (red). Compared to controls (**A**), mice lacking Insm1 (**B**), Neurod1 (**C**), and Pax6 (**D**) exhibit a decrease in the number of Ki67-positive endocrine cells. Arrows indicate cells co-expressing GFP and Ki67. (**E-H**) Immunofluorescence labeling of pancreatic tissues from E18.5 embryos with antibodies against GFP (green) and TUNEL assay (red) marking positive apoptotic events. Compared with $Insm1^{+/-}$ mice (**E**), mice lacking Insm1 (**F**), Neurod1 (**G**), and Pax6 (**H**) exhibit an increase in the number of endocrine cells positive for both TUNEL and GFP. Arrows indicate TUNEL positive cells co-expressing GFP. (**I**) Quantification of a percentage of Ki67-positive cells among GFP-positive cells demonstrates a proliferation defect in endocrine cells from Insm1, Neurod1 and Pax6 KO pancreata at E18.5. (**J**) Quantification of a percentage of TUNEL-positive cells among GFP-positive cells demonstrates increased apoptosis in endocrine cells in Neurod1 and Pax6 KO pancreata at E18.5. Error bars indicate SEM (n=3); P-values were determined by one-way ANOVA test. Asterisks indicate p-values of *<0.05, **<0.01, ***<0.001. Scale bars: 50 µm.

Pancreatic pre-endocrine cells lacking Insm1, Neurod1, and Pax6 have distinct transcriptional profiles. Next, to better understand how each factor contributes to the pro-endocrine cell transcriptome, I performed bulk RNA-Seq on FACS-purified cell populations isolated from E15.5 embryos. Cells were isolated at this earlier time point to minimize variance in gene expression that may be brought on by maturation towards specific endocrine subtypes. Hierarchical clustering and principal component analysis (PCA) of the 12 RNA-Seq samples revealed tight clustering by genotype, indicating the low variance between samples and that each condition analyzed has a unique transcriptional profile (**Figs. 2.6A and 2.6B**).

Dysregulated expression of genes involved in hormone secretion, RNA metabolism and processing, and cell development in Insm1 KO endocrine cells. We previously reported the gene expression profile of both *Insm1*$^{+/-}$ and *Insm1* KO mice at E15.5 using a legacy RNA amplification and sequencing method (Osipovich et al., 2014). To accurately compare the transcriptomes of the three TF KOs in this study, I collected new *Insm1* KO datasets using the methods, reagents, and instrumentation described herein. A comparison of the two different *Insm1*$^{+/-}$ and *Insm1* KO datasets indicated that they clustered first by amplification method (Nugen vs. Clontech) and secondarily by genotype (*Insm1*$^{+/-}$ control vs. *Insm1* KO), indicating a batch effect. However, a linear regression analysis of the $\log_2$-fold changes of differentially expressed genes from the old and new datasets showed a strong correlation (R^2=0.85) (**Figs. 2.7A-C**).

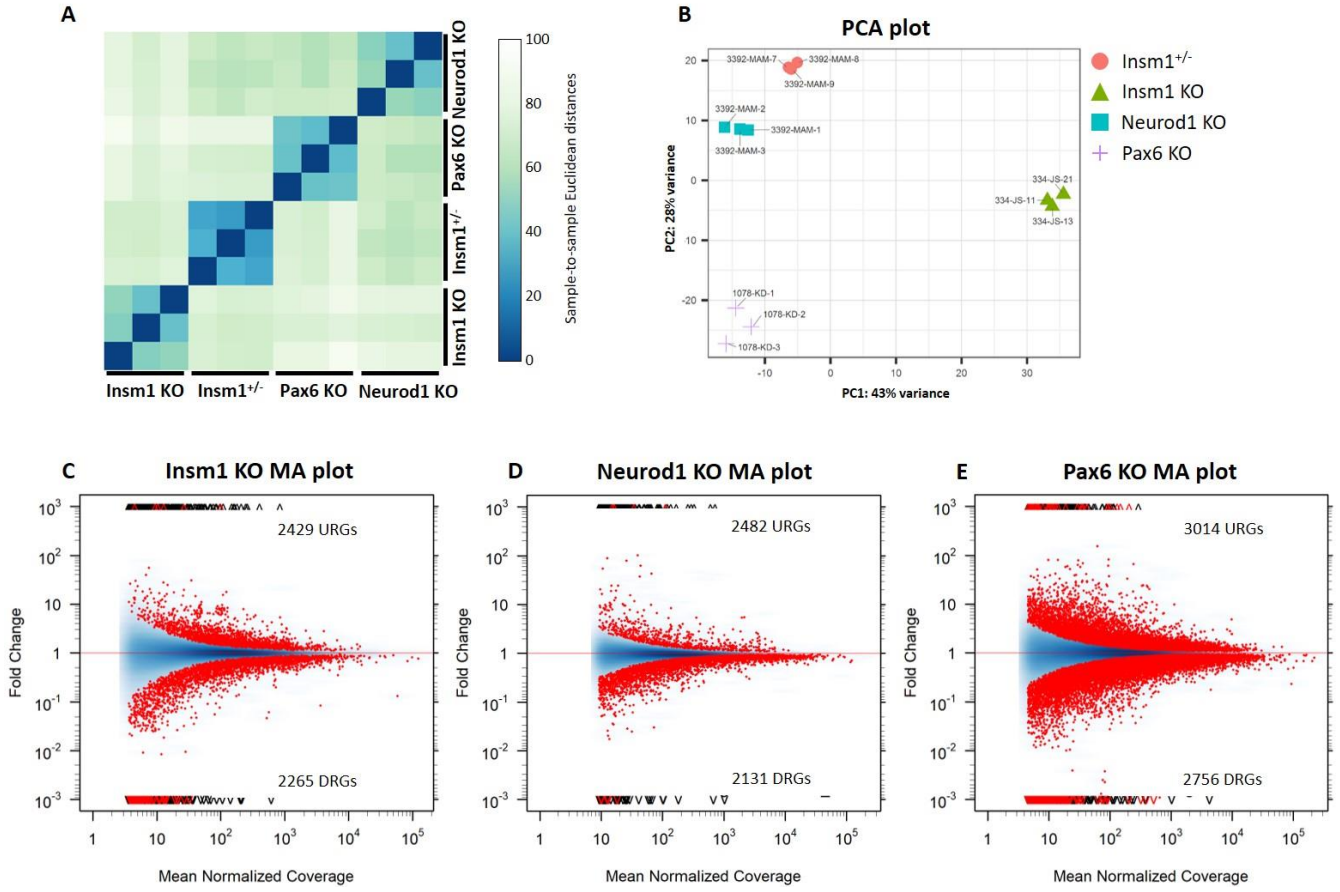


Figure 2.6. RNA-Seq Quality control analyses and comparison of *Insm1*+/-, *Insm1*-, *Neurod1*-, and *Pax6*-KO datasets reveal low sample variance within each of the RNA-Seq conditions. (**A**) Hierarchical sample clustering of all RNA-Seq datasets shows that replicates for each of the sequenced condition cluster together. The color key indicates the distance of similarity between samples. (**B**) Principal component analysis (PCA). The variance between all 12 samples is plotted across PC1 (y-axis) and PC2 (x-axis) which account for 43% and 28% of variability, respectively. (**C-E**) MA plots of Insm1, Neurod1, and Pax6 KO datasets compared to Insm1$^{+/-}$ controls. The $\log_2$ fold change (y-axis) is plotted against the mean normalized counts (x-axis). The provided number of genes changed is based on a p-value ≤ 0.05.

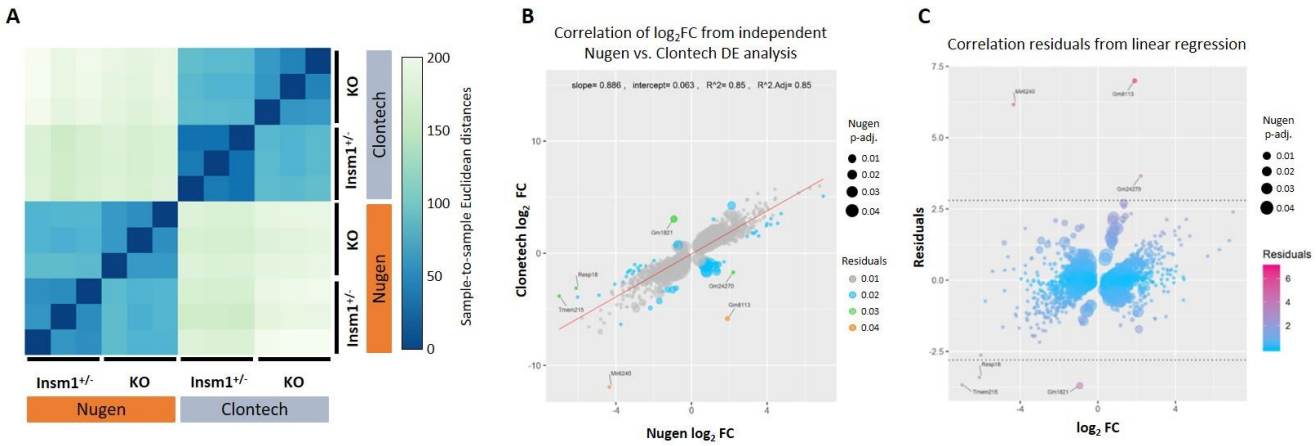
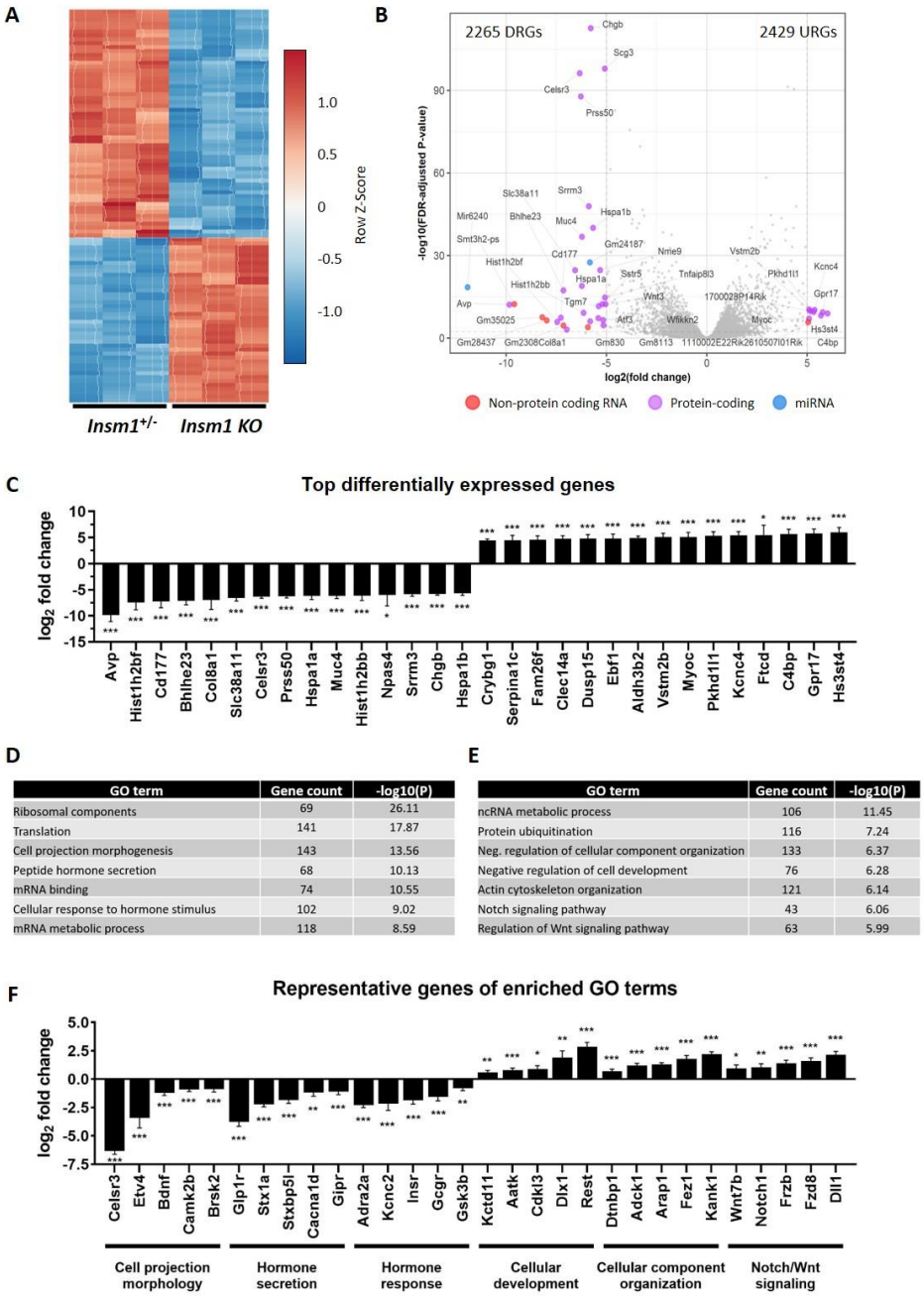


Figure 2.7. Comparative analysis of newly generated *Insm1* KO RNA-Seq datasets and previously published *Insm1* KO data. (**A**) Hierarchical clustering analysis of normalized expression data from Insm1$^{+/-}$ and Insm1$^{-/-}$ samples prepared using either the Ovation RNA-Seq system (NuGEN, CA) as in the previous study (Osipovich et al., 2014) or with the SMART-Seq® v4 Ultra® Low Input RNA Kit for Sequencing (Clontech/Takara), used in the present study. (**B**) Correlation analysis of original data vs. data presented in this study. Regression line representing the log_2 fold change of the original datasets (x-axis) against the log_2 fold change of the new datasets (y-axis) from independent DE analyses. Residual genes are highlighted and labeled. (**C**) Correlation residuals from the linear regression analysis. The resulting residual genes are plotted against the log_2 fold change in expression value. Residual outliers are highlighted and labeled. The size of the individual gene nodes corresponds to the adjusted p-value for that gene from $p < 0.01$ through $p < 0.05$.

Analysis of the new *Insm1*$^{+/-}$ and *Insm1* KO datasets by DESeq revealed a total of 4,694 differentially expressed genes, 2,265 were categorized as downregulated genes (DRGs) and 2,429 upregulated genes (URGs) in the KO compared to control embryos (p-value ≤ 0.05) (**Figs. 2.6C and 2.8A-B**). Among the top DRGs were numerous TFs known to control endocrine cell development, including both *Sox9* (McDonald et al., 2012) and *Mnx1,* which is necessary for endocrine cell fate allocation and in later stages for maintaining the β-cell fate (Pan et al., 2015; Sund et al., 2001; Wang et al., 2002). Also, among the DRGs were the TF *Bhlhe23*, the histone modifiers *Hist1h2bf* and *Histh2bb*, and genes involved in the trafficking and secretion of insulin (*Avp, Chgb*) (**Fig. 2.8C)**. Gene ontology (GO) and pathway enrichment analysis results from DRGs revealed enrichment of many genes, including those involved in cell projection morphogenesis (*Bdnf*, *Brsk2*, *Camk2b*, *Etv4*), cellular response to hormone stimulus (*Adra2a*, *Gcgr*, *Insr*, *Kcnc2*), hormone secretion (*StxbKp5l*, *Stx1a*, *Cacna1d*), and mRNA metabolic process (*Srsf2*, *Hnrnpc*, *Celf4*, *Celf6*) (**Figs. 2.8D and 2.8F**).

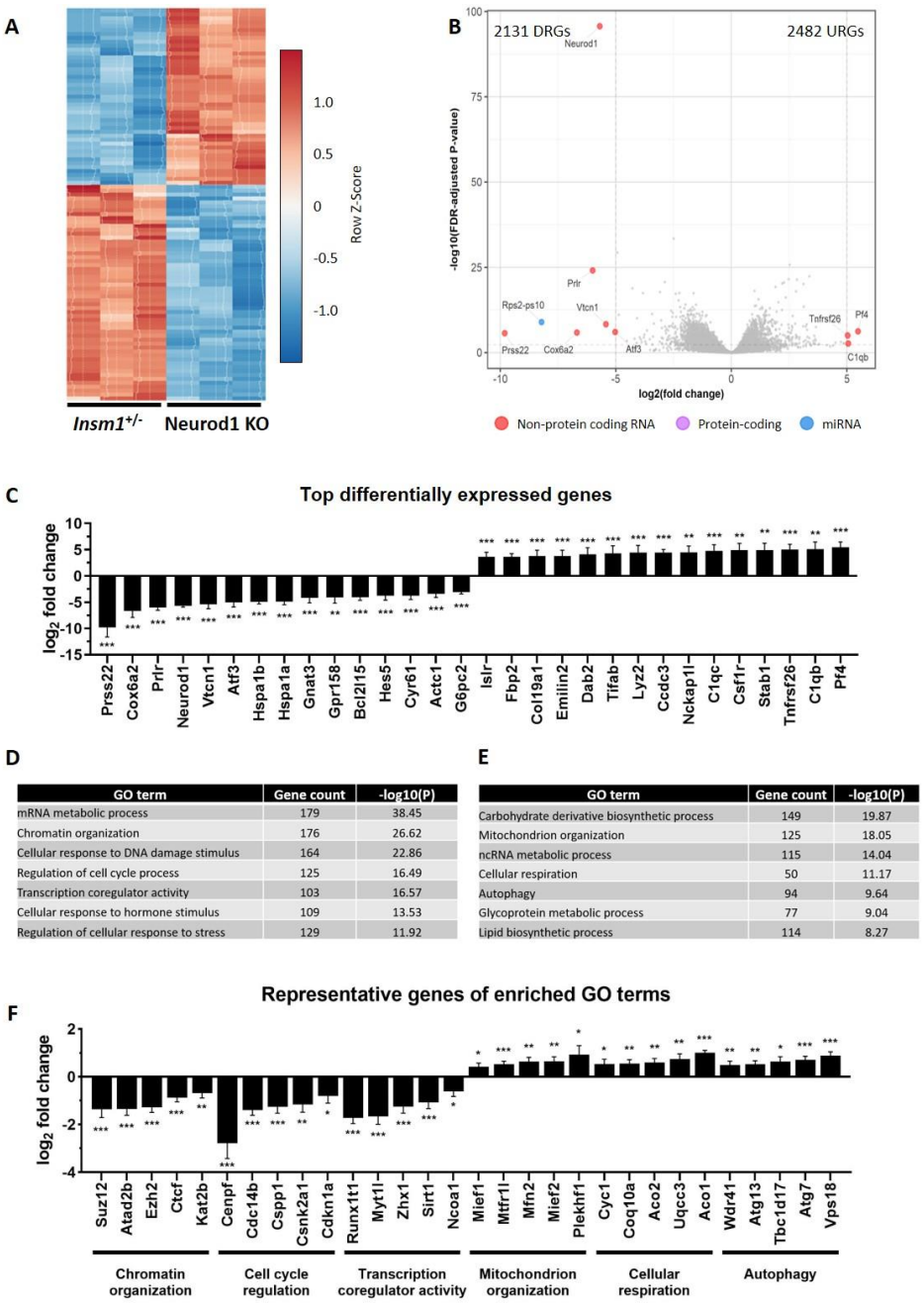
Among the top URGs, there were many genes necessary for the key functions of mature islet cells. Examples include *Kcnc4* (Kv3.4), a voltage-gated potassium channel expressed in β- and δ-cells; and *Gpr17*, an orphan G protein-coupled receptor known to inhibit neural cell maturation (Chen et al., 2009; Gopel et al., 2000; Jacobson and Philipson, 2007). GO and pathway enrichment analyses identified an enrichment in genes involved in the regulation of cellular component organization (*Adck1*, *Arap1*), Notch and Wnt signaling pathways (*Dll1*, *Notch1/2*, *Wnt7b*, *Frzb*, *Frz8*), and the negative regulation of cellular development (*Aatk, Dlx1, Rest, Kctd11*) (**Figs. 2.8E and 2.8F**). These results indicate that Insm1 stimulates many genes involved in hormone secretion and the function of mature endocrine cells while repressing genes associated with Notch and Wnt signaling.



GO term	Gene count	-log10(P)
Ribosomal components	69	26.11
Translation	141	17.87
Cell projection morphogenesis	143	13.56
Peptide hormone secretion	68	10.13
mRNA binding	74	10.55
Cellular response to hormone stimulus	102	9.02
mRNA metabolic process	118	8.59

GO term	Gene count	-log10(P)
ncRNA metabolic process	106	11.45
Protein ubiquitination	116	7.24
Neg. regulation of cellular component organization	133	6.37
Negative regulation of cell development	76	6.28
Actin cytoskeleton organization	121	6.14
Notch signaling pathway	43	6.06
Regulation of Wnt signaling pathway	63	5.99

Figure 2.8. Transcriptional profiling of *Insm1* KO pancreatic endocrine cells at E15.5. (**A**) Heat map representation of the top 100 variant genes (based on *p*-value) dysregulated in *Insm1 KO* relative to *Insm1*$^{+/-}$. (**B**) Volcano plot of differentially expressed genes (DEGs) between control and KO samples plotting the $\log_2$ fold change (*x*-axis) against the $-\log_{10}$ FDR-adjusted *p*-value (*y*-axis). Top differentially expressed genes (based on $\log_2$ fold change) are indicated by names. (**C**) Top 15 up and downregulated genes by $\log_2$ fold change. (**D**) Select enriched terms identified by Gene Ontology (GO) and pathway analyses of downregulated genes (DRGs) and (**E**) upregulated genes (URGs) in *Insm1* KO samples. The tables display enriched biological and molecular processes terms, the number of genes represented in that term, and their $\log_2$ *p*-value. Only genes with an adjusted *p*-value ≤ 0.05 were analyzed. The analyses were performed using Metascape. (**F**) Bar graph of the $\log_2$ fold change of select example genes from enriched functional categories.



GO term	Gene count	-log10(P)
mRNA metabolic process	179	38.45
Chromatin organization	176	26.62
Cellular response to DNA damage stimulus	164	22.86
Regulation of cell cycle process	125	16.49
Transcription coregulator activity	103	16.57
Cellular response to hormone stimulus	109	13.53
Regulation of cellular response to stress	129	11.92

GO term	Gene count	-log10(P)
Carbohydrate derivative biosynthetic process	149	19.87
Mitochondrion organization	125	18.05
ncRNA metabolic process	115	14.04
Cellular respiration	50	11.17
Autophagy	94	9.64
Glycoprotein metabolic process	77	9.04
Lipid biosynthetic process	114	8.27

Figure 2.9. Transcriptional profiling of *Neurod1* KO pancreatic endocrine cells at E15.5. (**A**) Heat map representation of the top 100 variant genes (based on *p*-value) dysregulated in *Neurod1* KO relative to *Insm1*$^{+/-}$. (**B**) Volcano plot of differentially expressed genes (DEGs) between control and KO samples plotting the $\log_2$ fold change (*x*-axis) against the $-\log_{10}$ FDR-adjusted p-value (*y*-axis). Top differentially expressed genes (based on $\log_2$ fold change) are indicated by names. (**C**) Top 15 up and downregulated genes by $\log_2$ fold change. (**D**) Select enriched terms identified by Gene Ontology (GO) and pathway analyses of downregulated genes (DRGs) and (**E**) upregulated genes (URGs) in *Neurod1*-null samples. The tables display enriched biological and molecular processes terms, the number of genes represented in that term, and their $\log_2$ *p*-value. Only genes with an adjusted *p*-value ≤ 0.05 were analyzed. The analyses were performed using Metascape. (**F**) Bar graph of the $\log_2$ fold change of select example genes from enriched functional categories.

Dysregulated expression of genes involved in chromatin organization, cell proliferation and mitochondrial function in Neurod1 KO endocrine cells. Comparing the *Neurod1* KO and control datasets revealed a different set of 4,613 dysregulated genes (2,131 DRGs and 2,482 URGs) (*p*-value ≤ 0.05) (**Figs. 2.6D and 2.9A-B**). The top DRGs included many factors known to impact islet cell function, including *Prlr, G6pc2, Hspa1a,* and *Hspa1b* (**Fig. 2.9C**). Transcriptional co-regulators (*Runx1t1, Myt1l*), cell cycle regulation (*Cdkn1a*, *Cdc14b)*, and chromatin organization (*Ctcf*, *Kat2b*) were enriched in DRGs in *Neurod1* KO (**Fig. 2.9D**).

The top URGs in this dataset included the NF-κB signaling regulators (*Tifab*, *Ccdc3*), extracellular matrix factors (*Emilin2*, *Col19a1*), and the TRAIL receptor *Tnfsrf26* **(Fig. 2.9C**). URGs were enriched in such functional categories as cellular respiration (*Coq10a*, *Uqcc3*), mitochondrial organization (*Mfn2*, *Meif1*, *Meif2*, *Plekhf1*), and autophagy (*Atg7, Atg13, Tbc1d17, Vps11)* (**Fig. 2.9D**). Moreover, genes involved in the regulation of cell death were also upregulated (*P2rx1*, *P2rx7, C6*). Identification of these enriched GO terms and pathways for the *Neurod1* KO mice parallels the changes we and others have observed in islet morphology, including reduced endocrine cell proliferation and increased cell death (Naya et al., 1997; Romer et al., 2019). These findings indicate that the loss of *Neurod1* causes a transcriptional response marked by the upregulation of genes involved in NF-kB signaling and energy metabolism and the downregulation of other pancreatic TFs and genes associated with cell cycle regulation.

Dysregulated expression of genes involved in cell cycle regulation, developmental growth and apoptosis in Pax6 KO endocrine cells. Comparison of the *Pax6* KO embryos and controls revealed a set of 5,770 genes (2,756 DRGs and 3,014 URGs) that were Pax6 dependent (*p*-value ≤ 0.05) (**Figs. 2.6E and 2.10A-B**). Among the top URGs were *Prlr*, *G6pc2*, *Gcg*, *Ptgdr*, and *Ccdc3* (**Fig. 2.10C**). Inspection of the enriched GO terms and pathways for Pax6-dependent DRGs and URGs suggests a critical role for Pax6 in stimulating cell growth and cell cycle and regulating ER stress and apoptosis in developing endocrine cells (**Figs. 2.10D-E**). Specifically, many genes involved in chromatin organization (*Ash1l, p300, Mettl4*), cell cycle regulation (*Ccnd1, Cdc27, Cdkn1b, Chek1*), and developmental cell growth (*Fgf9, Igf1r, Tgfb2*) were downregulated in *Pax6* KO animals (**Figs. 2.10D and 2.10F**). Consistent with the diminished number of alpha cells observed at E18.5, there was a marked reduction in *Gcg* expression, as well as dysregulation of other α-cell specific genes *Brn4* and *Pcsk2*. GO terms and pathways enriched in URGs include the cellular component organization (*Elmo1/2*, *Lima1*, *Myo7a*), response to ER stress (*Atf6b*, *Traf2*), and the apoptotic signaling pathway (*Casp2*, *Casp9*, *Madd*) (**Figs. 2.10E-F**). Analysis of dysregulated

genes in *Pax6* KO animals shows the downregulation of other TFs, cell cycle regulators, and genes important for cellular growth, and the upregulation of apoptotic signaling and genes involved in response to ER stress.

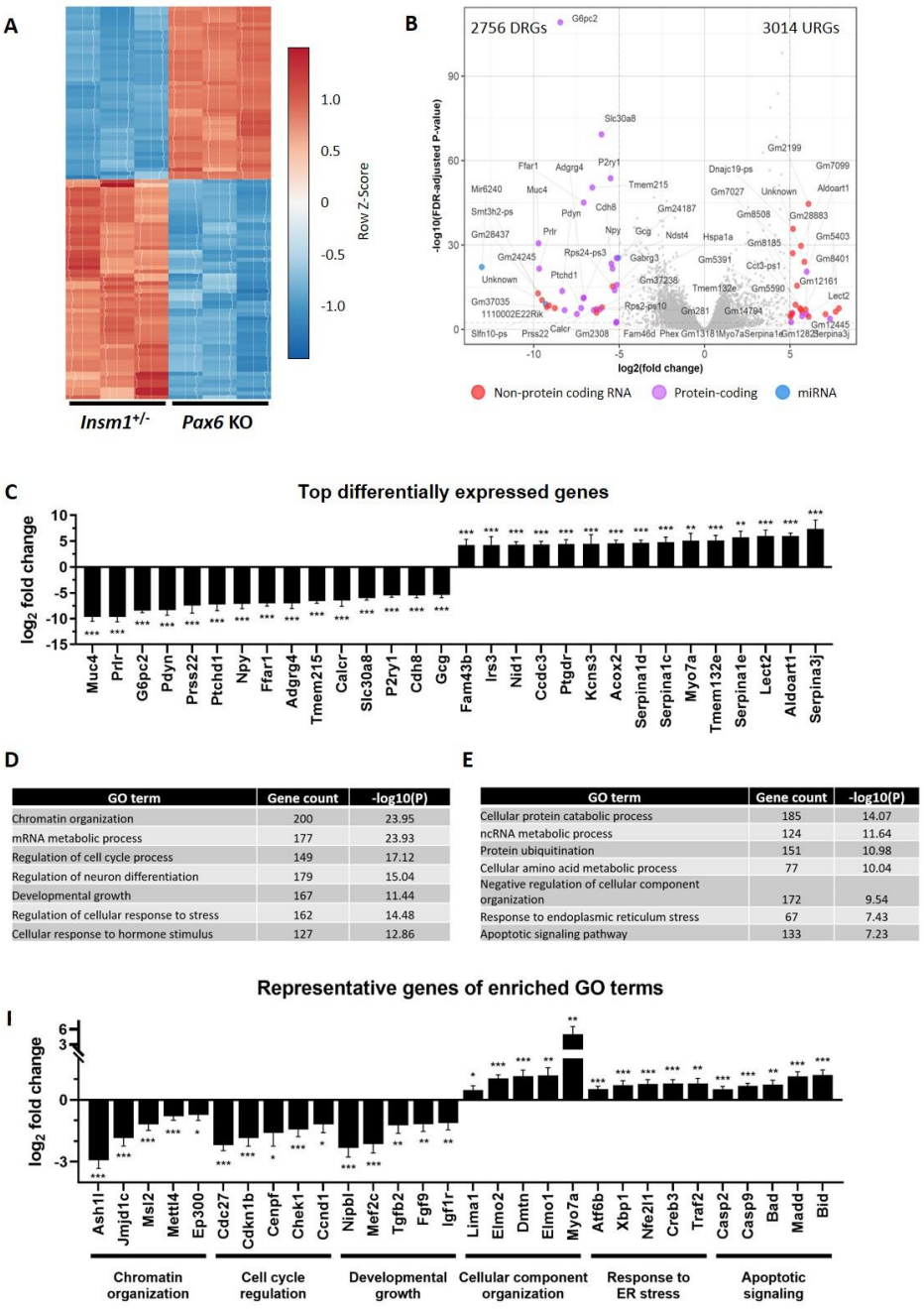


Figure 2.10. Transcriptional profiling of *Pax6* KO pancreatic endocrine cells at E15.5. (**A**) Heat map representation of the top 100 variant genes (based on p-value) dysregulated in *Pax6* KO relative to *Insm1*$^{+/-}$. (**B**) Volcano plot of differentially expressed genes (DEGs) between control and KO samples plotting the $\log_2$ fold change (*x*-axis) against the $-\log_{10}$ FDR-adjusted p-value (*y*-axis). Top differentially expressed genes (based on $\log_2$ fold change) are indicated by names. (**C**) Top 15 up and downregulated genes by $\log_2$ fold change. (**D**) Select enriched terms identified by Gene Ontology (GO) and pathway analyses of downregulated genes (DRGs) and (**E**) upregulated genes (URGs) in *Pax6*-null samples. The tables display enriched biological and molecular processes terms, the number of genes represented in that term, and their $\log_2$ *p*-value. Only genes with an adjusted *p*-value ≤ 0.05 were analyzed. The analyses were performed using Metascape. (**F**) Bar graph of the $\log_2$ fold change of select example genes from enriched functional categories.

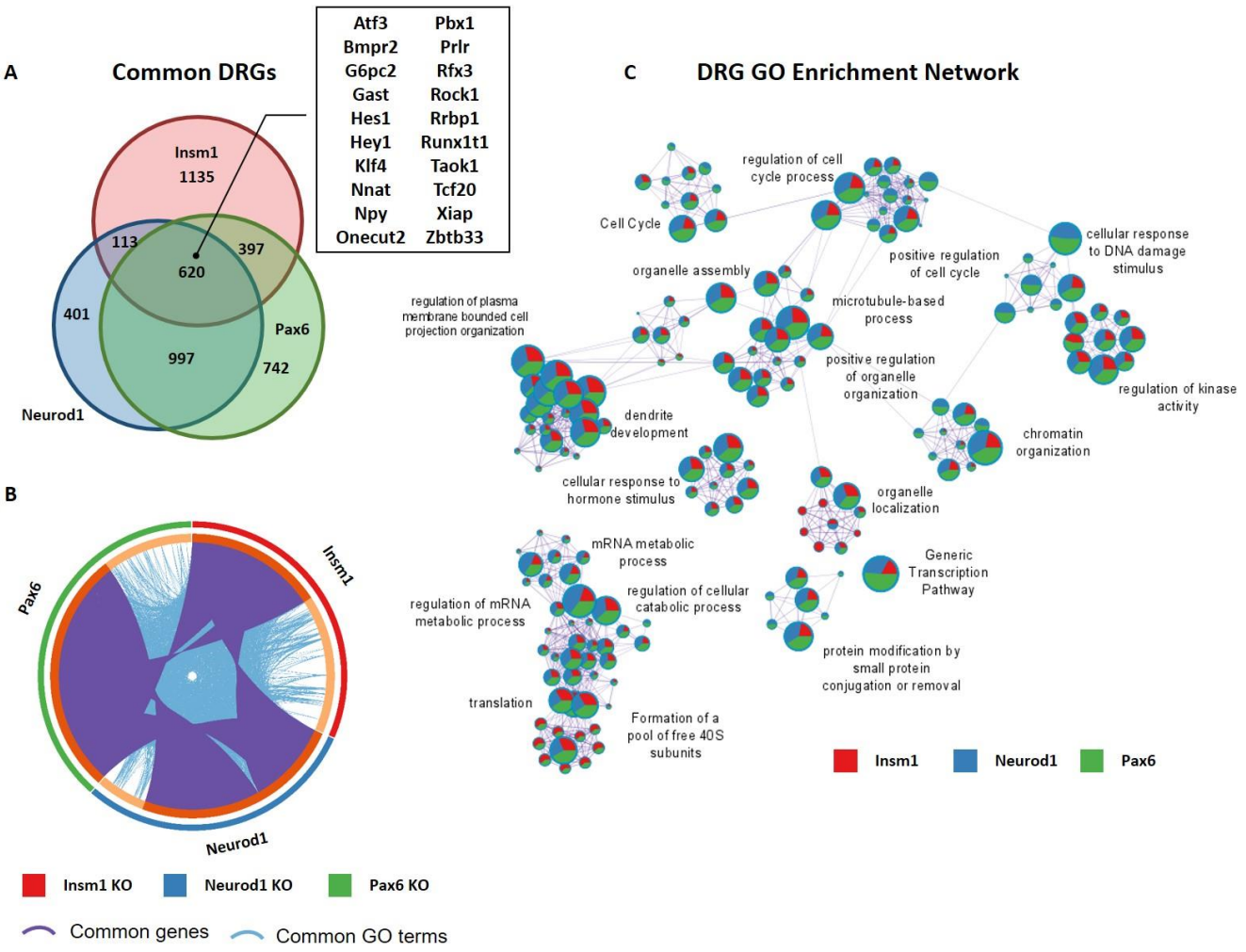


Figure 2.11. Enriched GO terms and pathways common for genes downregulated in *Insm1*, *Neurod1*, and *Pax6* KO pancreatic endocrine cells. (**A**) Venn diagram shows overlap between downregulated genes (DRGs) (*p*-value < 0.05 cut off) in each of the pairwise comparisons: *Insm1* KO, *Neurod1* KO, and *Pax6*- KO versus Insm1$^{+/-}$ datasets. A subset of commonly downregulated genes is highlighted within a boxed insert. (**B**) Circos plot representing genes (purple curves) and GO terms/pathways (blue curves) that are shared between DRGs from the three comparisons. (**C**) Network depiction of the enriched gene ontology terms shared between DRGs from the three comparisons. Node size is proportional to the number of genes in GO or pathway category, with pie charts indicating a proportion of genes from each comparison in that GO term. Metascape analyses were run using default parameters.

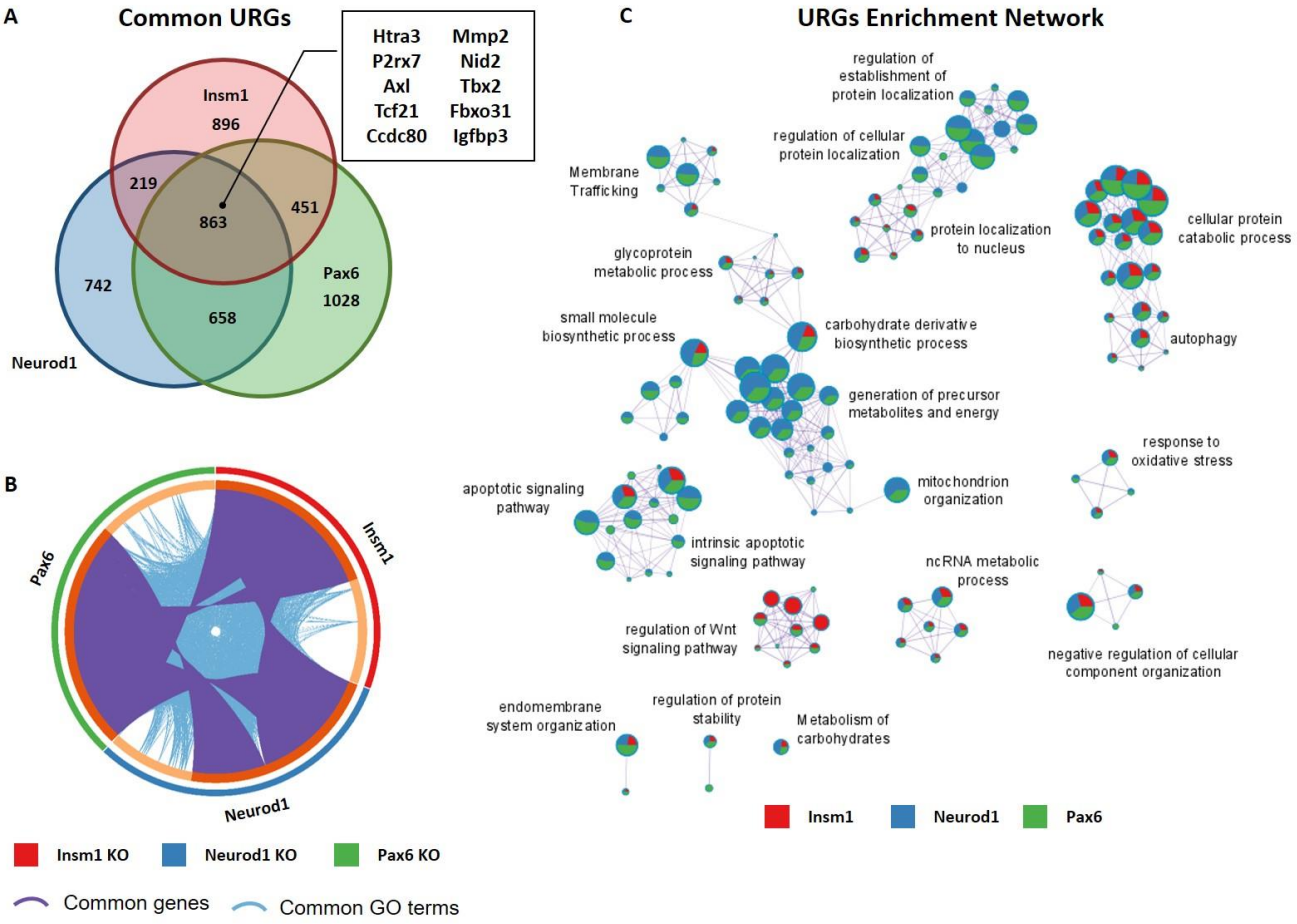


Figure 2.12. Enriched GO terms and pathways common for genes upregulated in Insm1, Neurod1 and Pax6 KO pancreatic endocrine cells. (**A**) Venn diagram shows overlap between upregulated genes (URGs) (*p*-value < 0.05 cuttoff) in each of the pairwise comparisons: *Insm1* KO, *Neurod1* KO, and *Pax6* KO versus *Insm1*$^{+/-}$ datasets. A subset of commonly downregulated genes is highlighted within a boxed insert. (**B**) Circos plot representing genes (purple curves) and GO terms/pathways (blue curves) that are shared between URGs from the three comparisons. (**C**) Network depiction of the enriched gene ontology terms shared between URGs from the three comparisons. Node size is proportional to the number of genes in GO or pathway category, with pie charts indicating a proportion of genes from each comparison in that GO term. Metascape analyses were run using default parameters.

Comparison of the gene sets dysregulated in the Insm1, Neurod1, and Pax6 KO endocrine cells. To further assess the similarities and differences in the genes affected by each gene KO, I started by comparing the gene lists. I found that of the 8,729 total dysregulated genes (based on adj. p-value ≤ 0.05) across all three KOs, 4,719 (54%) were affected by more than one TF (**Figs. 2.11A and 2.12A**). Additionally, 578 (6.6%) of genes were inversely regulated across TFs. For example, *Mnx1* is upregulated in *Pax6*- and *Neurod1* KO datasets and downregulated in *Insm1* KO samples. Conversely, *Rest* is downregulated in *Pax6*- and *Neurod1* KO samples and upregulated in *Insm1* KO data. Similarly, *Fev* is upregulated in the absence of *Pax6*, unchanged in *Neurod1* KO, and downregulated in *Insm1* KO datasets. This finding indicates that while Insm1, Neurod1, and Pax6 commonly regulate a sizeable fraction of genes, each TF has unique effects on the transcriptome. To quantify these findings and reflect the similarity of the different groups, I determined the percentage of shared genes relative to the total number of dysregulated genes. This analysis revealed a range of 2.6 – 22.6% similarity between different groups, with the highest similarity between *Neurod1* and *Pax6* DRGs (22.6%).

Next, to identify commonalities in the functional annotations of genes dysregulated in each gene set, I performed Metascape analyses for both the DRGs and URGs. These analyses revealed many genes, and functional GO terms and pathways shared between dysregulated genes from each gene set (**Figs. 2.11B and 2.12B**). Of the 620 commonly regulated DRGs (14.07% shared), the enriched GO term and signaling pathway network includes such categories as regulation of mRNA metabolic processes and translation (*Akap17b, Clk1, Srsf3, Srsf6*), cell cycle regulation (*Bbx, Cdk6, Cdc14b*), and chromatin organization (*Kmt2a*, *Setd5*, *Kdm7a*) (**Fig. 2.11C**). Other common DRGs included regulators of insulin secretion and signaling (*Ins1, Nnat, Ucn3, Insr*), transcriptional regulators (*St18, Neurod1, Rfx3, Myt1l, Runx1t1, Onecut2*), and kinases (*Akt3, Mapk8, Taok1, Prkcb*). While it is not surprising that these TFs regulate other TFs that are known to be involved in pancreas development and function, it is important to note their role in affecting the expression of cell cycle regulators and chromatin organization. Tight regulation of these processes has long been linked to cellular development, expansion, and differentiation, and their dysregulation likely contributes to the observed developmental delays and perturbed ratios of islet cell types (Boward et al., 2016; Dalton, 2015; Kim et al., 2015; Krentz et al., 2017; Soufi and Dalton, 2016).

The shared functional network for 863 URGs (17.7% shared) in all three KOs contains GO terms and pathways for ncRNA processing (*Aars*, *Ctu1*, *Lsm6*, *Tsen2*), apoptotic signaling pathway (*Casp9, Tradd, Tnfrsf21,*

Ddx47), protein localization (*Ipo4, Ipo9, Timm29*), and response to oxidative stress (*Pink1*, *Selenon*, *Fancc*) (**Fig. 2.12C**). Genes involved in Wnt-signaling pathways (*Frzb, Lzts2*) and other TFs (*Sox10, Tbx2, Foxa2*) were also upregulated. The upregulation of genes associated with apoptosis and oxidative stress response directly correlate with the increases in apoptosis in *Neurod1* and *Pax6* KO pancreata. Likewise, increases in the expression of members of the Wnt signaling pathway and early endocrine TFs are reminiscent of morphological, developmental delays (Sharon et al., 2019). This is particularly true of those observed in *Insm1* KO embryos which can generate endocrine cells, but many of those are unable to properly differentiate and mature into hormone expressing endocrine cells (Mellitzer et al., 2006; Osipovich et al., 2014).

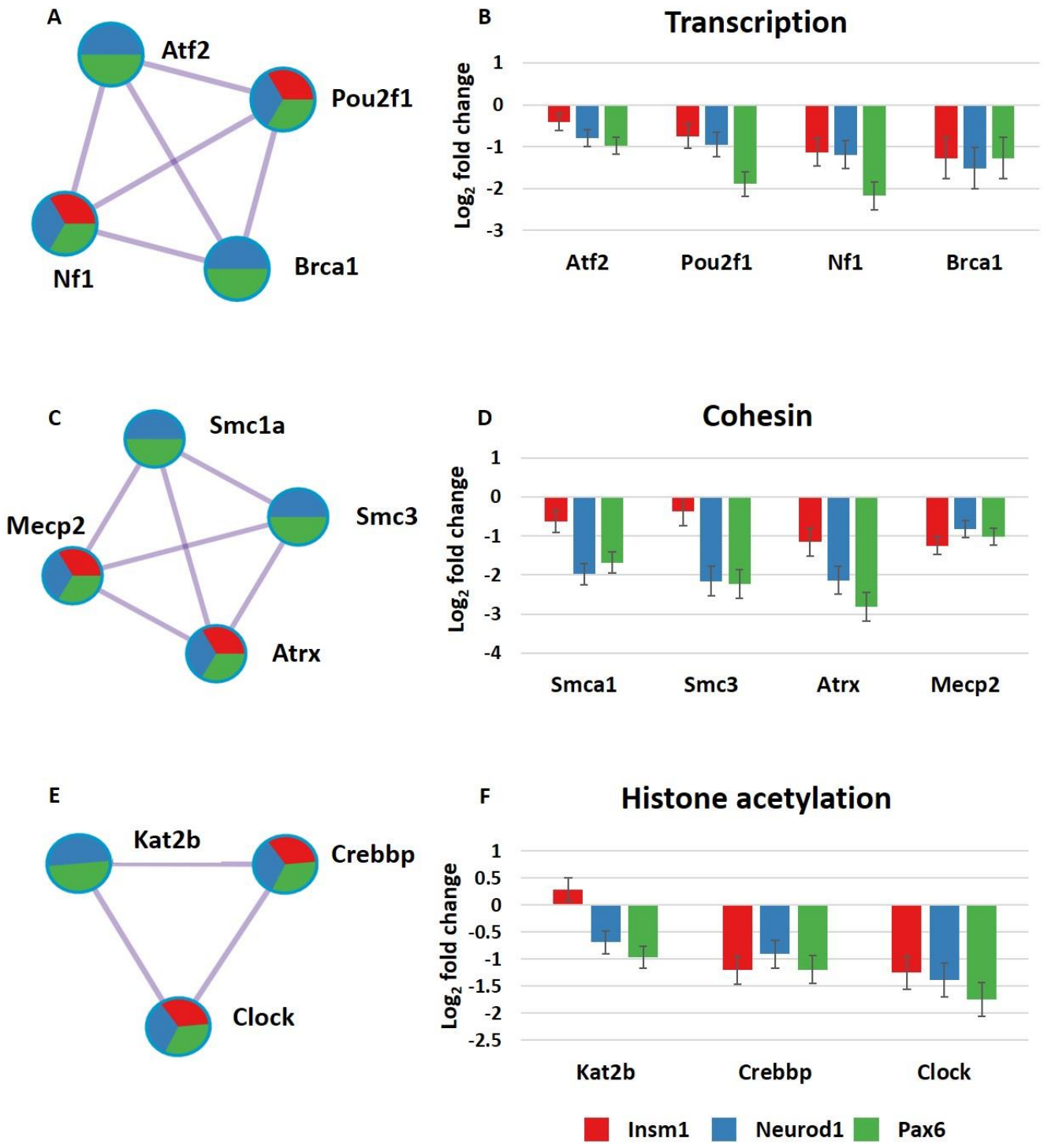


Figure 2.13. Metascape PPI MCODE analysis of DRGs in *Insm1* KO, *Neurod1* KO, and *Pax6* KO RNA-Seq datasets reveal an enrichment for transcriptional, cohesin, and histone acetylation complexes. Subset networks of the most enriched (based on p-value) modules of genes that form protein-protein interactions (PPI) specific for (**A**) transcriptional, (**C**) cohesin, and (**E**) histone acetylation complexes. (**B, D, F**) Bar graphs of the change in expression (as log_2 fold change) from our knock-out datasets of the genes included in each complex. Error bars indicate the SEM. Nodes are represented as pies in which the colored sectors correspond to their gene list.

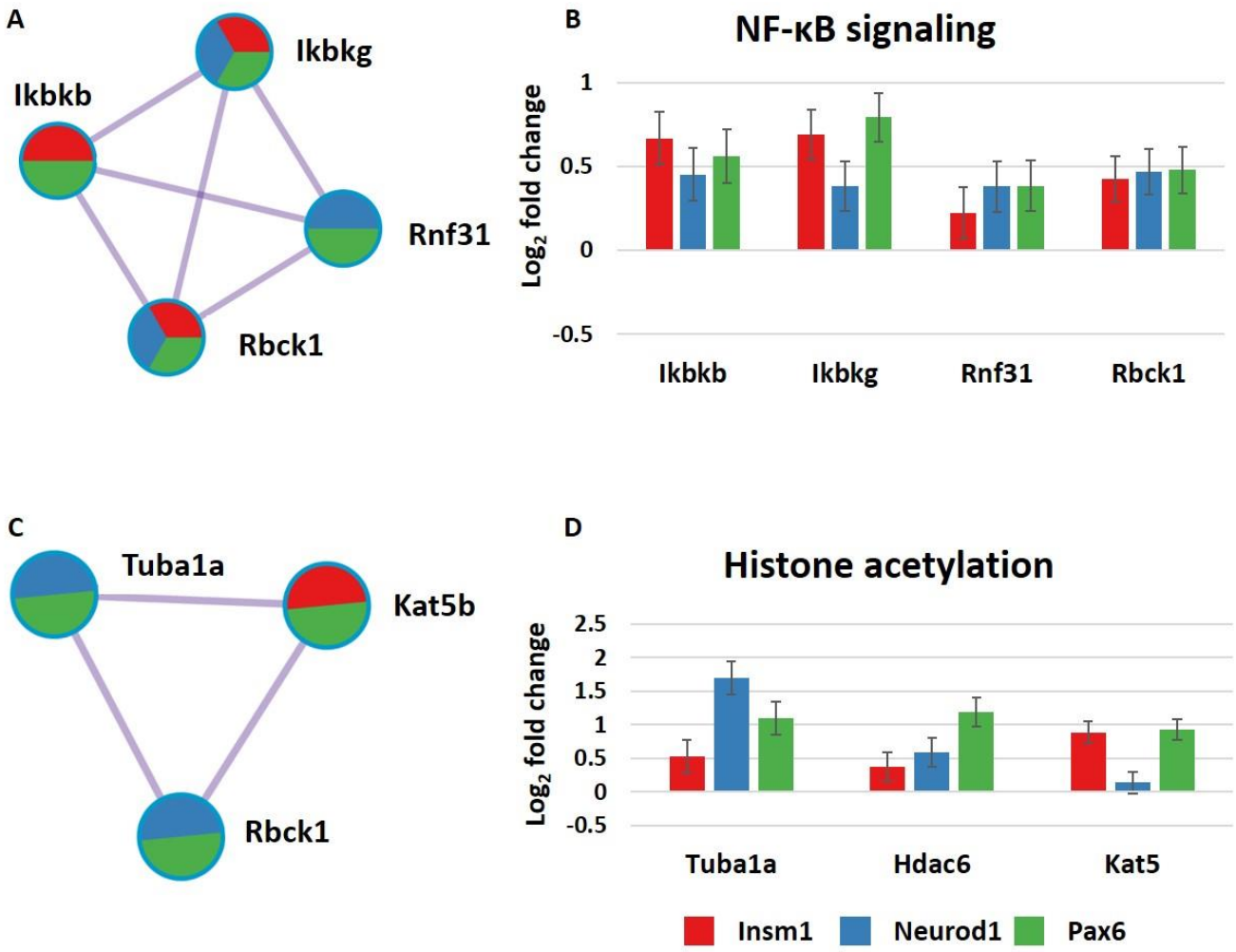


Figure 2.14. Metascape PPI MCODE analysis of URGs reveals enrichment for complexes involved in NF-κB signaling and histone acetylation. Subset networks of the most enriched (based on p-value) modules of genes that form PPIs specific for (**A**) NF-κB signaling and (**C**) histone acetylation complexes. (**B, D**) Bar graphs of the change in expression (as log_2 fold change) from our knock-out datasets of the genes included in each complex. Error bars indicate the std. error. Nodes are represented as pies in which the colored sectors correspond to their gene list. The color scheme is consistent with the previous figures.

Metascape analysis also identified protein-protein interaction (PPI) enrichments common to all three gene sets. Enriched subnetworks of DRGs known to form PPIs included members of the ATF2- *(Atf2, Pou2f1, Nf1, Brca1*), cohesin- (*Smca1, Smc3, Atrx, Mecp2*), and histone acetylation-complexes (*Kat2b, Crebbp, Clock*). (**Figs. 2.13A, 2.13C and 2.13E**). The PPI subnetworks of URGs involved in NF-κB signaling (*Ikbkb, Ikbkg, Rnf31, Rbck1*) and histone modification (*Tuba1a, Hdac6, Kat5b)* (**Figs. 2.14A and 2.14C**). Interestingly, NF-kB signaling has been recently shown to regulate β-cell proliferation and apoptosis, and in turn, β-cell mass, during development (Sever et al., 2021). Together, these analyses indicate that while there are differences in the genes dysregulated by Insm1, Neurod1, and Pax6, there are many shared functionalities, both from a gene ontology and PPI subnetwork perspective.

Prediction of direct targets of Insm1 and Neurod1 critical for endocrine cell development. To predict direct targets of Insm1 and Neurod1 in developing endocrine cells, I utilized previously reported ChIP-Seq datasets (Jia et al., 2015). Consistent with published data, our analysis showed that Insm1 and Neurod1 share a large proportion of binding sites (11,409 sites across the genome, 81.9% of total binding sites) (**Fig. 2.15A**). Integration of ChIP-Seq data with our RNA-Seq data on dysregulated genes suggests that 4,180 and 3,789 genes, respectively, may be direct targets of Insm1 and Neurod1 in endocrine progenitor cells (**Figs. 2.15B-C**). I further compared the lists of direct target genes and found 1,719 (27.5% shared) genes bound and dysregulated by both Insm1 and Neurod1. These included genes involved in peptide hormone response and secretion (*G6pc2, Glp1r, Sytl4, Stxbp5l*), zinc finger proteins (*Zfp326, Zbtb33, Zbdf2*), rhythmic processes (*Clock, Dhx9, Kcnd2*), and mRNA processing (*Celf4, Tra2a, Akap17b, Zfp871, Luc7l2, Ddx17*) (**Figs. 2.15D and 2.16A-D**). These findings indicate that of the genes dysregulated in our Insm1 and Neurod1 KO endocrine progenitor cells, 89% and 82% are predicted to be direct targets, respectively, based on the presence of binding sites within 5 kb of the transcriptional start site.

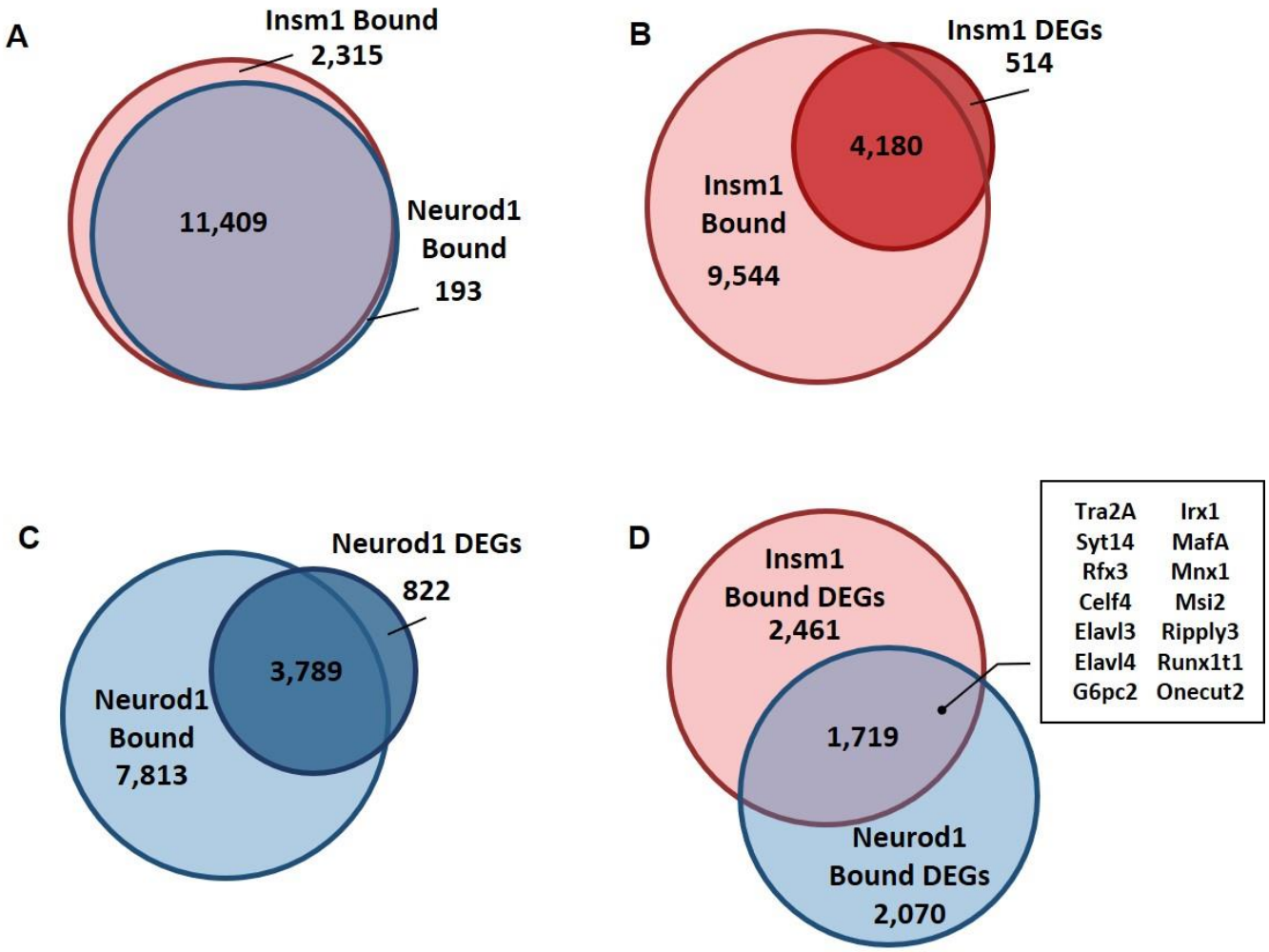


Figure 2.15. Synergistic gene regulation by Insm1 and Neurod1. Venn diagrams representing the overlap between (**A**) *Insm1* KO DEGs (*p*-value < 0.05 cutoff) and Insm1 bound genes, (**B**) *Neurod1* KO DEGs (*p*-value < 0.05 cutoff) and Neurod1 bound genes, and (**C**) genes bound by Insm1 and/or Neurod1. (**D**) Venn diagram showing the overlap of genes that are both bound and dysregulated by Insm1 and Neurod1, respectively. A subset of common direct target genes is highlighted within a boxed insert.

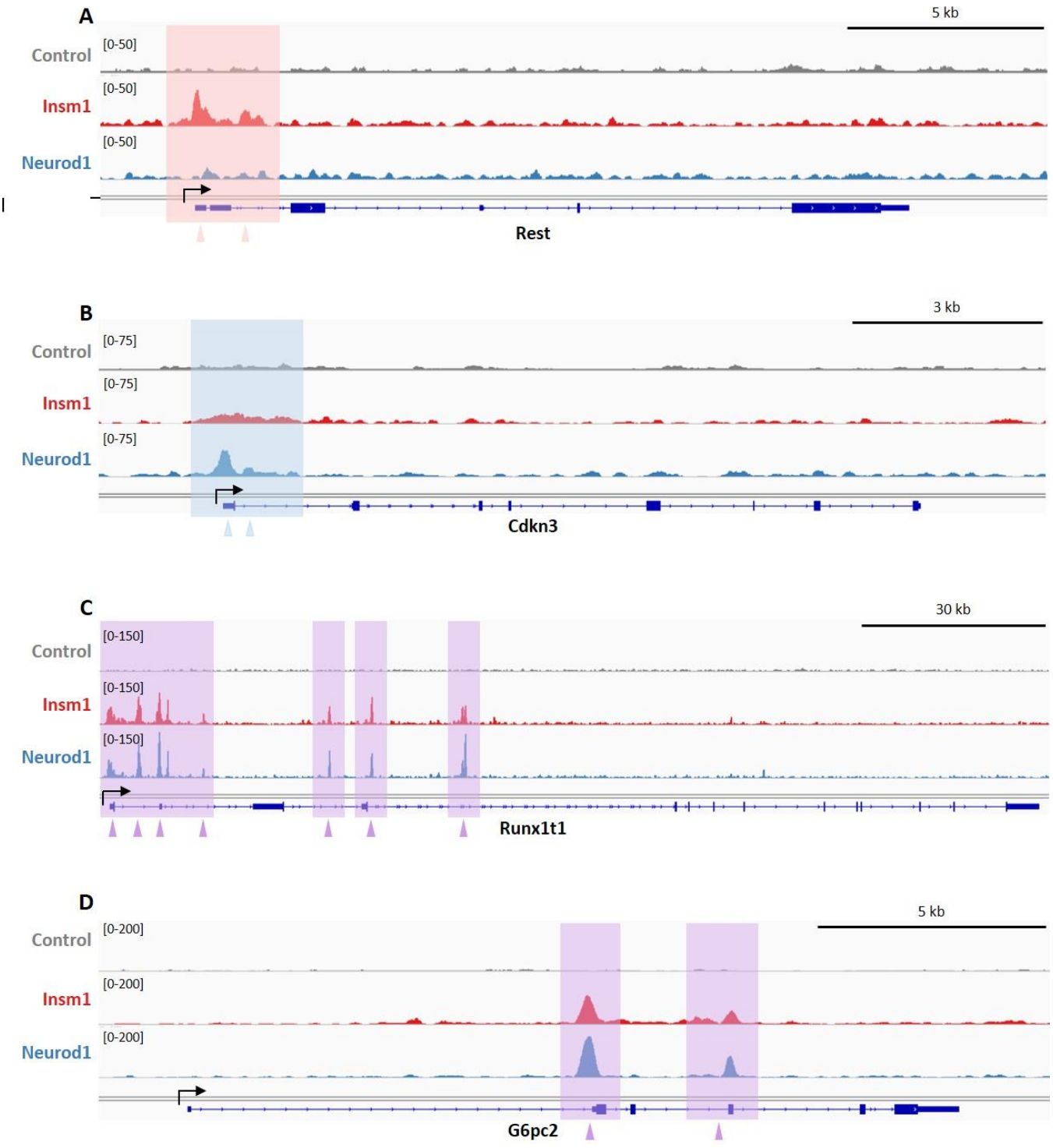


Figure 2.16. Example ChIP-Seq binding profiles of Insm1 and Neurod1 at dysregulated gene loci. Depicted are the ChIP-Seq coverage profiles for Insm1 and Neurod1 at (**A**) *Rest*, (**B**) *Cdkn3*, (**C**) *Runx1t1*, and (**D**) *G6pc2*. Strong TF binding by both Insm1 and Neurod1 can be detected around the transcriptional start sites and some internal regions of each of the genes. Insm1 binding profiles are in red, Neurod1 binding in blue. Regions with major peaks for both TFs are highlighted in purple and indicated by an arrow marker.

DISCUSSION

Insm1, Neurod1, and Pax6 both independently and coordinately govern pre-endocrine cell gene expression. *Insm1, Neurod1,* and *Pax6* are individually essential for the formation of functional pancreatic endocrine cells (Gu et al., 2010a; Jia et al., 2015; Mitchell et al., 2017; Naya et al., 1997; Osipovich et al., 2014; St-Onge et al., 1997). By systematically comparing the effects of deleting *Insm1, Neurod1,* and *Pax6* on the morphology, and gene expression of nascent pancreatic endocrine cells, we have obtained new insights into the complex and particularly important roles that each of these factors has in the development of pancreatic endocrine cells.

Role of Insm1. Consistent with previously published results, I showed that Insm1 KOs have defects in endocrine cell differentiation that were manifested in decreased numbers of Ins- and Gcg-positive and increased number of Ppy-positive endocrine cells. This was accompanied by decreased endocrine cell proliferation, as well as changes in the expression of genes that affect these processes (Gierl et al., 2006; Osipovich et al., 2014). In these studies, all analyzed mice carried a heterozygous *Insm1*^{GFPCre} allele which was also used as a control for all comparisons.

Our comparative transcriptional analysis revealed both similarities and differences in the genes regulated by Insm1 and Neurod1 and Pax6. For example, Insm1, but not Neurod1 or Pax6, activates the expression of *Fev* and *Mnx1* in endocrine cells, TFs that are known to influence the numbers and types of endocrine cells, as well as islet morphology. Moreover, Insm1 is specifically important for the expression of many genes involved in the inducible secretory function of endocrine cells, including protein secretion and Ca^{2+} regulated vesicle exocytosis. Our new data, together with a re-analysis of previously reported ChIP-Seq data, suggests that 89% of genes dysregulated in endocrine progenitor cells of Insm1 KO mice contain binding sites for Insm1 with 5 kb of the transcriptional start site.

Role of Neurod1. Like *Insm1* knockouts, mice that lack *Neurod1* exhibit an increased number of Ppy-expressing endocrine cells, but the proportion of other hormone–expressing *Insm1*^{GFP}-positive cells did not change significantly. While proportions of most of the differentiated endocrine cell types did not change, I saw a drastic decrease in proliferation and a relatively strong increase in apoptosis of endocrine cells that resulted in a marked

reduction in the overall number of endocrine cells. These findings are consistent with previous studies, supporting the hypothesis that a lack of proliferative cells and an increase in apoptosis are primary causes for the reduced total number of endocrine cells observed (Naya et al., 1997; Romer et al., 2019). Likewise, Romer et al., 2019 recently reported a decrease in α- and β- cell numbers only after the time ~P0 stage. Their quantifications at E17 show no differences in the cell numbers or ratios of Ins- or Gcg-expressing cells, indicating that the specific reduction in α- and β-cells occurs during late embryonic to early postnatal stages after islet differentiation has roughly concluded, which is consistent with our findings (Romer et al., 2019).

Our transcriptional profiling results also suggest that Neurod1 promotes proliferation and maturation through the stimulation of genes important for cell cycle regulation and chromatin modifications, processes that are important for cell differentiation (Chen and Dent, 2014; Dowen et al., 2014; Krentz et al., 2017; Pauklin and Vallier, 2013). Simultaneously, the observed upregulation of apoptotic and autophagy genes likely contributes to an increase in cell death, further decreasing the total number of endocrine cells. Additionally, ChIP-Seq data suggest that Neurod1 directly targets 82% of the genes dysregulated in this analysis, highlighting that it is likely the primary effector of those changes. Together these findings indicate the critical role of Neurod1 in propagating enough fully functional islet cells during development via the promotion of cellular proliferation and activation of key TFs.

Role of Pax6. A role for Pax6 in driving the specification of α-cells is well established, so the marked reduction of Gcg-positive cells (3.3% vs. 15% in controls) in the *Pax6* KO animals is not a surprise (Gosmain et al., 2010). Our data indicate that endocrine cells lacking *Pax6* not only have a decrease in the expression of *Gcg*, but also dysregulation in other genes important for α-cell development and function, such as *Pou3f4* (or *Brn4*), *Pcsk2*, *Slc30a8*, *Irx2*, and *Smarca1*. Besides its function in α-cells, Pax6 is important for the development and function of adult β-cells, which is supported by our findings of decreased expression of *Ins1*, *Ins2*, and *Insr* in Pax6 KOs (Gosmain et al., 2010; Mitchell et al., 2017; Swisa et al., 2017). While the morphological phenotypes of the islets I observed are in line with previously published results, I am aware that I cannot exclude the possibility of a combinatorial effect caused by potentially suboptimal levels of *Insm1* in our GFP reporter model (Gosmain et al., 2010; Sander et al., 1997; St-Onge et al., 1997).

Additionally, it was previously shown that Pax6 is alternatively spliced and that α- and β-cells preferably express alternate isoforms (Singer et al., 2019). I found that both Insm1, and Pax6 itself, regulate *Pax6* mRNAs

splicing, both potentially contributing to its cell-type-specific diversity. The downregulation of genes involved in chromatin organization, cell cycle regulation, and cell growth in *Pax6* KOs, similarly to *Neurod1* KOs, likely contribute to the reduction in endocrine cell expansion, while an increase in apoptotic genes correlates to the increase in cell death.

Regulatory overlap of Insm1, Neurod1, and Pax6. Recently we have shown that *Insm1* and *Neurod1* exhibit a high degree of co-expression during endocrine differentiation and that *Pax6* is similar except that it continues to be expressed in adult α-cells (Osipovich et al., 2021). These co-expression relationships are consistent with *Insm1, Neurod1,* and *Pax6* being important components of the gene regulatory network in nascent pancreatic endocrine cells. The genes regulated by these three factors also share many functional assignments, such as mRNA processing and alternative splicing, and cell cycle regulation. These similarities likely underpin the broadly similar morphological phenotypes when each is knocked out, which includes a reduced overall number of pro-endocrine cells, reduced endocrine cell proliferation, and increased apoptosis, as well as defects in differentiation towards different kinds of hormone-expressing cells.

Our findings also suggest epistasis between *Insm1* and *Pax6*. Specifically, I had difficulty obtaining $Insm1^{GFP.Cre/+}$; $Pax6^{fl/fl}$ animals at E18.5. This suggests an *Insm1* gene dosage effect that becomes more pronounced in the absence of *Pax6*. A recent report showed that mice with only a single functional *Insm1* allele ($Insm1^{+/-}$) have impaired β-cell proliferation that results in impaired glucose tolerance in adult animals (Tao et al., 2018). However, the authors observed no impairments in β-cell mass at birth and suggested that *Insm1* haplo-insufficiency has little to no effect on embryonic endocrine development (Tao et al., 2018). Previous *Pax6* KO studies have also not reported any unexpected difficulties in obtaining *Pax6* null embryos, and animals died just after birth (Ashery-Padan et al., 2004; Heller et al., 2004; Sander et al., 1997; St-Onge et al., 1997). Therefore, the reduced number of *Pax6* KO embryos I observed at E18.5 may be due to a detrimental synergistic effect, or negative epistasis, between *Insm1* and *Pax6* that impairs embryo survival.

In addition to the early lethality of *Pax6* KO embryos, our data show an increase in both apoptotic events at E18.5 and an increase in the expression of pro-apoptotic genes at E15.5. Previous studies of global *Pax6* KO mice have all observed a marked decrease in hormone-positive endocrine cells, but none quantified apoptosis events (Ashery-Padan et al., 2004; Heller et al., 2004; Sander et al., 1997; St-Onge et al., 1997). Moreover, Ashery-

Padan et al., who utilized a Cre-Lox-based lineage tracing strategy to quantify the proportion of *Pax6* expressing cells relative to total pancreatic tissue at E18.5, reported the loss of hormone expression but no difference in total cell number (Ashery-Padan et al., 2004). While apoptosis was not specifically quantified, their study suggests that pro-endocrine cells form and expand but then fail only to express islet hormones (Ashery-Padan et al., 2004). If so, then our observations suggest that *Insm1* and *Pax6* may have redundant functions, perhaps in the suppression of pro-apoptotic genes.

In conclusion, the present study provides additional insights into the roles of *Insm1*, *Neurod1*, and *Pax6* in driving endocrine development and differentiation. While each of these TFs has its own unique regulatory functions, there is also functional overlap between these factors based on the large number of commonly dysregulated genes represented in our data. Finally, our data demonstrate the broad impact of eliminating each TF, further illustrating their individual roles and importance in establishing the GRN within pancreatic endocrine cells.

CHAPTER III - NOVEL ROLES OF INSM1, NEUROD1, AND PAX6 IN REGULATION OF ALTERNATIVE SPLICING DURING ENDOCRINE CELL DEVELOPMENT

INTRODUCTION

Mechanisms and functions of alternative splicing.

The tight regulation of gene and protein expression necessary for establishing or maintaining gene regulatory networks goes beyond transcriptional activation and repression. While there are actually many different layers of regulation, the process of alternative splicing of mRNA is of note. Following transcription, nascent pre-mRNA must undergo splicing, which is the process of removing introns and the subsequent joining of exons to create a mature mRNA molecule (Dlamini et al., 2017). Alternative splicing (AS) occurs when varying combinations of exons, or introns, are included and excluded then spliced together to form multiple isoforms from the same gene transcript (**Fig. 3.1**) (Feero et al., 2010). AS leads to a significant enhancement of transcriptome diversity and the accompanied expansion of the protein repertoire, providing another mechanism for regulating gene expression (Dlamini et al., 2017). To illustrate this point, there are a predicted ~22,000 protein-coding genes in the human genome, but there are over 215,000 observed isoforms (Lee and Rio, 2015). This vast difference between gene and gene isoform numbers is due to the fact that as many as ~95% of genes are alternatively spliced. In comparison, the process of AS is evolutionarily old (Irimia et al., 2007; Kelemen et al., 2013), the relative frequency of AS events increase with organismal complexity (Kim et al., 2007a; Merkin et al., 2012). Species such as mice, flies, or worms have a similar number of genes relative to humans, but they all have a lower percentage of alternatively spliced genes relative to humans (~65%, 45%, and 25%, respectively) (Lee and Rio, 2015). Additionally, comparative analysis of gene expression and AS profiles in different organs across different vertebrates revealed that overall organ-specific AS patterns evolved much more rapidly than organ-specific gene expression patterns (Barbosa-Morais et al., 2012). Having played such a crucial part in evolution, AS must have significant roles in organ development and/or function.

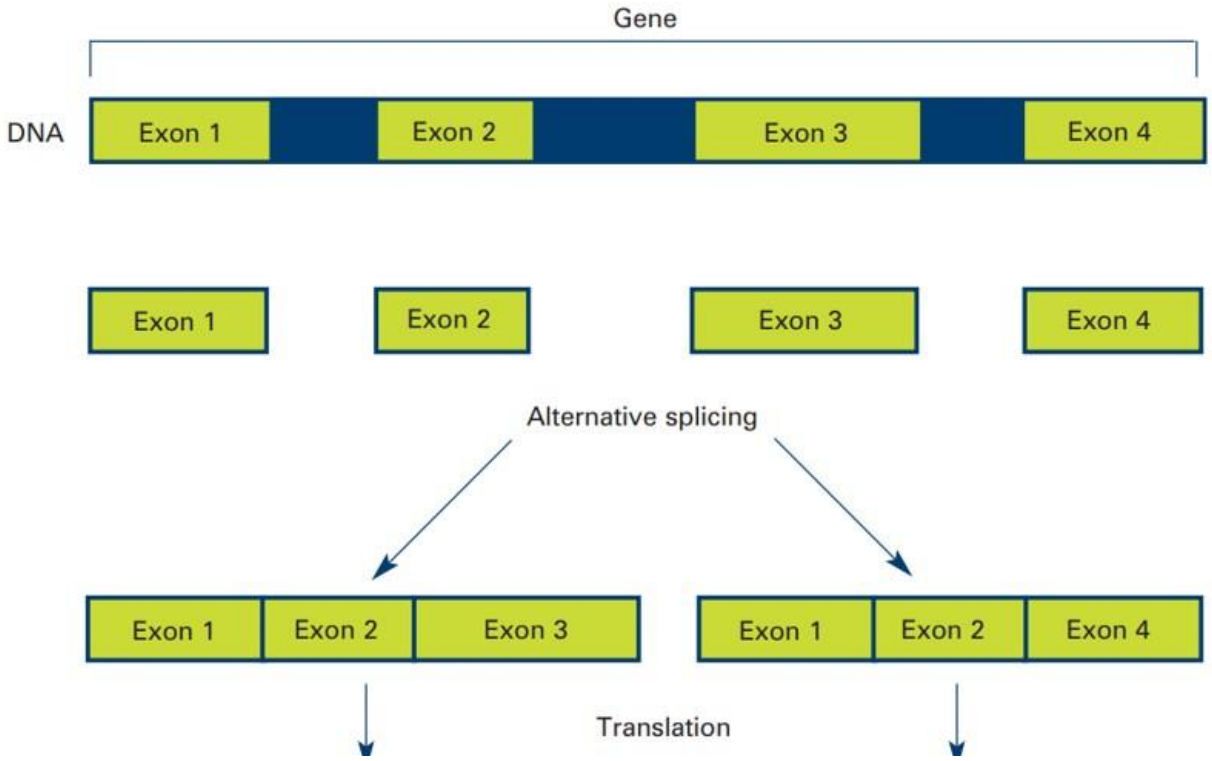


Figure 3.1. Pre-mRNA splicing products. Splicing of pre-mRNA generally involves the excision of introns, followed by the ligation of exons. Differential inclusion or exclusion of exons and introns from a single gene product results in multiple mRNA isoforms, and upon translation, multiple protein products. This figure shows the alternative inclusion and exclusion of exons 3 and 4. The first mRNA product is marked by the ligation of ex. 1, ex. 2, and ex. 3. The second mRNA product skips exon 3, and includes instead ex. 1, ex. 2, and ex. 4. For each, a unique protein structure is formed, labeled here as Protein A and Protein B. Figure modified and reprinted with permission from The National Institute of Health (NIH), NIH Publication No. 10-662, 2010.

In general, mRNA splicing is primarily mediated by the spliceosome. The spliceosome is a machinery complex made up of hundreds of individual components, including the NineTeen Complex (NTC), small nuclear ribonucleoprotein complexes (snRNPs) (U1, U2, U4, U5, U6), associated small nuclear RNAs (snRNA), along with other factors (Dvinge, 2018). The exon-introns boundaries of splice sites are marked by short sequence motifs at the upstream exon (5' splice site), the intronic branch point, and the downstream exon (3' splice site) (Dvinge, 2018). The spliceosome binds and assembles at these defined sites, and through a series of highly dynamic interactions, it forges the intron into a lariat-type loop and then catalytically cleaves the intron at the splice sites allowing the exons to be joined together (Gehring and Roignant, 2021).

In addition to acting as recognition sequences for the spliceosome, *cis*-acting sequences within the pre-mRNA are also important for the recruitment of splicing enhancers or repressors (Dvinge, 2018). The interaction of these sequences with RNA-binding protein (RBP) splicing factors constitutes a mechanism for the regulation of alternative splicing patterns (Lee and Rio, 2015). There are hundreds of RBPs that regulate binding sites and spliceosome activity which fall into three categorical groups. Namely, RBPs can be of the heterogeneous nuclear ribonucleoprotein (hnRNP) type, serine and arginine-rich (SR) proteins, or tissue-specific RBPs (**Fig. 3.2**) (Lee and Rio, 2015; Licatalosi and Darnell, 2010). SR proteins have a C-terminal domain enriched with the Arginine and Serine amino acid sequences and an N-terminal RNA recognition domain (Jeong, 2017; Zhou and Fu, 2013). SR proteins are typically considered to function as splicing activators, promoting the splicing at binding sites (Anko, 2014). Conversely, hnRNP particles often function as repressors of alternative splicing (**Fig. 3.2**). The third class of proteins consists of tissue-specific RBPs. These include factors such as Nova1, Tra2A/B, PTB, Msi1/2, and members of the Rbfox family (Biamonti et al., 2019; Lee and Rio, 2015; Witten and Ule, 2011).

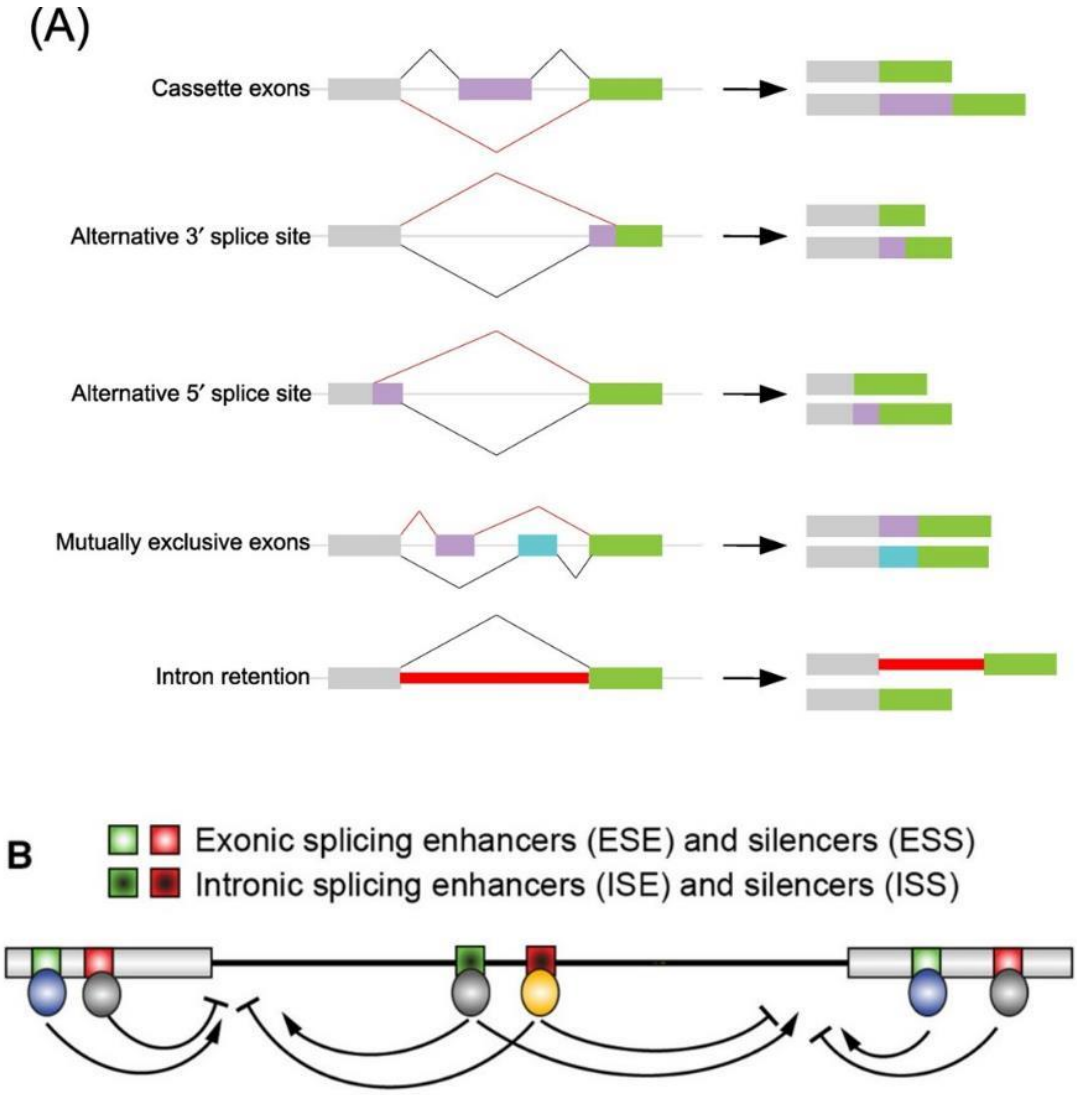


Figure 3.2. Alternative splicing events promotes transcriptome and proteome diversity through key regulatory mechanisms. There are five main types of alternative splicing events. The most common is exon cassettes that lead to an exon being excluded from mature mRNA. Inclusion of an intron in the mRNA transcript is defined as intron retention (red box) and subsequently undergoes nonsense-mediated decay. Alternative 5′ or 3′ splice site leads to inclusion of a part of an intron or exclusion of segment of an exon. Mutually exclusive exons are characterized by one exon that is spliced out while the other one is retained. The grey, purple, and green boxes represent different exons. The black and red lines represent different splicing events. Alternative splicing is regulated by *cis*-elements and RNA binding protein splicing factors. Location of cis-acting splicing enhancer or silencer motifs, and a simplified schematic of the trans-acting splicing regulators that bind them: exonic splicing enhancers (ESE), exonic splicing silencers (ESS), intronic splicing enhancers (ISE), and intronic splicing silencers (ISS). SR proteins and hnRNPs are illustrated as being splicing enhancers and repressors, although their regulatory role is highly complex and depends on multiple factors, such as the location of their binding site and expression of other splicing factors. They are likewise not restricted to binding motifs within either exons or introns specifically. This figure was modified and reprinted with permission from Dvinge, 2018; and Wang & Aifantis, 2020.

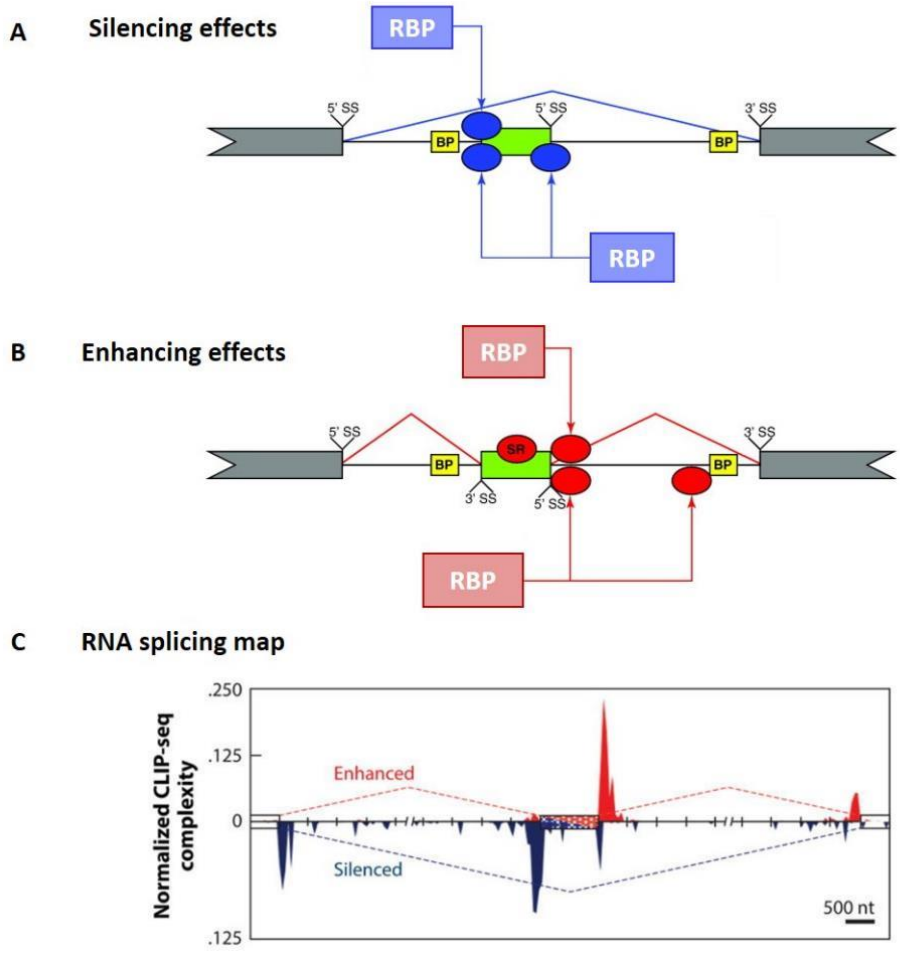


Figure 3.3. RNA binding proteins can function to both enhance and silence alternative splicing sites. (**A**) Positions in silenced exons with the enriched binding of different RBPs. Shown above the transcript diagram, RBP (blue ellipse) binding to a single site close to 3′ SS silences exon inclusion (blue line). Even though not shown in this diagram, such a site can also lie within the exon or close to the 5′ SS. Shown below the transcript diagram, some RBPs bind at intronic positions close to 3′ and 5′ splice sites of the exon to efficiently silence exon inclusion. The arrows indicate that binding at the different positions is achieved by the different RBPs or a hnRNP C. (**B**) Positions of enhanced exons with the enriched binding of different RBPs. Shown above the transcript diagram, RBP (red ellipse) binding downstream of the 5′ splice site of a cassette exon promotes its inclusion (red line). Shown below the transcript diagram, RBP binding at multiple positions at both ends of the downstream intron enhances exon inclusion. The arrows indicate that interaction between the RBPs bound at both sides of the intron might be required for the enhancing effect. In contrast to the RBPs studied so far, SR proteins enhance inclusion when bound within exons. Upstream and downstream exons (grey boxes) and potential branch points (yellow boxes) are also indicated for positional reference. (**C**) A Nova RNA splicing map for cassette exons generated by integrating the HITS-CLIP experimental identification of Nova-binding sites and splice-junction microarray data. This figure was modified and reprinted with permission from Witten and Ule, 2011, Lee and Rio, 2015, and Licatalosi, 2008.

Interestingly, RBPs are not the only factors that can promote activation or repression of splicing activity at a given pre-mRNA splice site. Other *cis*-acting sequences are known as splicing regulatory elements (SREs), which can be located within exons or introns, enhance or inhibit splicing through the recruitment of *trans*-acting RBP factors (Wang and Wang, 2014). These regulatory elements are known as exonic and intronic splicing silencers (ESS and ISS), and exonic and intronic splicing enhancers (ESE and ISE), respectively (Lee and Rio, 2015). As their names suggest, ESS and ISS motifs silence splicing activity, while ESE and ISE motifs promote splicing activity. Consistent with the general activating or repressive functions of the previously mentioned RBP types, SR proteins often bind ESE and ISE sites, and ESS and ISS motifs are generally bound by hnRNPs (**Fig. 3.2**) (Anko, 2014; Long and Caceres, 2009; Martinez-Contreras et al., 2007). While specific classes of RBPs tend to have enhancing or silencing functions, most can function in both manners depending on the motifs to which they are bound (**Fig. 3.3A-B**). These context-dependent interactions are often depicted as RNA splicing maps, in which the binding locations and resulting isoform are represented (**Fig. 3.3C**).

Alternative splicing in development.

It is well established that as cells transition from a pluripotent state into more differentiated cell types during development, key switches in the gene expression programs are necessary. The progressive shift in transcriptional patterns is driven by the activity of lineage-specific TFs, which narrow the cells' lineage potential and push them towards a defined cell type (Fiszbein and Kornblihtt, 2017). However, there is emerging evidence that, in addition to transcriptional regulation, mRNA processing is also important for developmental programming. Research has shown that splicing is critical for establishing gene regulatory networks, as switches in AS patterns closely correlate with cellular differentiation (Fiszbein and Kornblihtt, 2017).

Many examples outlining the importance of alternative splicing events can be found in the development of various tissue types, such as during neurogenesis, adipogenesis, and immune cell differentiation. One study demonstrated that during neuronal development, neural-specific Ser/Arg repeat-related protein of 100 kDa (nSR100/SRRM4) governs the alternative splicing of the RE1 Silencing Transcription Factor (Rest), which represses the expression of genes involved in neurogenesis (Norris and Calarco, 2012; Raj et al., 2011). They found that nSR100 promotes the splicing of the *Rest4* isoform, a shorter version of the gene that has impaired

repressive functions, allowing for the upregulation of necessary pro-neural genes in neural cells. These factors form a feedback loop, where in non-neuronal cell types, the full isoform of *Rest* represses expression of nSR100 and is able to fully repress neural-specific genes (Raj et al., 2011).

Until recently, the highly conserved RBP splicing factor Nova1 has been considered a brain-specific factor, playing an important part in neuronal development and function, and has even been shown to be critical in the evolution of brain development in humans from Neanderthals (Trujillo et al., 2021). However, Eizirik et al. demonstrated that it is also preferentially expressed in human and rodent islets (Eizirik et al., 2012). β-cell-specific knockdown of Nova1 resulted in alternative splicing of ~5k isoforms, of which only 154 were shared with transcripts found in the brain. Alternatively spliced genes included those involved in calcium signaling, exocytosis, insulin signaling, apoptosis, and splicing and transcription. Importantly, Nova1 knockdown resulted in differential splicing of *Insr*, *Foxo1*, *Foxo3a*, *Pax6*, *Snap25*, and *Sf1* (**Fig. 3.4A**) (Eizirik et al., 2012). Furthermore, Singer *et al.* recently reported that the alternative splicing of *Pax6*, which requires recruitment of RBPs by the lncRNA *Paupar*, alters both its transcriptional activity and DNA binding specificity (Singer et al., 2019). *Paupar* knockout experiments indicated that *Paupar* promotes splicing of the alpha-cell specific isoform of *Pax6*, modulating its transcriptional activity. The α-cell-specific *Pax6* isoform is necessary for the proper expression of α-cell-specific genes, thus influencing its developmental program and function (**Fig. 3.4B**) (Singer et al., 2019).

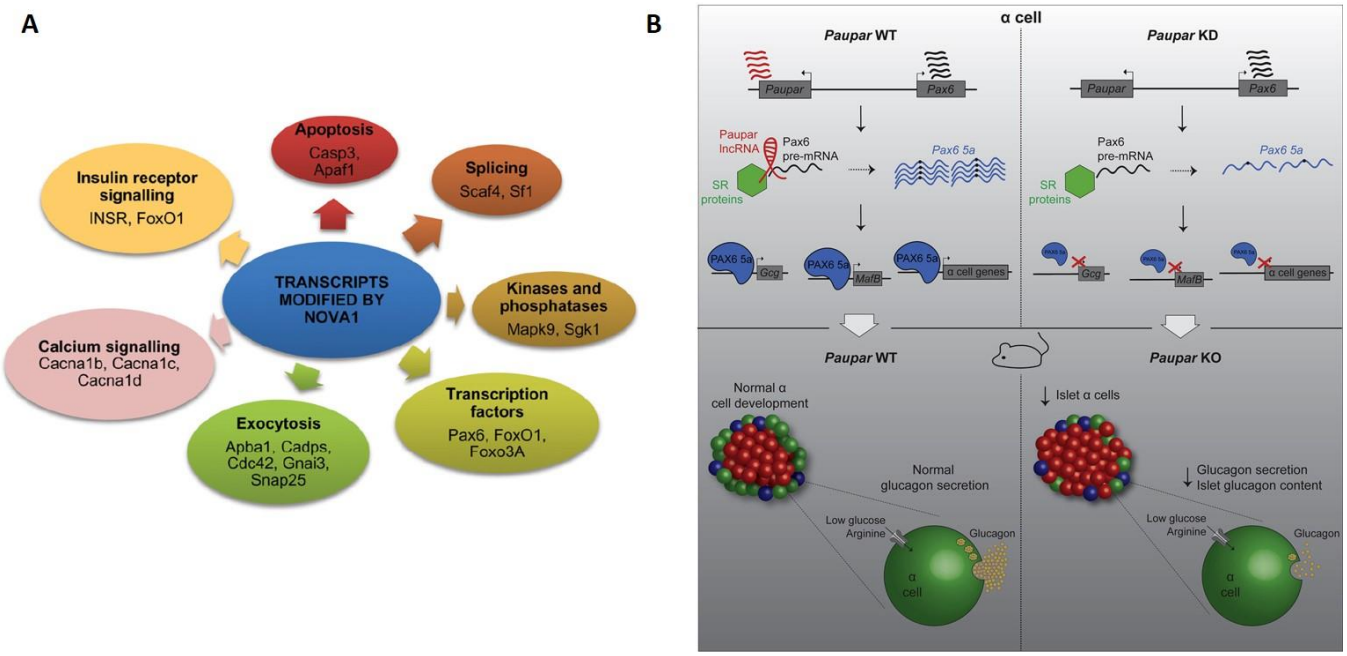


Figure 3.4. Alternative splicing of Pax6 disrupts its transcriptional activity in pancreatic alpha-cells. (**A**) Pauper regulates AS of Pax6. Mis-splicing of Pax6 disrupts is ability to activate expression of alpha-cell specific genes. (**B**) RNA binding protein Nova1 has been found to modulate alternative splicing of genes important for beta-cell function. This Fig. was modified and reprinted with permission from Singer et al., 2019 and Villate, et al., 2014.

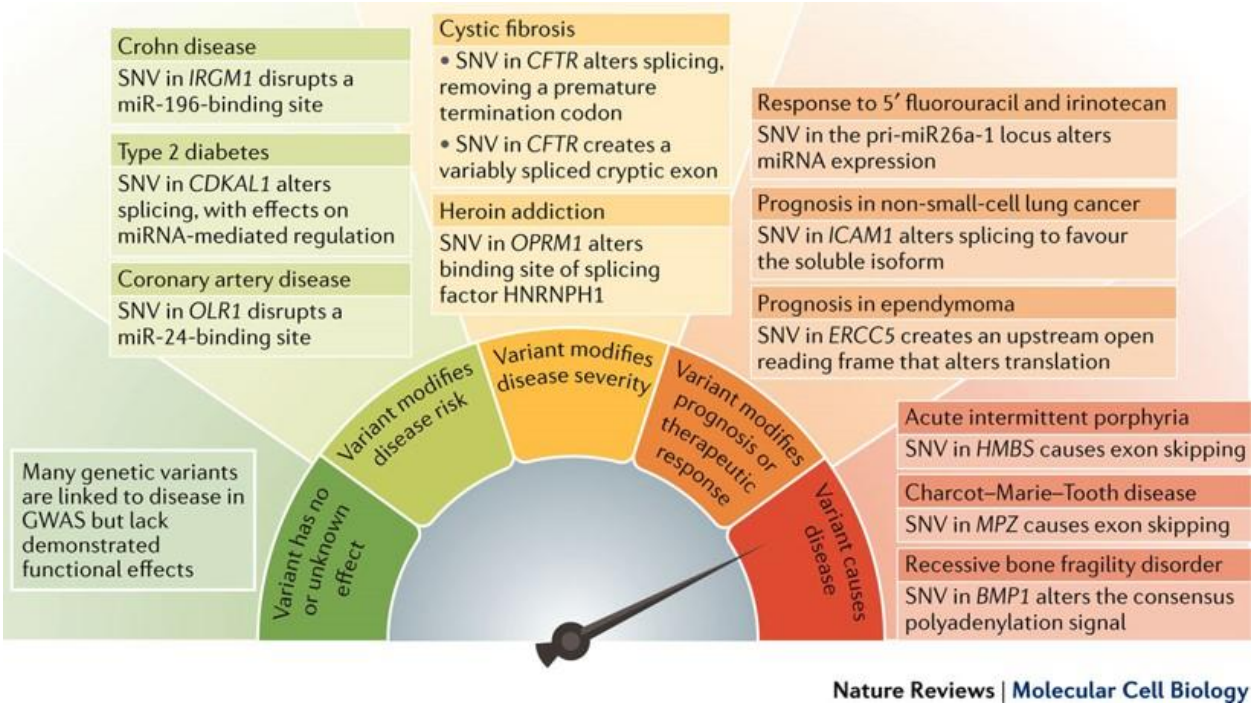


Figure 3.5. Genetic variants that affect RNA processing can have a range of effects on human health. These effects range from benign or unclassified variants to pathological variants that cause monogenic diseases. Selected examples of functional genetic variants that affect RNA processing are depicted, which have a range of consequences, such as causing disease, altering prognosis or therapeutic response, modifying disease severity or modifying disease risk. BMP1, bone morphogenetic protein 1; CDKAL1, CDK5 regulatory subunit-associated protein 1-like 1; CFTR, cystic fibrosis transmembrane conductance regulator; ERCC5, ERCC excision repair 5, endonuclease; GWAS, genome-wide association studies; HMBS, hydroxymethylbilane synthase; HNRNPH1, heterogeneous nuclear ribonucleoprotein H1; ICAM1, intercellular adhesion molecule 1; IRGM1, immunity-related GTPase M member 1; miRNA, microRNA; MPZ, myelin protein zero; OLR1, oxidized low-density lipoprotein receptor 1; OPRM1, opioid receptor mu 1; pri, primary; SNV, single-nucleotide variant. Figure modified and reprinted with permission from Manning and Cooper, 2016.

Alternative splicing in disease.

Dysregulation of AS events has also been linked to various diseases and can have varying degrees of consequences (**Fig. 3.5**) (Manning and Cooper, 2017; Pihlajamaki et al., 2011). Disruptions can be caused by both *cis*- and *trans*-mutations, such as those at splice sites and recognition sequences, as well as in RNA binding proteins or the members of the spliceosome themselves (Lee and Rio, 2015). A growing body of evidence has linked the development of multiple cancer types with disturbances in alternative splicing. The perturbed function of splicing factors like *Srfs1*, *Esrp*, *Rbfox2,* and *Rbm47*, as well as differential splicing of targets such as *Numb*, *Ron*, *Tcf4*, and *mTORC1/2,* can affect cell plasticity and promote tumorigenesis (Biamonti et al., 2019). Many reviews have nicely outlined the role of AS in cancer (Biamonti et al., 2019; Bonnal et al., 2020; Cieply and Carstens, 2015; Pradella et al., 2017).

Importantly, mutations leading to aberrant splicing of genes important for islet cell development and function have been linked to maturity-onset diabetes of the young (MODY) (Dlamini et al., 2017). *Hnf1a* (*hepatic nuclear factor 1 homeobox*) is another MODY causal gene, and multiple splicing mutations (including *IVS4nt-1G>T*) in this gene have led to instability, premature degradation, and altered function, including a decrease in the amount of insulin made in beta-cells (Capelli et al., 2009; Harries et al., 2006; Harries et al., 2008). Tissue-specific splicing patterns of *glucokinase* (*Gck*) that were associated with altered functions were previously discovered (Liang et al., 1991). In islet cells, Gck facilitates the formation of glucose-6-phosphate (G6P) through phosphorylation of glucose and regulates insulin secretion (Dlamini et al., 2017; Liang et al., 1991). Mutations in *Gck* have also been identified in individuals with MODY (Costantini et al., 2011; Lorini et al., 2009).

Still, our understanding of the role of AS, particularly within the pancreas, remains incompletely understood, and much more research is needed. Thus, gaining a better understanding of RBPs and alternative splicing events in the developing and mature pancreas is of great importance. Here, I identify a novel role for *Insm1*, *Neurod1*, and *Pax6* in regulating alternative splicing events. Elimination of Insm1, Neurod1, and Pax6 impaired expression of many RNA binding proteins, thereby altering RNA splicing events, including for Syt14 and Snap25, two genes required for insulin section. All three factors are necessary for normal splicing of Syt14, and both Insm1 and Pax6 are necessary for the processing of Snap25. Collectively, these data provide a deeper understanding of how Insm1, Neurod1, and Pax6 are essential for the formation of functional endocrine cells.

MATERIALS AND METHODS

Differential splice variant analysis. DSV analysis was performed using the previously described RNA-Seq datasets from control and KO embryos at E15.5. Snakemake (5.2.4), a workflow management system, was used to call on five programs necessary to process the samples. FastQC (0.11.8) returned a quality score report on each sample's sequences before and after Trimgalore (0.5.0) trimmed the sequences. Trimmed paired-end sequences were passed to STAR (2.6.0), an RNA seq aligner, to align against the mouse genome (gencode 17; GRCm38), and gene counts for each sample were returned. The average uniquely mapped reads count per sample was 52M (89%). QoRTs (Quality of RNA-Seq toolset) converts the gene counts format for analysis in JunctionSeq. Lastly, Multiqc (1.7) compiled the final summary output from the other jobs into a single report. Gene counts for each sample were passed to R (3.5.3) to do a quality control assessment before analysis. Quality control assessment included analyzing a box plot of normalized sample counts, PCA plot, sample to sample heat map clustering, and a density plot of the normalized counts. All samples passed the quality control assessment. Each sample's QoRTS formatted gene counts were run through JunctionSeq (1.12.1) to determine any differential gene expression through alternative splicing between the knock-out and control groups.

Sashimi plots and exon junction visualization. Plots of read coverage across exons and corresponding sashimi plots were generated using the Integrative Genomics Viewer (IGV) platform (Robinson et al., 2011). Indexed BAM files from each KO dataset were uploaded directly to IGV for analysis, and mm10 was used as the reference genome. All tracks are set to the same read count scale for a given locus.

Identification of putative RNA binding protein binding sites. Putative binding sites of RBPs at alternatively spliced genes were identified using the online oRNAment database (http://rnabiology.ircm.qc.ca/oRNAment) (Benoit Bouvrette et al., 2020). Specific gene loci were queried for ELAVL4, NOVA1, and TRA2A binding sites using the search tool. Putative binding sites for each locus were visualized using the Integrative Genomics Viewer (IGV) platform (Robinson et al., 2011).

Insm1, Neurod1, and Pax6 KO endocrine cells exhibit dysregulated RNA splicing.

The regulation of RNA metabolic processes, which includes the regulation of alternative splicing, was among the top five enriched GO terms and signaling pathways of dysregulated genes found in all three of the KO gene sets. To determine whether mRNA splicing was affected in the *Insm1*, *Neurod1*, and *Pax6* KO cells, I first examined the expression of several RNA binding proteins (RBPs) known to regulate mRNA splicing. I found several RBPs with similarly dysregulated expression patterns, including *Elavl3*, *Elavl4*, *Tra2A,* and *Tra2B* (**Fig. 3.6**). The expression of other differential RNA splicing factors such as serine/arginine-rich (SR) proteins that generally act as activators of splicing (*Srsf1*, *Srsf3*), heterogeneous nuclear ribonucleoproteins (hnRNPs) with repressive functions (*Hnrnpa0*, *Hnrnph1*, *Hnrnpha3*), and members of the spliceosome itself (*Sfb1, Sfb2, Sfb3, Sfb4*) are also dysregulated. Importantly, ChIP-Seq data indicate that Insm1 and Neurod1 bind at the promoter regions of many of the same RBPs whose expression levels are dysregulated in our KO data. For example, *Elavl3/4, Nova1, Tra2a/b*, and *Srsf1*, in addition to being downregulated, are all bound by both Insm1 and Neurod1 at their promoter regions (**Fig. 3.7A-D**).

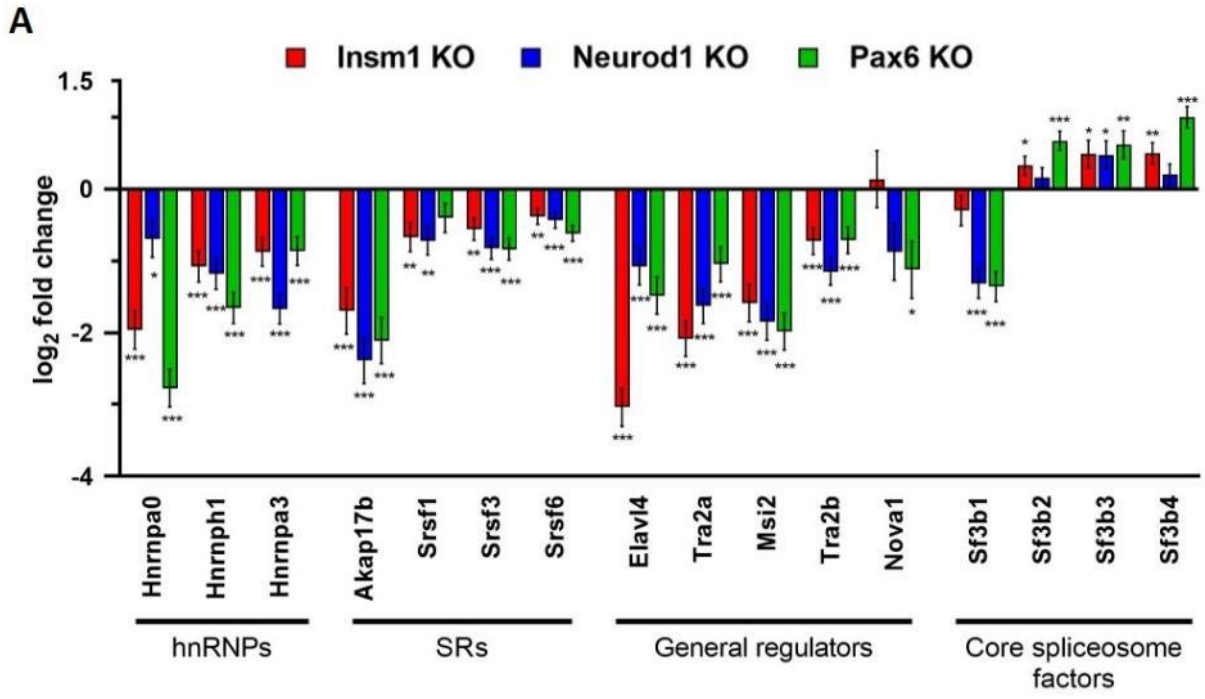


Figure 3.6. Insm1, Neurod1 and Pax6 regulate alternative splicing in pancreatic endocrine cells. (**A**) Many important RNA binding proteins (RBPs) are dysregulated in in *Insm1*, *Neurod1* and *Pax6* KOs. Bar graph of the $\log_2$ fold change of a subset of dysregulated RBPs. Asterisks indicate *p*-values of *<0.05, **<0.01, ***<0.001. (**B**) Venn diagram demonstrates overlaps between differentially spliced genes in *Insm1*, *Neurod1,* and *Pax6* KO RNA-Seq datasets. Examples of alternatively spliced genes are highlighted with a box insert. Differential splicing events between the knock-out and control groups were identified by performing JunctionSeq analysis on the RNA-Seq datasets. A cutoff of *p*-adj. value ≤ 0.01 and $\log_2$ fold change ≥ |1| was used.

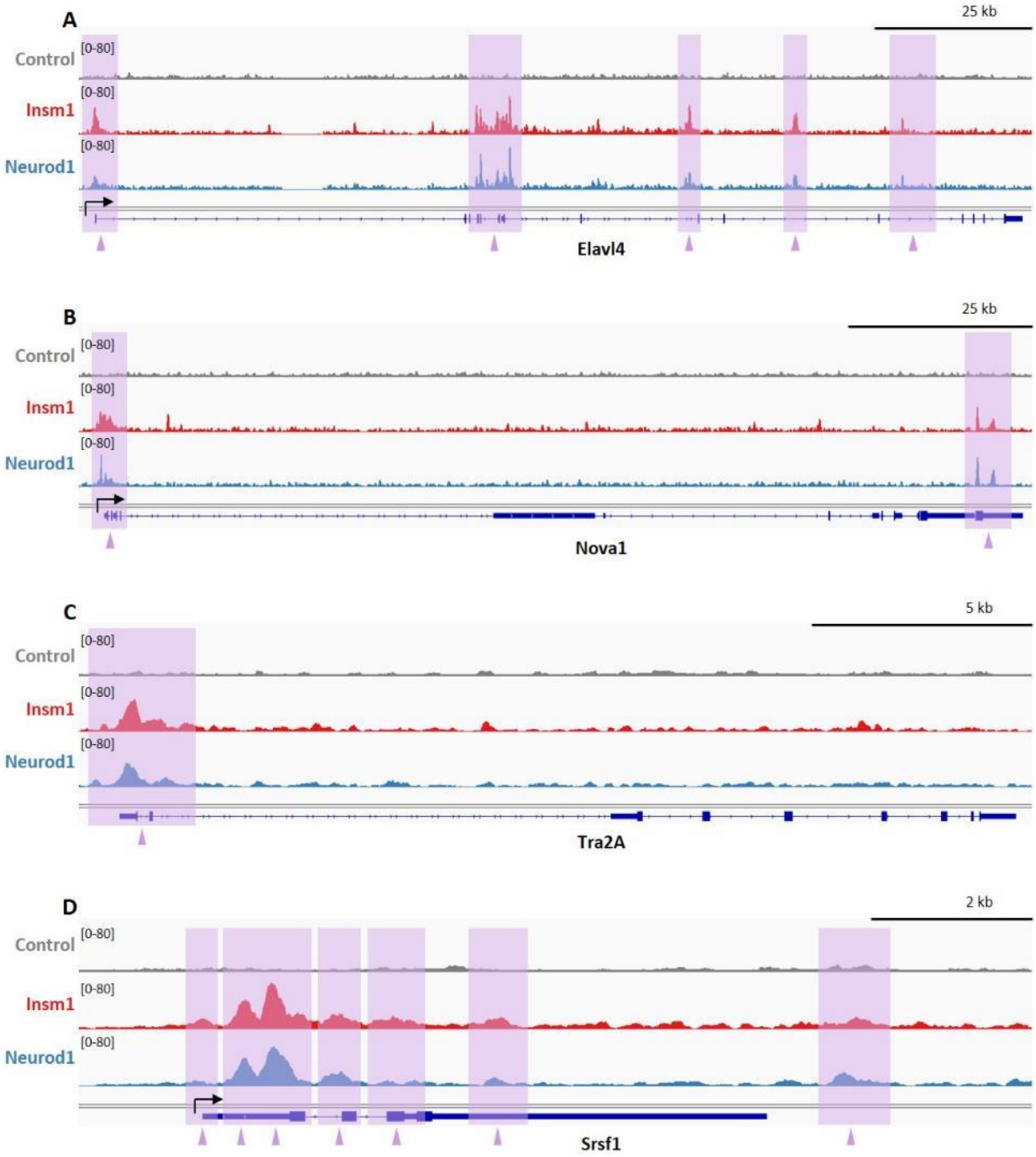


Figure 3.7. Direct binding of dysregulated RNA binding proteins by Insm1 and Neurod1. Depicted are the ChIP-Seq coverage profiles for Insm1 and Neurod1 at (**A**) *Elavl4*, (**B**) *Nova1*, (**C**) *Tra2A*, and (**D**) *Srsf1*. Strong TF binding by both Insm1 and Neurod1 can be detected around the transcriptional start sites and some internal regions of each of the genes. Insm1 binding profiles are in red, Neurod1 binding in blue. Regions with major peaks for both TFs are highlighted in purple and indicated by an arrow marker.

Differential splicing analysis reveals unique alternative splicing events in knockout models. To determine if the dysregulation of RBP expression in the absence of *Insm1*, *Neurod1*, or *Pax6* was correlated with differential splicing events, I used our RNA-Seq datasets to perform a computational alternative splicing analysis using JunctionSeq. The results indicated hundreds of differential splicing events in *Insm1* KO (356), *Neurod1* KO (212), and *Pax6* KO (885) datasets (adj. *p*-value ≤ 0.01, $\log_2$ FC > |1|) (**Fig. 3.8A-B**). To determine the commonalities and differences in the sets of differentially spliced genes from each KO, I identified those genes that were AS in more than one KO dataset. Of the 1,262 total differentially spliced genes, 172 (13.6%) were shared between pairs of TFs, with *Pax6* itself being alternatively spliced in *Insm1* and *Pax6* KO datasets (**Fig. 3.8A**). However, only 19 (1.5%) differential splicing events were common to all three KOs. This subset included genes such as *Ncoa4*, *Syt14*, *Tnfrsf12a*, *Mov10, Eif2s2, Robo2,* and *Ccdc6* (**Fig. 3.8A**).

Genes uniquely AS in *Insm1* KO animals included the TF and ubiquitin-protein *Rbck1*, the RBP *Tra2a*, and several transporter genes such as *Slc7a6*, *Slc14a2*, and *Slc4a10*. Those specific to *Neurod1* included genes important for islet cell development and function like *Pdx1*, as well as *Pax4*, *Rest*, and *Gck. Pax6* KO embryos exhibited AS of many genes, including those important for islet cell function like *Ins1*, *Insr*, and Sstr3, apoptotic genes such as *Casp3*, *Casp8*, and *Bcl2l1*. The transcription factors *Tcf7l2* and *Ripply3* and zinc finger protein *Jazf1* were also differentially spliced in the *Pax6* KO animals.

To get an overview of the biological functions of the alternatively spliced genes, I performed enrichment analysis for gene ontology and pathway terms using Metascape (p-adj. < 0.01) (**Fig. 3.9**). It is important to note that because enrichment analysis platforms can only process a maximum of 3000 genes in a given list, only the top 3000 most significantly AS genes (based on *p*-adj. value) from the *Pax6* KO dataset were included in these analyses. Genes found to be differentially spliced across the three KO datasets were involved in protein localization (*Rab34, Ran, Ipo5, Xpo1*), regulation of protein catabolism (*Apoe, Cst3, Nedd4, Atxn3*), and mRNA metabolic process (*Srsf10, Hnrnpa2b1, Hnrnph1, Snrnp200*) which includes ribonucleoprotein complex biogenesis (**Fig. 3.9**). The enrichment of alternatively spliced genes being involved in these mRNA processes further supports the notion that *Insm1*, *Neurod1*, and *Pax6* have a role in the regulation of splicing.

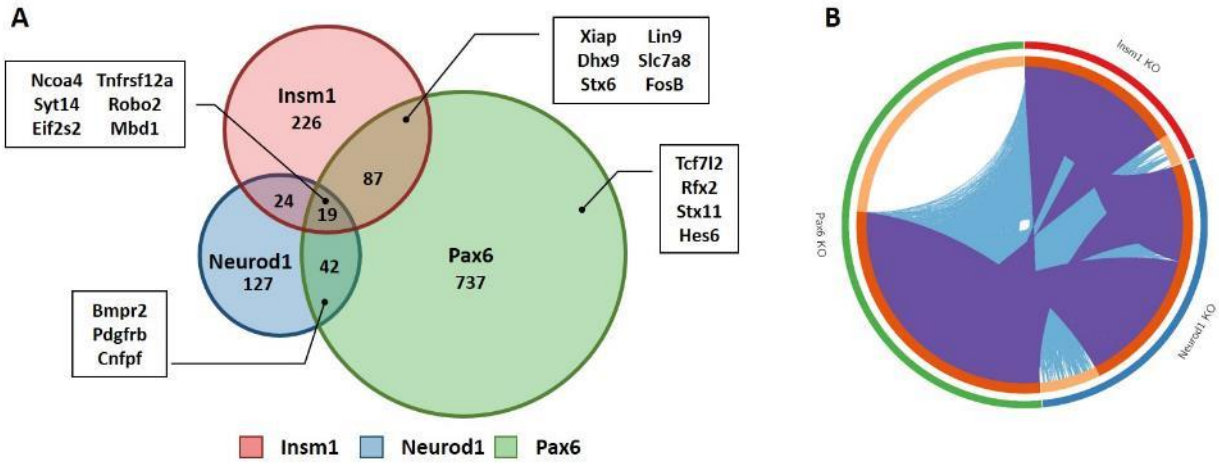


Figure 3.8. Venn diagram demonstrates overlaps between differentially spliced genes in Insm1, Neurod1, and Pax6 KO RNA-Seq datasets. (**A**) Venn diagram shows overlap of alternatively spiced genes in each of the pairwise comparisons: *Insm1* KO, *Neurod1* KO, and *Pax6* KO versus *Insm1*$^{+/-}$ datasets. Select examples of alternatively spliced genes is highlighted within a boxed insert. (**B**) Circos plot representing genes (purple curves) and GO terms/pathways (blue curves) that are shared between alternatively spliced genes from the three comparisons. Differential splicing events between the knock-out and control groups were identified by performing JunctionSeq analysis on the RNA-Seq datasets. A cutoff of p-adj. value ≤ 0.01 and $\log_2$ fold change ≥ |1| was used.

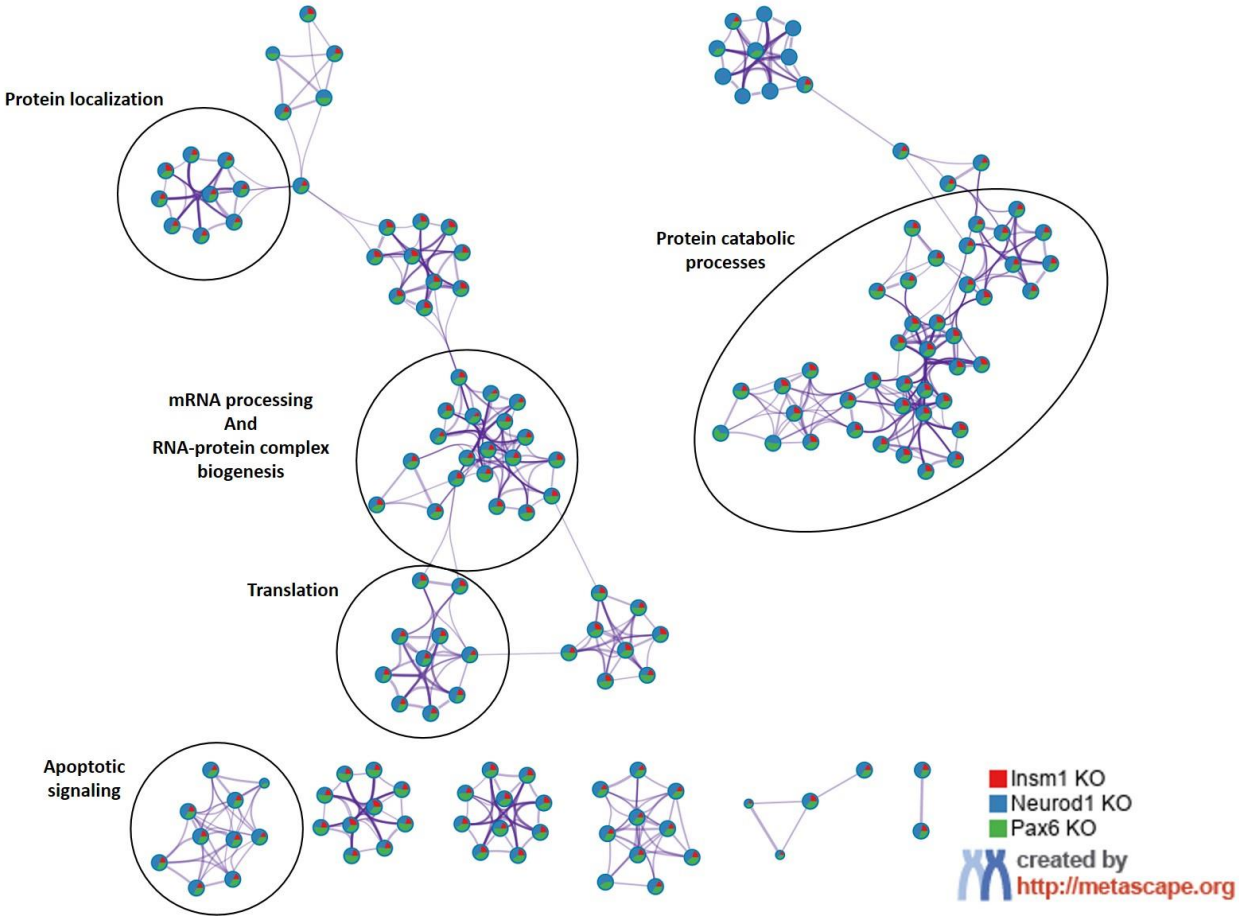


Figure 3.9. Enriched GO terms and pathways common for genes differentially spliced in *Insm1*, *Neurod1* and *Pax6* KO pancreatic endocrine cells. Network depiction of the enriched gene ontology terms shared between DRGs from the three comparisons. Node size is proportional to the number of genes in GO or pathway category, with pie charts indicating a proportion of genes from each comparison in that GO term. Metascape analyses were run using default parameters.

Examining alternative splicing events of select genes in pancreatic endocrine cell development.

Since some genes have several possible transcriptional isoforms, interpreting alternative splicing data is currently challenging. Therefore, I chose to take a closer look at a few genes that were differentially spliced in our KO datasets, and which have only a few RNA splicing variants. *Syt14* is a Ca^{2+}-independent synaptotagmin required for the exocytosis and trafficking of secretory vesicles. Interestingly differential splicing patterns were revealed for *Syt14* across all three KO datasets, specifically a decrease in expression of the Ensemble splice variant 205. To quantitatively visualize these differences, I generated splice junction plots (Sashimi plots) that tally the exon junction read counts. Sashimi plots of *Syt14* reveal a marked decrease of junction reads at exon 4 of the *Sty14*-205 splice variant (**Fig. 3.10A**). To determine if dysregulated RBPs show binding preferences near the differentially spliced loci might cause these splicing differences, I used the online oRNAment database to identify local binding motifs (Benoit Bouvrette et al., 2020) and found that TRA2A, NOVA1, and ELAVL4 all show binding patterns at exon 4, as well as adjacent exons 3 and 5 (**Fig. 3. 11A**). Importantly, expression of *Tra2A* and *Elavl4* are downregulated in all three KO gene sets, and *Nova1* downregulated in *Pax6* KO suggesting their potential involvement in differential splicing of *Syt14* in developing endocrine cells.

Similarly, *Snap 2*5 (*Synaptosomal-Associated Protein, 25kDa*) is a critical SNARE protein involved in regulating exocytosis of synaptic vesicles in the brain and hormones in pancreatic islet cells (Kadkova et al., 2019; Liang et al., 2020). Two isoforms, *Snap25a* and *Snap25b* have been identified that differ based on the alternative usage of two different exon 5 sequences (5a and 5b), with corresponding proteins differing in only nine amino acids (Kadkova et al. 2019). I observed that the *Snap25b* isoform is increased in the *Pax6* KO (p-value < 0.001) and decreased in *Insm1* KO (*p*-value < 0.01) datasets, as indicated in the associated Sashimi plots (**Fig. 3.10B**). RBP binding site analysis of this locus indicated TRA2A and NOVA1 binding sites at both 5a and 5b exons and ELAVL4 sites at nearby upstream exons (**Fig. 3.11B**). Together, these findings indicate that *Insm1, Neurod1,* and *Pax6* are required for the expression of RBPs that bind in the vicinity of exons whose proper splicing may be essential for the function of pancreatic endocrine cells.

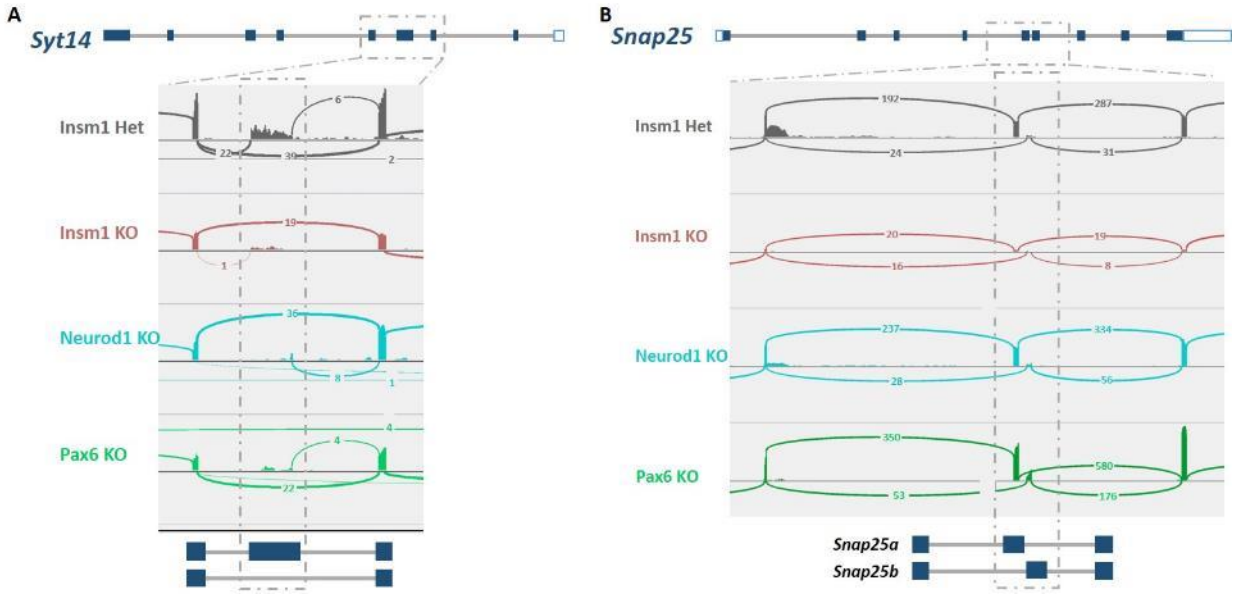


Figure 3.10. Identification of differential splicing events in *Syt14* and *Snap25* genes. The gene structures of (**A**) *Syt14* and (**B**) *Snap25* are illustrated above corresponding Sashimi plots of differentially spliced regions. Sashimi plots show the representative splice junctions as arcs that connect the exons in each of the datasets. The numbers found on each of the corresponding arcs indicate the junction depth or reads spanning the exon junctions. Dashed boxes highlight the regions of alternative splicing, which are again depicted below the Sashimi plot to provide a visual reference for read alignments. Sashimi plots were generated using the Integrated Genomic Viewer (IGV).

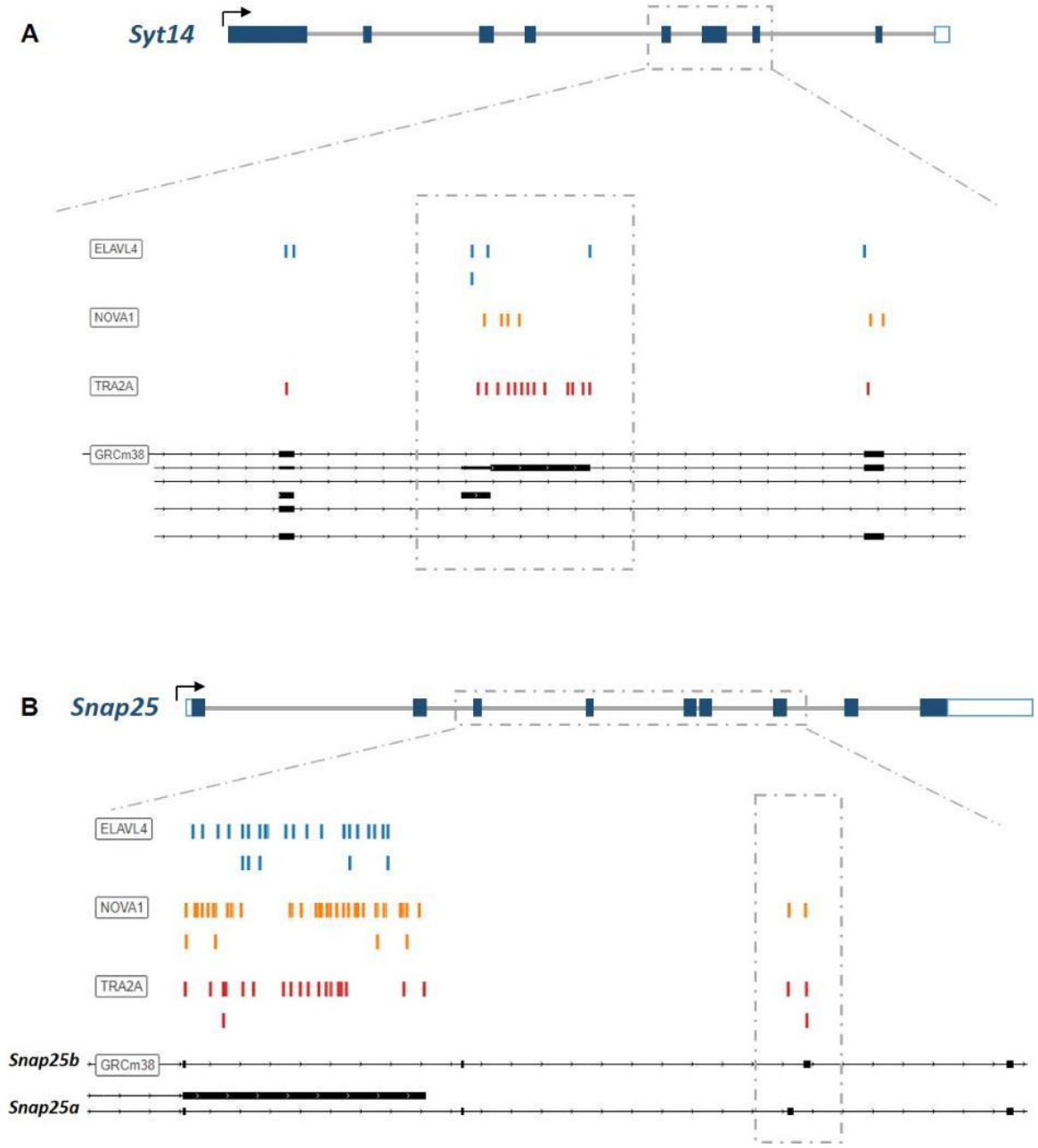


Figure 3.11. Differentially spliced regions of *Syt14* and *Snap25* genes are bound by dysregulated RNA binding proteins. (**A**) Schematic of *Syt14* gene structure and corresponding binding sites of the RBPs ELAVL4, NOVA1, and TRA2A near the differentially spliced regions. (**B**) Schematic of *Snap25* and corresponding binding sites of ELAVL4, NOVA1, and TRA2A at the differentially spliced site. Colored dashes above the indicated gene region represent the putative binding site of the given RPB. Binding sites were identified via the online oRNAment database.

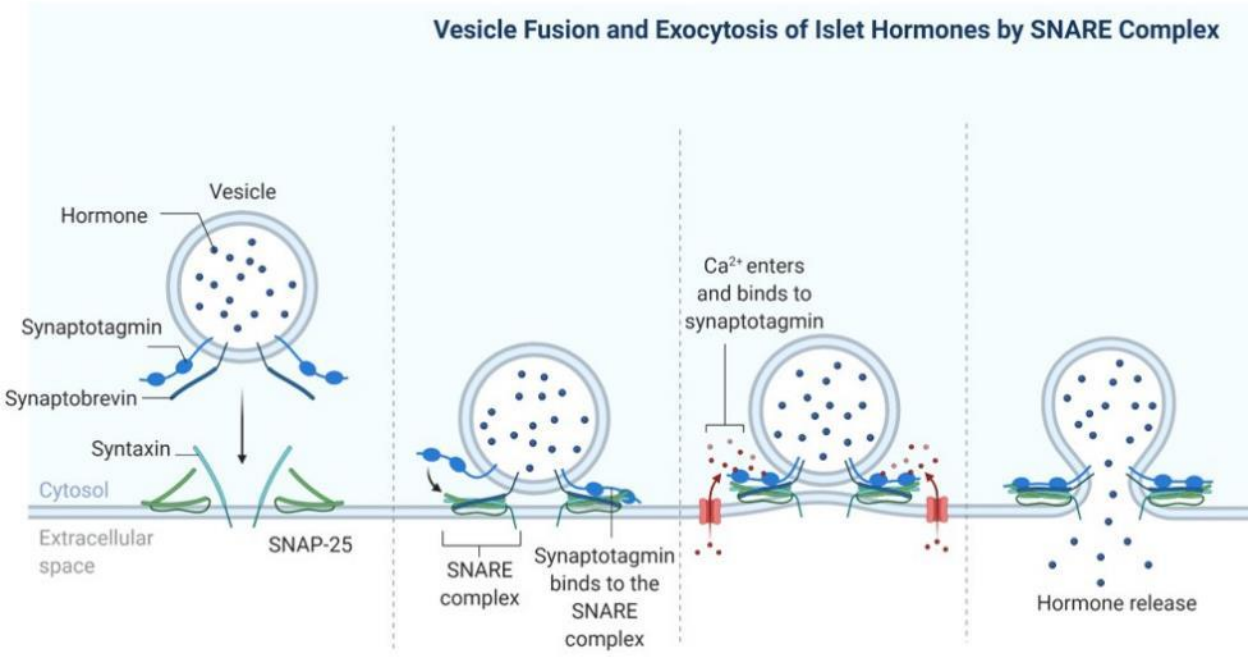


Figure 3.12. *Syt14* and *Snap25* are components of the SNARE complex important for hormone secretion. Schematic representation of the SNARE complex involved in exocytosis of insulin. Syt14 and Snap25 proteins are indicated. This original work was generated using BioRender.com.

DISCUSSION

Dysregulation of RNA binding protein gene expression and identification of alternative splicing events in Insm1, Neurod1 and Pax6 knockout animals. Alternative RNA splicing enables a single gene to give rise to multiple mRNA products that encode different protein products. This process is critical for the diversification and specification of gene (isoform) expression during the development and function of an organism. The specific function of AS protein products can vary greatly and impact cellular function in unique ways (Alvelos et al., 2018). Recent studies have suggested the importance of the unique splicing program innate to pro-endocrine cells and how the functional outputs help steer the developmental process (Alvelos et al., 2021; Colli et al., 2020; Juan-Mateu et al., 2018; Singer et al., 2019). However, the precise mechanisms regulating alternative splicing in endocrine cells remains incompletely understood, and much work is needed to shed light on how AS affects development and function in islets.

Our data indicate Insm1, Neurod1, and Pax6 regulate the expression of RNA binding proteins (RBPs) that mediate different RNA splicing events. RBP genes dysregulated include the expression of typically considered splicing enhancers (*Srsf1, Srsf3, Srsf6*) and splicing silencers (*hnrnpa0, hnrnpa3, hnrnph1*). Expression of tissue-specific RBPs such as *Elavl4*, *Tra2A*, *Msi1* and *Msi2*, and *Nova1* was also perturbed in the TF knockouts. The loss of *Nova1* has been directly demonstrated to perturb beta-cell function and shown to orchestrate AS of *Snap25* and *Insr*, highlighting its importance (Villate et al., 2014). Given the large number of RBPs dysregulated from classes known to both promote and inhibit splicing, *Insm1*, *Neurod1*, and *Pax6* appear to broadly regulate the expression of key alternative splicing factors.

Consistent with these differences in RBP gene expression, I identified multiple global differential slicing events within KO datasets. Interestingly, while many genes were differentially spliced overall, only a small fraction (1.5%) of those events were shared in each of the KO datasets. This suggests that even though these three TFs regulated the expression of common AS genes, they seem to have unique regulatory roles of alternative splicing patterns during development. Additionally, their unique AS gene sets are enriched for functionally different terms and pathways. *Insm1-, Neurod1-,* and *Pax6*-dependent AS genes were involved in actin cytoskeleton organization

(log_{10}P -4.91) and EGFR1 signaling (log_{10}P -5.38); adaptive immune system (log_{10}P -7.42) and chromatin modifiers (log_{10}P -6.30); and cell projection organization (log_{10}P -7.4) and insulin signaling (log_{10}P -4.69), respectively.

Alternative splicing of important endocrine-specific genes. It has been recently reported that in addition to the dysregulated expression of many RBP genes, diabetic patients also have differential splicing in as many as 26% of genes (Jeffery et al., 2019). Many of the same AS genes linked to diabetes in that study were differentially spliced in our datasets. The genes *Pax4, Rest,* and *Gck,* AS in *Neurod1* KO samples all have well-established roles in the development and function of islet cells. They each also have isoform-specific functions and/or links to T2D (Dlamini et al., 2017; Liang et al., 1991; Norris and Calarco, 2012; Raj et al., 2011; Sujjitjoon et al., 2016). MODY type 9 is marked by a mutation in the *Pax4* gene (*IVS7-1G>A* mutation), which results in the mis-splicing of *Pax4*, impairing its repressive functions and increasing apoptosis of beta-cells in high-glucose conditions (Sujjitjoon et al., 2016). Mutations in or near splice sites of the critical *Gck* gene (*c.459T>G;* p.Pro153Pro) have been identified in individuals with MODY (Costantini et al., 2011). *In silico* analyses of another splice site mutation (*c.45G>A;* p.Glu17SerfsX161) revealed aberrant splicing, leading to pre-mature protein degradation, and both have been shown to contribute to hyperglycemia (Costantini et al., 2011; Dlamini et al., 2017; Garin et al., 2008; Lorini et al., 2009).

Pax6-dependent AS of genes involved in insulin signaling was of note. For example, differential splicing of *Insr* results in two primary isoforms, IRA and IRB, and differ by the inclusion or exclusion of exon 11 (Moller et al., 1989; Seino et al., 1989; Whittaker et al., 2002). Both isoforms are expressed in islet cells and play important roles in islet development and function. In human islets stressed by chronically high levels of glucose, an increase in the IRB (Ex. 11^{+}) isoform expression is observed. This shift in isoform preference is linked to the progressive impairment of beta-cell function and their ability to produce and secrete insulin (Escribano et al., 2017; Hribal et al., 2003). The transcription factor *Tcf7l2* gene is also differentially spliced in *Pax6* KOs. Like *Insr*, *Tcf7l2* is known to have isoform-specific effects on β-cell function and has strong polymorphism associations with T2D (Le Bacquer et al., 2011; Osmark et al., 2009; Zhang et al., 2020).

Alternative splicing of Syt14 and Snap25. To gain explore the potential biological consequences of AS genes, I specifically looked at the *Syt14* and *Snap25* genes, both of which are components of the SNARE complex (**Fig. 3.12**). Synaptotagmin 14 (*Syt14)* is a member of the SYT family that is necessary for exocytosis of secretory

vesicles, and mutations in this gene have been associated with neurodegenerative disorders in humans (Doi et al., 2011). Our data revealed that *Syt14* is differentially spliced in all three of the KO mice studied. In each of the datasets, there appeared to be an increase of exon-4 skipping, which is predicted to result in a shorter isoform. While other SYT family factors and related members of the SNARE complex (e.g., *Syt4, Syt7, Stx4*) are known to have specific functions in islet cells, direct evidence implicating the involvement of *Syt14* is lacking (Gilbert and Blum, 2018; Huang et al., 2018; Jewell et al., 2010; Yang et al., 2001).

Similarly, *Synaptosome Associated Protein 25*, or *Snap25,* is a member of the SNARE complex. *Snap25* is necessary for insulin secretion, known to be downregulated in T2D islet cells, and is differentially spliced in both *Insm1* and *Pax6* KO endocrine cells (Daraio et al., 2017; Gonelle-Gispert et al., 1999; Jewell et al., 2010; Ostenson et al., 2006; Sadoul et al., 1995). It is well established that *Snap25* has two primary mRNA isoforms that have slightly different localization patterns and functional traits, with a developmental switch from *Snap25a* to -2*5b* being observed in the brain (Bark et al., 2004; Bark et al., 1995; Boschert et al., 1996; Daraio et al., 2017; Irfan et al., 2019; Jacobsson et al., 1999; Nagy et al., 2005). Both isoforms are expressed in islet cells, but *Snap25a* is dominant in beta-cells (Gonelle-Gispert et al., 1999; Jeans et al., 2007), and subtle differences in functional dynamics exist between the two. *Snap25*-deficient mice display alterations in insulin secretion, asynchronous fluctuations in intracellular Ca^{2+} concentrations, and become obese and develop diabetes (Daraio et al., 2017; Valladolid-Acebes et al., 2015). I show here that *Snap25* exons 5a/5b and surrounding sequences contain putative NOVA1 binding sites. *Nova1*-KD in FACS sorted rat beta cells and in INS-1E cells results in a decrease of the *Snap25b* isoform, correlated by perturbed glucose-stimulated insulin secretion (Villate et al., 2014). This *Nova1*-dependent mechanism of *Snap25* AS in endocrine cells is consistent with the downregulation of *Nova1* gene expression and shift in Snap25 isoforms observed in our data.

In conclusion, the present study provides evidence for a novel role of *Insm1*, *Neurod1*, and *Pax6* in regulating alternative splicing patterns during endocrine cell development. I show that the differential expression of RBP splicing factors correlates with alternative splicing events. This includes the AS of key endocrine genes, such as *Pax4*, *Gck,* and *Insr,* highlighting the importance of alternative splicing in endocrine cell development and function. Closer inspection of differential splicing in both *Syt14* and *Snap25* suggests a potential mechanism of contributing to the impaired function of KO islets in TF KOs. Finally, our data demonstrate the broad impact of

eliminating each TF on both gene expression and AS, further illustrating their individual roles and importance in establishing the GRN within pancreatic endocrine cells.

CHAPTER IV - ROLE AND REGULATION OF ZINC FINGER PROTEINS IN ENDOCRINE DEVELOPMENT

INTRODUCTION

Discovery and classification of zinc finger proteins.

Zinc finger proteins (ZFPs/ZNFs) constitute a class of proteins that contain an array of zinc finger domains that are stabilized through interactions between cysteine and histidine residues with zinc ions. These makeup one of the most abundant classes of proteins in the human genome and are comprised of over 1500 gene members (**Table 4.1**) (Cassandri et al., 2017). However, our understanding of these genes remains incompletely understood. The first ZFP, Transcription factor IIIA (TFIIIA), was discovered in the *Xenopus* model. It was originally noted that TFIIIA's interaction with DNA was dependent on the presence of zinc (Hanas et al., 1983), and shortly after, its protein structure was described and defined by the presence of 9 C_2H_2 zinc finger domains (Miller et al., 1985). This discovery ultimately led to the identification of many new ZFP members containing the same or similar protein domains required for binding to DNA sequences.

Zinc finger proteins come in an estimated 30 different ZFP types (**Table 4.1**) (Cassandri et al., 2017). These subtypes vary based on the domain structures of the proteins, which differ in the combinations of cysteine/histidine residues that interact with zinc ions (Cassandri et al., 2017). Examples of the various zinc finger domains are depicted in **Figure 4.1**. Important subtypes include C_2H_2, *GATA zinc-finger domain-containing* (GATAD), *really interesting new gene* (RING), ZF class homeoboxes and pseudogenes, LIM (*LIN-11*, *Isl-1,* and *MEC-3*), and *plant homeodomain* (PHD). To add to the complexity, a given ZFP can also contain multiple types of ZFP domains in varying numbers, as well as other non-zinc finger domains like the Kruppel-associated box domain (KRAB). For example, some can be very simple in structure, such as *Klf4,* which contains only three C_2H_2 domains (SMART database - Q60793). Others can be highly complex, such as *Zfhx4*, which contains as many as 23 C_2H_2, 7 U1-like, and 4 HOX domains (SMART database - Q9JJN2).

The functional roles of ZFPs are just as broad as the range of subtypes. Besides just DNA binding, many ZFP domains can also mediate interactions with RNA or with other proteins. Thus, ZFPs have been shown to be involved in transcriptional regulation (*Insm1, GATA4/6, Isl1*), ubiquitylation (*Rbx1, Cul7*), and chromatin modifications (*Prdm9, Zfp-1, KMT2H/Ash1/Ash1l, KMT8A/Prdm2*) (Mihola et al., 2009; Singh et al., 2016), among others. The functional role of a protein is often influenced by the type of ZFP domains a protein contains (**Fig. 4.2**) (Vilas et al., 2018). For example, C_2H_2-type domains are often associated with transcriptional regulation (Brayer and Segal, 2008; Bruno et al., 2019; Vaquerizas et al., 2009), RING domain-containing genes frequently have E3-ubiquitin ligase activity (Cassandri et al., 2017; Joazeiro and Weissman, 2000), and PHD-containing proteins often function as epigenetic “readers” that can affect chromatin organization and epigenetic modifications (Jain et al., 2020; Sanchez and Zhou, 2011).

Table 1. Types of zinc-finger proteins

Type name	*Zinc-finger structure*	*Number of genes*	*Number of TF*	*Important members*
Zinc fingers C2H2-type (ZNF)	C-x-C-x-H-x-H	720	372	*KLF4, KLF5, EGR3, ZFP637, SLUG, ZNF750, ZNF281, ZBP89, GLIS1, GLIS3*
Ring finger proteins (RNF)	C-x-C-x-C-x-H-xxx-C-x-C-x-C-x-C	275	12	*MDM2, BRCA1, ZNF179*
PHD finger proteins (PHF)	C-x-C-x-C-x-C-xxx-H-x-C-x-C-x-C	90	0	*KDM2A, PHF1, ING1*
LIM domain containing	C-x-C-x-H-x-C-x-C-x-C-x-C-x-(C,H,D)	53	1	*ZNF185, LIMK1, PXN*
Nuclear hormone receptors (NR)	C-x-C-x-C-x-C-xxx-C-x-C-x-C-x-C	50	47	*VDR, ESR1, NR4A1*
Zinc fingers CCCH-type (ZC3H)	C-x-C-x-C-x-H	35	2	*RC3H1, HELZ, MBNL1, ZFP36, ZFP36L1*
Zinc fingers FYVE-type (ZFYVE)	C-x-C-x-C-x-C-xxx-C-x-C-x-C-x-C	31	0	*EEA1, HGS, PIKFYVE*
Zinc fingers CCHC-type (ZCCHC)	C-x-C-x-H-x-C	25	2	*CNBP, SF1, LIN28A*
Zinc fingers DHHC-type (ZDHHC)	C-x-C-x-H-x-C-xxx-C-x-C-x-H-x-C	24	0	*ZDHHC2, ZDHHC8, ZDHHC9*
Zinc fingers MYND-type (ZMYND)	C-x-C-x-C-x-C-xxx-C-x-C-x-H-x-C	21	4	*PDCD2, RUNX1T1, SMYD2,SMYD1*
Zinc fingers RANBP2-type (ZRANB)	C-x-C-x-C-x-C	21	3	*YAF2, SHARPIN, EWSR1*
Zinc fingers ZZ-type (ZZZ)	C-x-C-x-C-x-C	18	3	*HERC2, NBR1, CREBBP*
Zinc fingers C2HC-type (ZC2HC)	C-x-C-x-H-x-C	16	2	*IKBKG, L3MBTL1, ZNF746*
GATA zinc-finger domain containing (GATAD)	C-x-C-x-C-x-C	15	15	*GATA4, GATA6, MTA1*
ZF class homeoboxes and pseudogenes	C-x-C-x-H-x-H	15	10	*ADNP, ZEB1, ZHX1*
THAP domain containing (THAP)	C-x-C-x-C-x-H	12	3	*THAP1, THAP4, THAP11*
Zinc fingers CXXC-type (CXXC)	C-x-C-x-C-x-C-xxx-C-x-C-x-C-x-C	12	2	*CXXC1, CXXC5, MBD1,DNMT1*
Zinc fingers SWIM-type (ZSWIM)	C-x-C-x-C-x-H	9	0	*MAP3K1, ZSWIM5, ZSWIM6*
Zinc fingers AN1-type (ZFAND)	C-x-C-x-C-x-C-xxx-C-x-H-x-H-x-C	8	0	*ZFAND3, ZFAND6, IGHMBP2*
Zinc fingers 3CxxC-type (Z3CXXC)	C-x-C-x-H-x-C	8	0	*ZAR1, RTP1,RTP4*
Zinc fingers CW-type (ZCW)	C-x-C-x-C-x-C	7	0	*MORC1, ZCWPW1,KDM1B*
Zinc fingers GRF-type (ZGRF)	C-x-C-x-C-x-C	7	0	*TTF2, NEIL3, TOP3A*
Zinc fingers MIZ-type (ZMIZ)	C-x-C-x-H-x-C	7	1	*PIAS1, PIAS3, PIAS4*
Zinc fingers BED-type (ZBED)	C-x-C-x-H-x-H	6	2	*ZBED1, ZBED4, ZBED6*
Zinc fingers HIT-type (ZNHIT)	C-x-C-x-C-x-C-xxx-C-x-C-x-H-x-C	6	0	*ZNHIT3, DDX59, INO80B*
Zinc fingers MYM-type (ZMYM)	C-x-C-x-C-x-C	6	6	*ZMYM2, ZMYM3, ZMYM4*
Zinc fingers matrin-type (ZMAT)	C-x-C-x-H-x-H	5	0	*ZNF638, ZMAT1, ZMAT3, ZMAT5*
Zinc fingers C2H2C-type	C-x-C-x-H-x-H	3	3	*MYT1, MYT1L, ST18*
Zinc fingers DBF-type (ZDBF)	C-x-C-x-H-x-H	3	0	*DBF4, DBF4B, ZDBF2*
Zinc fingers PARP-type	C-x-C-x-H-x-C	2	1	*LIG3, PARP1*

Table 4.1 Types of zinc finger proteins. A list of zinc finger types with a brief description of the zinc-finger domain structure, the number of genes included, and the most studied members is summarized. The associate protein structures of these domains are depicted in Figure 4.1. This table was reprinted with permission from Cassandri et.al., 2017.

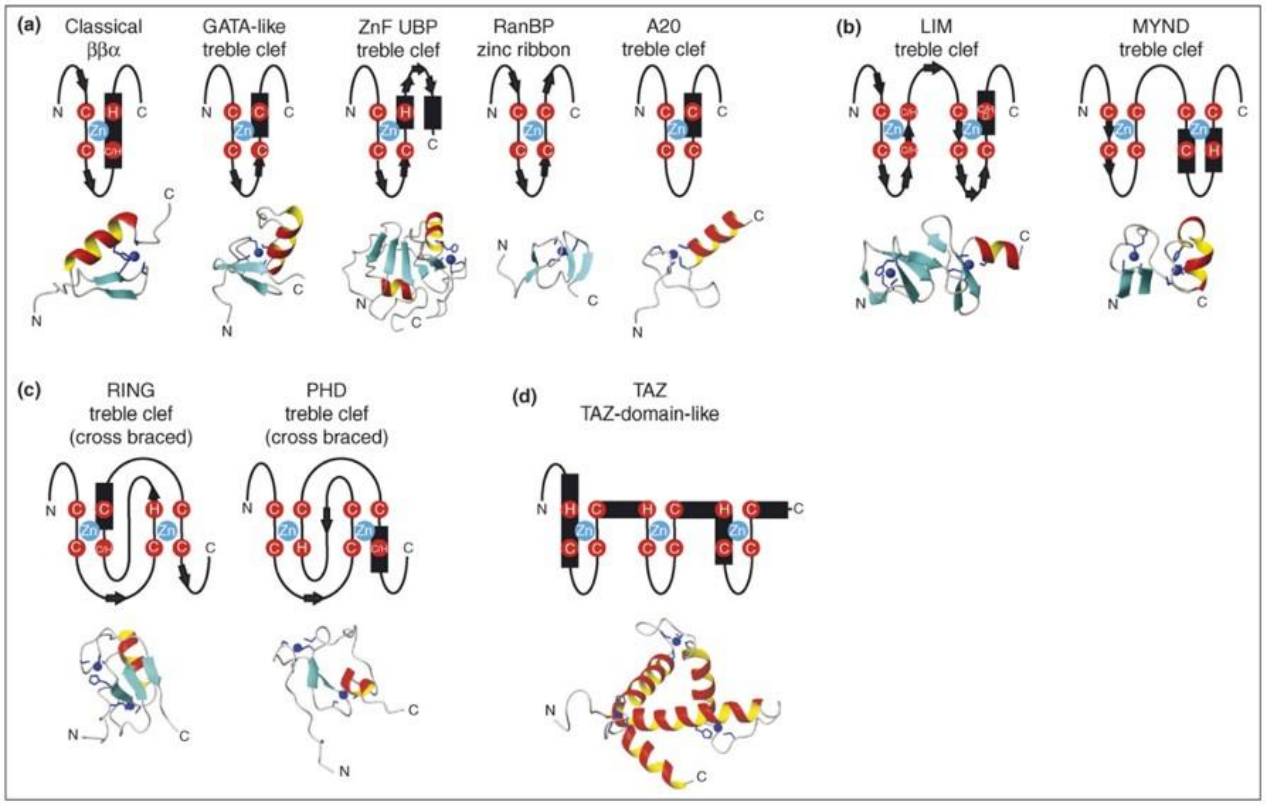
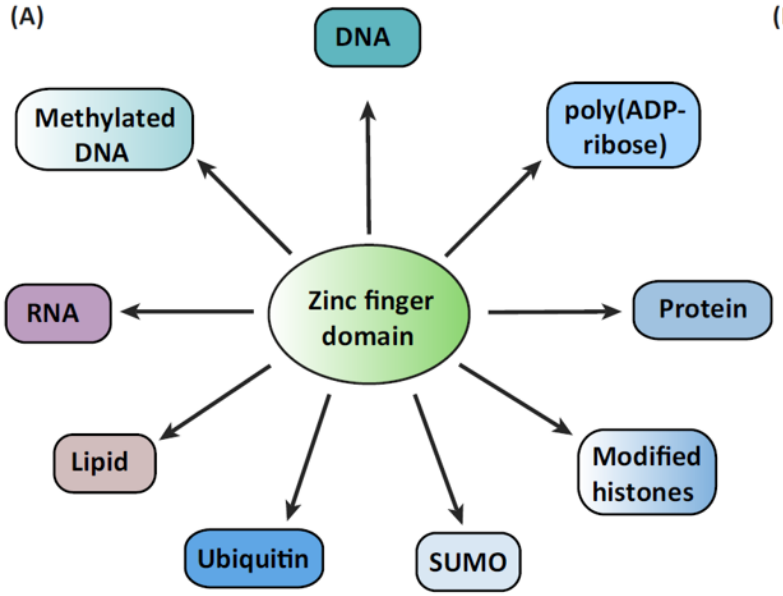


Figure 4.1. Topology and structures of ZnF domains. 3-dimentional renditions of the secondary and tertiary structures of zinc finger domain subtypes. (**A**) Classical (PDB code 1FU9), GATA (1GAT), ZnF UBP (2G43), RanBP (1Q5W), A20 (2FID). (**B**) LIM (1A7I) and MYND (1FV6). (**C**) RING (1CHC) and PHD (1XWH). (D) TAZ (1R8U). This figure was modified and reprinted with permission from Gamsjaeger, et al., 2007.



(B)

Binding substrate	Zinc finger domain
DNA	C2H2, KRAB
Methylated DNA	C2H2, FYVE
RNA	C2H2, CHHC
Lipid	C2H2, FYVE
Protein	C2H2, RING, ZZ -type
Ubiquitin	KRAB, PBZ, UBP
SUMO	MYM-Type
Modified Histones	PHD, C5HC2
Poly(ADP-ribose)	PBZ

Figure 4.2. Functional diversity and macromolecular binding specificities of zinc finger domains. (**A**) ZnF domains bind to several types of substrates in addition to nucleic acids. (**B**) List of ZnF domains and their binding substrates. This figure was modified and reprinted with permission from Vilas, et. al., 2018.

Another ZFP subtype, Krüppel-associated box domain zinc finger proteins (KRAB-ZFPs), are characterized by the presence of a KRAB domain at its N-terminal and C_2H_2-type zinc fingers at its C-terminal. These proteins typically function as transcriptional repressors. Interestingly, zinc finger proteins have expanded greatly throughout evolution and are one of the most rapidly evolving classes of genes, but KRAB-ZFPs appears to have arisen with mammals (Bruno et al., 2019; Emerson and Thomas, 2009; Mackeh et al., 2018; Thomas and Emerson, 2009). Furthermore, emerging data has pointed to the evolution of KRAB-ZFPs coinciding with that of transposable elements (TEs) (**Fig. 4.3**) (Bruno et al., 2019; Castro-Diaz et al., 2014; Ecco et al., 2017; Imbeault et al., 2017; Jacobs et al., 2014; Thomas and Schneider, 2011). TEs are regions of DNA capable of "jumping" around the genome and of acquiring additional mutations along the way. These mobile genetic elements are thought to be important for the evolution process and makeup as much as 50% of the human genome (Bruno, 2019). Recent reports have identified KRAB-ZFPs as being important for the silencing of TEs in the genome, and the genetic drift of KRAB-ZFPs parallels that of their TE targets (Bruno et al., 2019; Ecco et al., 2017). The repression of TEs by KRAB-ZFPs has been tied to the regulation of cell proliferation, differentiation, and apoptosis, and binding to TEs functions in the tissue-specific regulation of expression of nearby genes (Imbeault et al., 2017; Lupo et al., 2013). Thus, the evolution of ZFPs has been instrumental to the development of advanced species and can take on specialized roles.

While ZFPs remain generally understudied, many that have been characterized have been shown to play a role in various aspects of development and disease states. Through the regulation of gene expression, ZFPs regulate several processes, including differentiation, homeostasis, apoptosis, and proliferation in most cell types (**Fig. 4.4**) (Cassandri et al., 2017). For example, one of the most well-studied ZFPs is the CCCTC-Binding Factor, CTCF. CTCF is a highly conserved zinc finger protein that contains 11 C_2H_2 domains that is ubiquitously expressed and was originally identified as a repressor *c-myc* expression (Filippova et al., 1996; Klenova et al., 1993). It has multiple functional roles and is involved in chromatin modifications, cell cycle regulation, transcriptional activity, and alternative splicing (**Fig. 4.5**) (Arzate-Mejia et al., 2018; Franco et al., 2014; Pongubala and Murre, 2021). However, one of its most important roles is as an insulator protein and in coordinating intrachromosomal interactions over long distances (Agrawal and Rao, 2021). By binding at super-enhancers, CTCF facilitates chromosomal looping to bring promoter regions in proximity of the tissue-specific super-enhancer regions, thus regulating tissue-specific gene expression during development and in mature cells (Alpsoy et al., 2021; Arzate-Mejia et al., 2018; Kang and

Lee, 2021). Knockout studies have shown that animals lack lacking CTCF die during early development, at E5.5 in mice and ~24hpf in zebrafish (Carmona-Aldana et al., 2018; Moore et al., 2012). In the islet cells, CTCF has been shown to regulate glucose homeostasis through governing the expression of *Ins* and *Gcg*, as well as the expression of *Pax6* (Fang et al., 2014).

Other ZFPs are also known to have functional roles during pancreas development. For example, *Insm1* is C_2H_2-containing TF critical for proper endocrine cell expansion and differentiation (Osipovich et al., 2014). The TFs *Glis3* (*GLIS Family Zinc Finger 3*) and *Isl1* (*ISL LIM Homeobox 1*) have both been implicated in the development and function of pancreatic endocrine cells (Ediger et al., 2014; Guo et al., 2011; Kang et al., 2009; Kang et al., 2016; Scoville et al., 2019). *Glis3* belongs to the Krüppel-like family of proteins and contains five C_2H_2 domains (Jetten, 2018). In developing endocrine cells, Glis3 directly regulates expression of *Neurog3,* and KO animals become severely hyperglycemic, dying by P11 (Dimitri et al., 2011; Habeb et al., 2012; Senee et al., 2006). Pancreas-specific KO models also develop diabetes in adulthood and reveal that Glis3 regulates the expression of genes critical for beta cell identity and function (*Ins, Mafa, Slc2a2, Abcc8, Isl1*) (**Fig. 4.4**) (Scoville et al., 2019). Similarly, *Isl1* is a LIM-type ZFP that contains two LIM ZF domains and one homeobox domain. *Isl1* is important during both developing and mature endocrine cells, functioning to promote proliferation during development and to maintain cell identity and function in mature cells (Ediger et al., 2014; Guo et al., 2011).

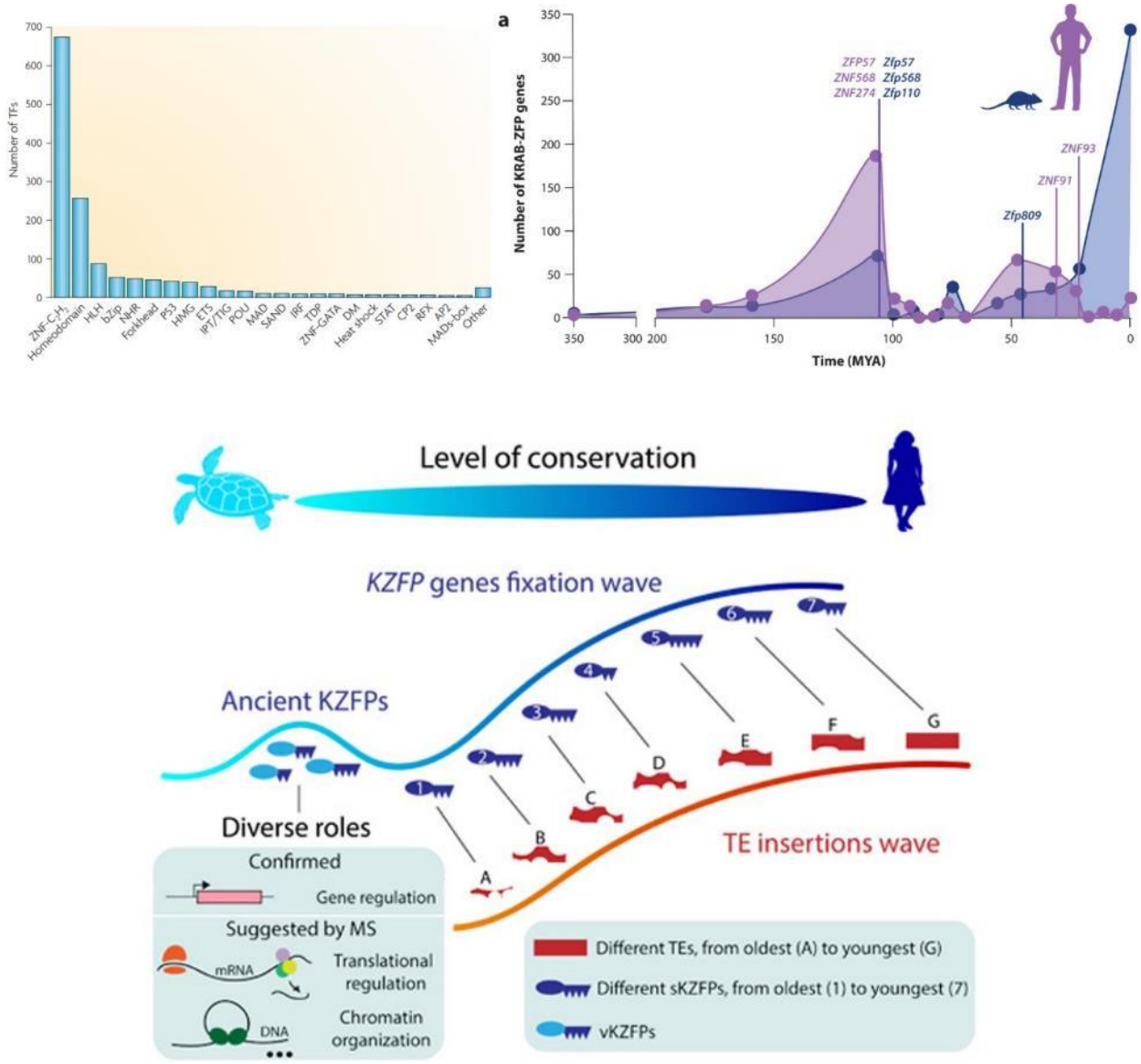


Figure 4.3. Evolution of KRAB-ZFP genes. Histogram of the approximate ages of all existent KRAB-ZFP genes in mice and humans. The approximate age of each KRAB-ZFP gene was determined first by identifying orthologs based on zinc fingerprint alignments as described in Reference 42, and then by estimating evolutionary distances between species using the TimeTree database. A noticeable burst of new KRAB-ZFP genes occurred prior to the mammalian radiation over 105 MYA, with separate bursts along the human tree prior to the split from New World monkeys (68 MYA). In the mouse lineage a recent burst of KRAB-ZFP genes since the split with rats has dramatically increased the number of KRAB-ZFP genes. This figure was modified and reprinted with permission from Bruno, et al., 2019.

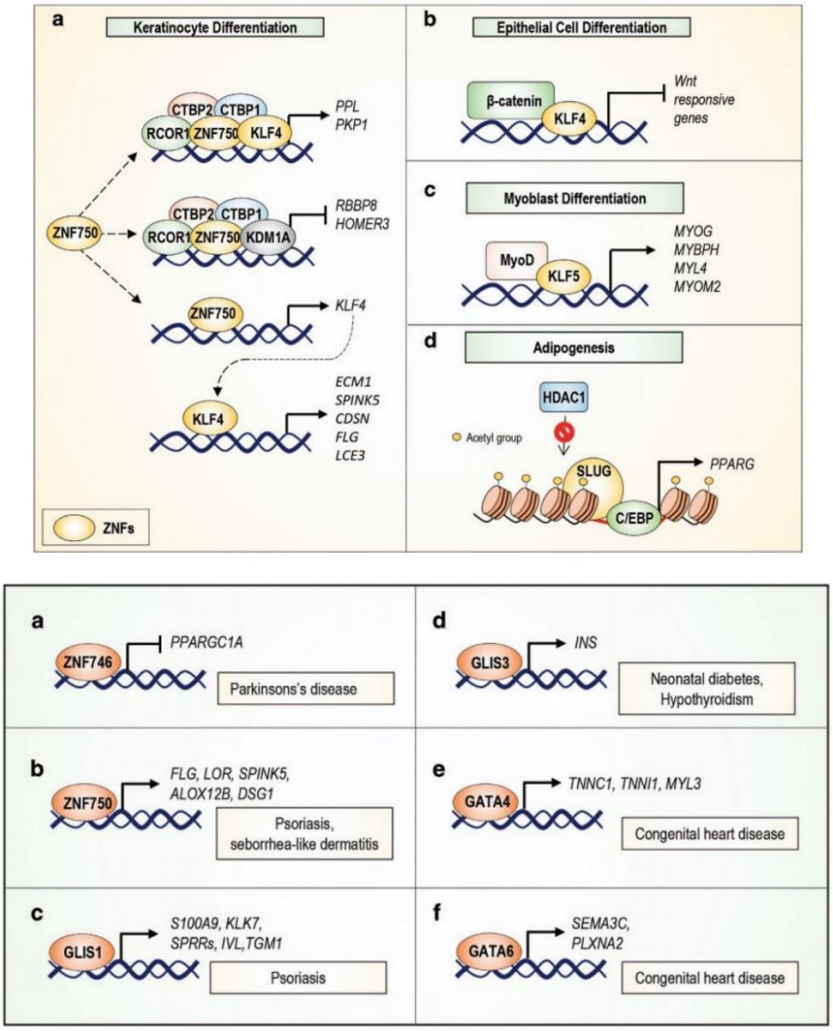


Figure 4.4. Molecular pathways regulated by ZNFs in physiological conditions. TOP: (**A**) ZNF750 regulates keratinocytes terminal differentiation by interacting with KLF4 and chromatin regulators. This interaction leads to the positive regulation of genes (PPL, PKP1) involved in differentiation. In addition, ZNF750 interacts with KDM1A and negatively regulates progenitor gene expression (RBBP8, HOMER3). ZNF750 directly regulates the expression of KLF4, which subsequently modulates the expression of the indicated genes. (**B**) KLF4 regulates epithelial cell differentiation by interacting with β-catenin and repressing the WNT signalling pathway. (**C**) KLF5 is involved in myoblast differentiation, acting as a co-factor for MyoD. This action leads to the upregulation of the indicated genes. (**D**) The presence of SLUG on the PPARG promoter reduces HDAC1 recruitment, leading to C/EBP-mediated activation of PPARG expression. This effect promotes adipogenesis. **BOTTOM**: (**A**) ZNF746 represses the expression of PGC-1α, resulting in the loss of dopaminergic neurons in the substantia nigra of Parkinson's patients. (b) ZNF750 regulates the expression of epidermal differentiation markers, such as FLG, LOR, SPINK5, ALOX12B, and DSG, which are altered in human skin diseases. (**C**) Glis1 regulates transcription of several genes involved in the differentiation of epidermal keratinocytes, including cornifin, involucrin, and transglutaminase 1. The expression of these genes is altered in psoriasis. (**D**) Glis3 modulates expression of the insulin gene, contributing to the pathogenesis of neonatal diabetes and hypothyroidism. (**E**) Troponin C and I and myosin light chain-3 genes are induced during cardiac hypertrophy due to overexpression of the GATA4 transcription factor. (**F**) The expression of SEMA3C and its receptor PLXNA2 is downregulated by GATA6 mutations, resulting in the development of OFT defects associated with CHDs. This figure was modified and reprinted with permission from Cassandri, et. al., 2017.

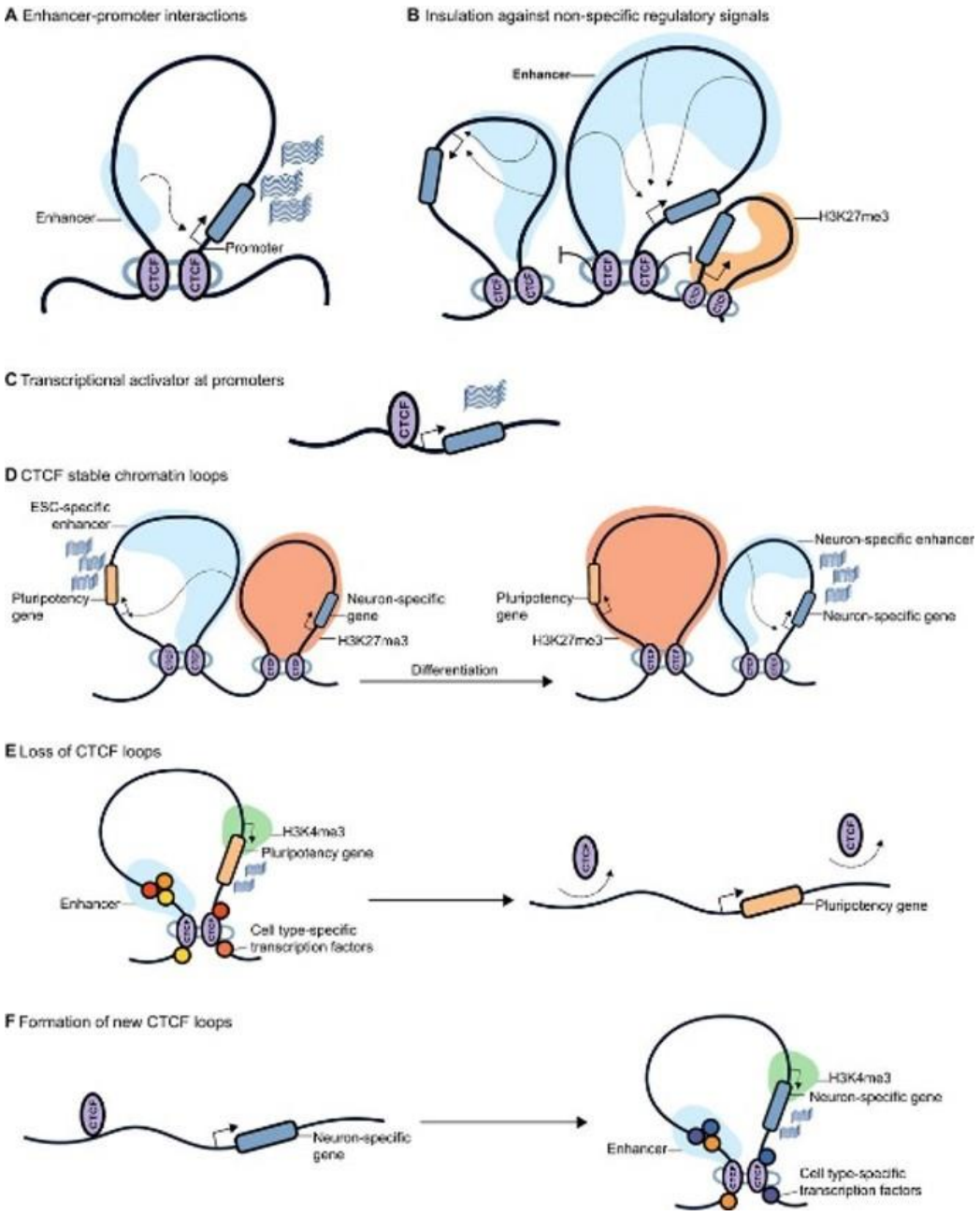


Figure 4.5. Modes of CTCF function during cell differentiation and development. (**A-C**) CTCF can promote the physical interaction between enhancers and promoters (**A**), insulate genes from neighboring regulatory signals by chromatin looping (**B**) or mediate transcriptional activation by binding at the promoter regions of certain genes (**C**). (**D**) During cell differentiation, CTCF can also promote the establishment of a basal, stable topology upon which regulatory transactions can take place. Some chromatin loops can be formed early in development and remain unchanged during differentiation, although genes inside could be subject to differential gene expression owing to the binding of specific transcription factors. (**E,F**) In addition, some chromatin loops can be lost (**E**) or formed de novo by recruitment of CTCF and perhaps cell type-specific transcription factors (**F**). This figure was modified and reprinted with permission from Rodrigo, et. al., 2018.

Clever methodologies for harnessing the specificity of ZFPs have been employed to facilitate genetic manipulations (Urnov et al., 2010). Due to the pronounced specificity of ZFPs by their zinc finger domains, researchers can engineer a series of ZF domains to bind at a specific region of interest in the genome. By attaching these artificial proteins to other domains with different functional roles, targeted mutagenesis and transcriptional regulation could be achieved. For example, by incorporating nucleases (such as *FokI*), zinc-finger nucleases were born (ZFNs) and could introduce mutations at the targeted loci. The addition of other functional domains could also be added to achieve other genetic manipulations, such as the addition of transcriptional regulators to regulate gene expression (**Fig. 4.6**) (Gupta and Musunuru, 2014; Jamieson et al., 2003). While conceptually powerful, the generation of these could be cumbersome and efficient targeting difficult. ZFNs were soon replaced by more efficient TALENs (transcription activator-like effector nucleases), and most recently, by CRISPR-Cas9 technology (Carroll, 2017).

In this study, I have investigated the role of select ZFPs in pancreas development. The five ZFPs discussed here are members of a developmental gene co-expression network (Osipovich et al., 2021) and co-expressed with multiple endocrine-specific transcription factors. I found that two of these ZFPs, *Jazf1,* and *Zfp800*, play particularly important roles in development and warrant further study.

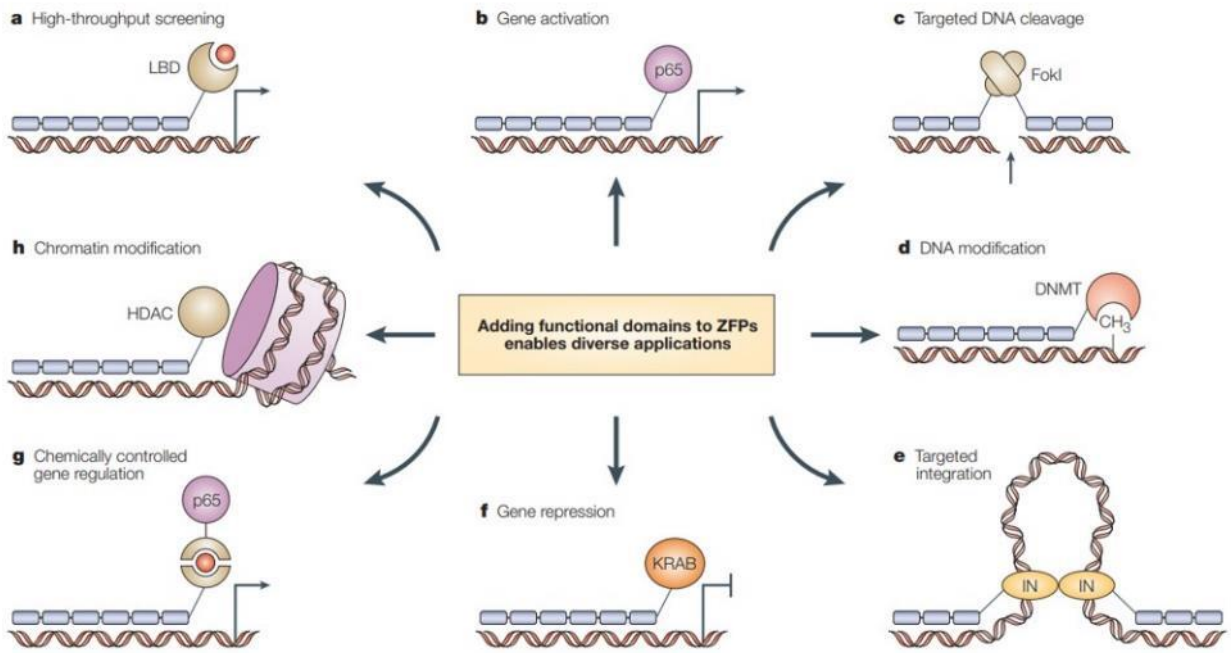


Figure 4.6. Harnessing the specificity of zinc finger proteins to facilitate genetic manipulations. An engineered zinc-finger protein (ZFP) can be combined with different functional domains for many different applications. The ZFPs are depicted as horizontal arrays of grey rectangles containing either three or six fingers. Pictorial representations of functional domains are attached to the ZFPs by short lines, representing linkers. (**A**) A ZFP combined with a ligand-binding domain (LBD), for use in high-throughput screening. (**B**) A ZFP combined with the p65 domain for activating transcription. (**C**) A ZFP combined with a nuclease domain from a Type II restriction enzyme, such as FokI, for targeted DNA cleavage. (**D**) A ZFP combined with a DNA methyltransferase (DNMT) domain for DNA modification. (**E**) A ZFP combined with a retroviral integrase (IN), for inserting exogenous sequences into DNA. (**F**) A ZFP combined with a repression domain, such as the KRAB domain of the KOX protein, for repressing transcription. (**G**) A ZFP transcription factor that only fully assembles in the presence of a small-molecule drug. (**H**) A ZFP combined with a chromatin-modifying domain, such as a histone deacetylase (HDAC), for small-molecule screening. This figures was modified and reprinted with permission from Jamieson, et al., 2003.

MATERIALS AND METHODS

Generation of knockout mice. *Zfp800*, *Jazf1, Zfp92, Zfp329,* and *Zfhx4* knockout mice were obtained by microinjecting target-specific guide RNA (gRNA) complexes into Cas9-expressing embryos obtained from the mating of homozygous B6J/129(Cg)- *Gt(ROSA)26Sor*$^{tm1.1(CAG-cas9*,-EGFP)Fezh/J}$ mice (JAX 026179, MGI: J:213550) with C57BL/6N (JAX 000664) mice. For each target gene, a pool of three gRNA complexes (crRNA+tracrRNA) (**Table 4.2**) were microinjected into single-cell embryos at a final concentration of 20 µM. The guide RNA complexes were made by first diluting crRNA (IDT) and tracrRNA (IDT) to a 100 µM stock in IDT Duplex Buffer. A mixture of 2 µl of crRNA (100 µM), 2 µl of tracrRNA (100 µM) and 6 µl of water was then heated to 95°C for 5 min and then slowly cooled to 25°C to anneal. To make the final pooled injection mixture, 1.5 µl (~675 ng) of each complex was mixed with the injection buffer (1 mM Tris HCl, pH 7.5, 0.1 mM EDTA, filtered through 0.2 µm) to a final concentration of 20 µM, or approximately 10 ng/µl of each gRNA complex. Founder animals resulting from the gRNA complex microinjections were screened for INDEL mutations via T7-assays (NEB) and by determining heterodimer formation via PAGE-gel analyses. For screening by PAGE-gel, PCR products using primers flanking the targeted site were heated to 95°C for 5 min and then slowly cooled to room temperature to reanneal. PCR products were then run on a 10% TBE PAGE gel (Invitrogen). PCR products for mutant founders were then analyzed by Sanger sequencing (GENEWIZ). All microinjections and founder screenings were performed by the Vanderbilt Gene Editing Resource. Founder animals were then bred to C57BL6/J mice for seven generations while the Cas9-expressing ROSA26 allele was segregated out. Pups were genotyped by PCR using primers surrounding the deletion sites (**Fig. 4.11 and Table 4.2**).

Body weight and composition measurements. Animal body weights were initially taken at the time of weaning (3 weeks old) and weekly thereafter. Nuclear magnetic resonance (NMR) analysis was performed on 12-week old animals to obtain body composition measurements of control and knockout mice using a Bruker Mini Spec Body Composition Analyzer. Quality control checks for NMR parameters and calibrations were performed before each testing using a standard provided by the manufacturer. Live, conscious mice were placed into a clear, cylindrical tube and immobilized using a plunger. The tube was placed into the analyzer for scanning for ~2 minutes. This assay was terminal, and mice were euthanized following data collection.

Glucose tolerance testing. Intraperitoneal glucose tolerance tests (IGTTs) were performed at 12 weeks of age for *Jazf1* and *Zfp92* KO animals following a 16-h overnight fast. Blood glucose concentrations were measured at 0, 15, 30, 60, and 120 min after administering d-glucose (2 mg/g body mass) using an Accu-Check glucometer.

Plasma insulin and IGF-1 and GH measurements. Plasma was collected from the tail vein of *Jazf1* and *Zfp92* KO and WT animals at 12 weeks following a 16-h overnight fast. The plasma sample was put into a purple top EDTA coated tube (Sarstedt Inc. #16.444.100) and centrifuged for ~10 minutes at 4°C, then stored at -80°C until assayed. Magnetic Luminex Assays (using x-map technology via the MagPix system) were performed according to the manufacturer's instructions to measure IGF-1 (R&D Systems #LXSAMSM-01, IGF-1) and GH (R&D Systems #LXSAMSM-01, GH) levels using targeted antibodies. Protein concentrations were compared after normalization by the total protein concentration of the extract. This assay was performed with the assistance of the VUMC Hormone Assay and Analytical Services Core which is supported by NIH grants DK059637 and DK020593.

Oligo name	Sequence (5' to 3')	PCR product size	Application
Zfp800 crRNA-1	GTTCGCAGCATCCGTGATGA	Not applicable	CRISPR/Cas9-mediated disruption of *Zfp800* gene
Zfp800 crRNA-2	CAAACGGACCACCATCATCA		
Zfp800 crRNA-3	CGCAGCATCCGTGATGATGG		
Jazf1 crRNA-1	CTCGATGAGGTCGGCCAGCG	Not applicable	CRISPR/Cas9-mediated disruption of *Jazf1* gene
Jasf1 crRNA-2	TCCAACACCTGCCGCTTCGG		
Jazf1 crRNA-3	CGACCTCATCGAGCACATCG		
Zfp92 crRNA-1	GTTCGCAGCATCCGTGATGA	Not applicable	CRISPR/Cas9-mediated disruption of *Zfp92* gene
Zfp92 crRNA-2	CAAACGGACCACCATCATCA		
Zfp92 crRNA-3	CGCAGCATCCGTGATGATGG		
Zfp329 crRNA-1	ACATGGAAGCCGTTTGTCCC	Not applicable	CRISPR/Cas9-mediated disruption of *Zfp329* gene
Zfp329 crRNA-2	CCCCATTTTAGGTGATAATT		
Zfp329 crRNA-3	TCCCAATTATCACCTAAAAT		
Zfhx4 crRNA-1	GCGTCTCCGCCATAATACAG	Not applicable	CRISPR/Cas9-mediated disruption of *Zfhx4* gene
Zfhx4 crRNA-2	CGGAGACGCATTTATTACTG		
Zfhx4 crRNA-3	GCCAATCCCCTGTATTATGG		
Zfp800-Fwd	ATGGGTTTGTGCTTCTCTGTCC	Targeted: 172 bps	Genotyping of Zfp800^{em1Mgn} mice
Zfp800-Rev	GTGAAGGAGAGCAAAGACACTGG	Wild type: 188 bps	
Jazf170-D	GCTCTCGATGTAGCACCATGACAG	Targeted: 126 bps	Genotyping of Jazf1^{em1Mgn} mice
Jazf170-R	CACCGATGTGGTTGTCCTCGATG	Wild type: 134 bps	
KD2Zfp92-Fwd	AATTCCCGACCACATAACTG	Targeted: 165 bps	Genotyping of Zfp92^{em1Mgn} mice
KD2Zfp92-Rev	GCAAGAAAGTTCCAAAGCAGAGTC	Wild type: 172 bps	
Zfp329-Fwd	ATGGAGGGATTTACAAGAGAGG	Targeted: 262 bps	Genotyping of Zfp329^{em1Mgn} mice
Zfp329-Rev	GGCAACTATGTAAGGTTTGGTC	Wild type: 270 bps	
Zfhx4-Fwd	CGATGGTTGTGTTAGTGATGG	Targeted: 272 bps	Genotyping of Zfhx4^{em1Mgn} mice
Zfhx4-Rev	GGATCAGGTCCTATGAGGTTTG	Wild type: 279 bps	

Table 4.2 Oligonucleotides used in the study. This table lists oligonucleotide names, sequences, PCR product sizes and applications. Oligonucleotides were used either in the generation of ZFP knockout animals or for genotyping purposes.

Enrichment of C_2H_2-zinc finger proteins in developmental gene co-expression network.

To gain insights into pancreatic β-cell development, our lap generated a lineage-directed gene co-expression network (GCN) based on gene expression levels at various developmental stages (Osipovich et al., 2021). The network is comprised of 33 RNA-Seq datasets representing 11 distinct cell populations from E8.0 to P60 (**Fig 4.7A**). FACS sorted cells from phenotypically normal animals were taken from the endoderm at E8.0 and E8.5 ($Sox17^{GFPCre/+}$ expressing cells), the ventral foregut endoderm at E9.5 ($Sox17^{GFPCre/+}$ expressing and EPCAM-positive immunoreactive cells), the foregut endoderm at E9.5 ($Pdx1^{CFP/+}$ expressing cells), pancreatic multipotent progenitor cells at E10.5 ($Ptf1a^{YFP/+}$ expressing cells), endocrine progenitor cells at E15.5 (two different populations from $Neurog3^{GFP/+}$ or $Insm1^{GFPCre/+}$ expressing cells), and β-cells at E16.5 and P60 (Ins1 (MIP)-GFP expressing cells). To identify Neurog3- or Insm1-dependent genes, two aberrant populations from E15.5 were also taken ($Neurog3^{GFP/GFP}$ or $Insm1^{GFPCre/GFPCre}$ expressing cells) (Osipovich et al., 2021) (**Fig. 4.7A**).

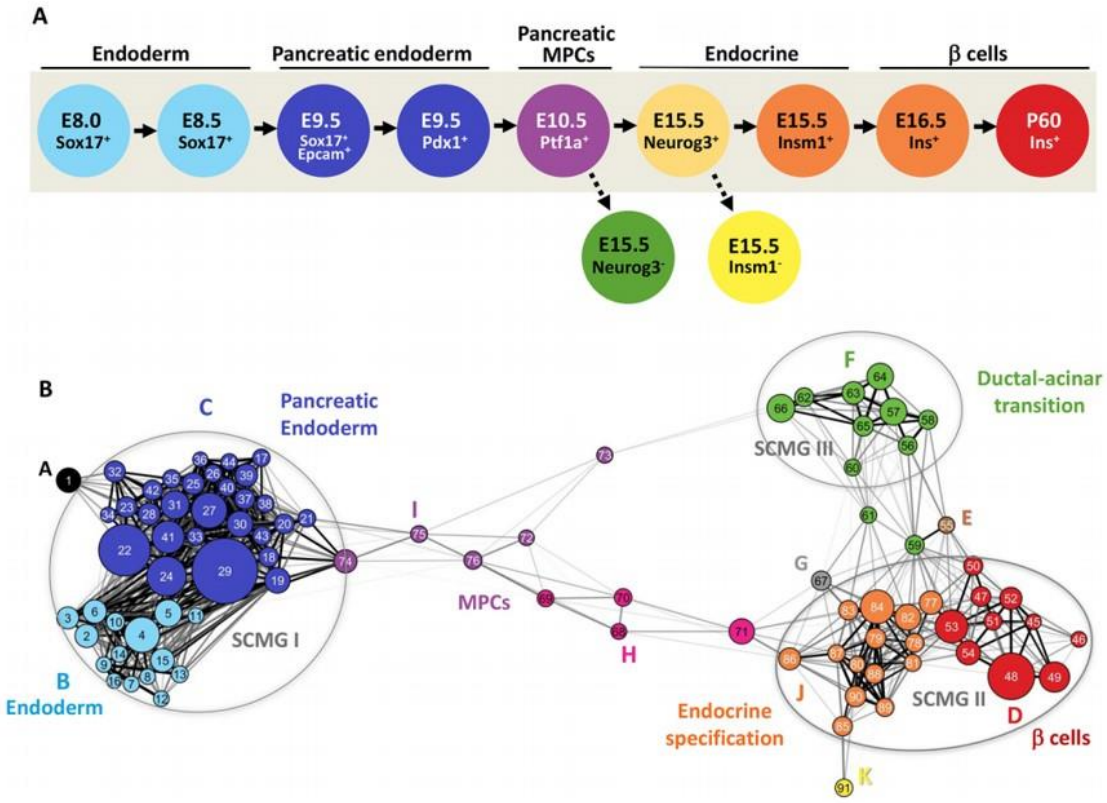


Figure 4.7. Schematic of the developmentally oriented gene co-expression network for β-cells. (**A**) A developmental scheme outlining 11 populations representing different stages of β-cell development profiled by RNA-Seq. Nine murine cell populations along the β-cell lineage and two additional mutant populations were profiled. Each profiled cell population has been shown in genetically altered mice to reflect a defined progenitor cell population that precedes the formation of mature β-cells, the populations are: gut tube endoderm (Sox17+/− at E8.5), posterior foregut endoderm (Pdx1+/− at E9.5), pancreatic multipotent progenitor cells (MPCs) (Ptf1a+/− at E10.5), endocrine progenitor cells (Neurog3+/− and Insm1+/− at E15.5), nascent β-cells (Ins+ at E16.5), and adult β-cells (Ins+ at P60). Also profiled were two mutant conditions for endocrine progenitor cells (Neurog3−/− and Insm1−/− at E15.5). (**B**) A meta-network view of GCN constructed by applying iterative WGCNA analysis on the obtained developmental RNA-Seq profiles. Each node in the meta-network represents a module of highly co-expressed genes; module numbers are indicated inside the node; node sizes are proportional to the number of genes in the module. Node color represents meta-module memberships and is coordinated with colors of developmental stages (as in A) it represents. The meta-network is defined by correlations between module eigengenes and partitions modules into three distinct strongly connected module groups (SCMGs). Edge lengths are inversely proportional to the topological overlap between modules; two modules with a high degree of topological overlap are strongly connected to the same group of modules. The strongest connections (topological overlap>0.45) are further highlighted by heavy black lines. A high-resolution version of this image is provided in Fig. S4, as is an interactive Cytoscape file for the entire network as well as individual meta-modules via https://markmagnuson.github.io/BetaCell-GCN/. This figure was modified and reprinted with permission from Osipovich, et al., 2021.

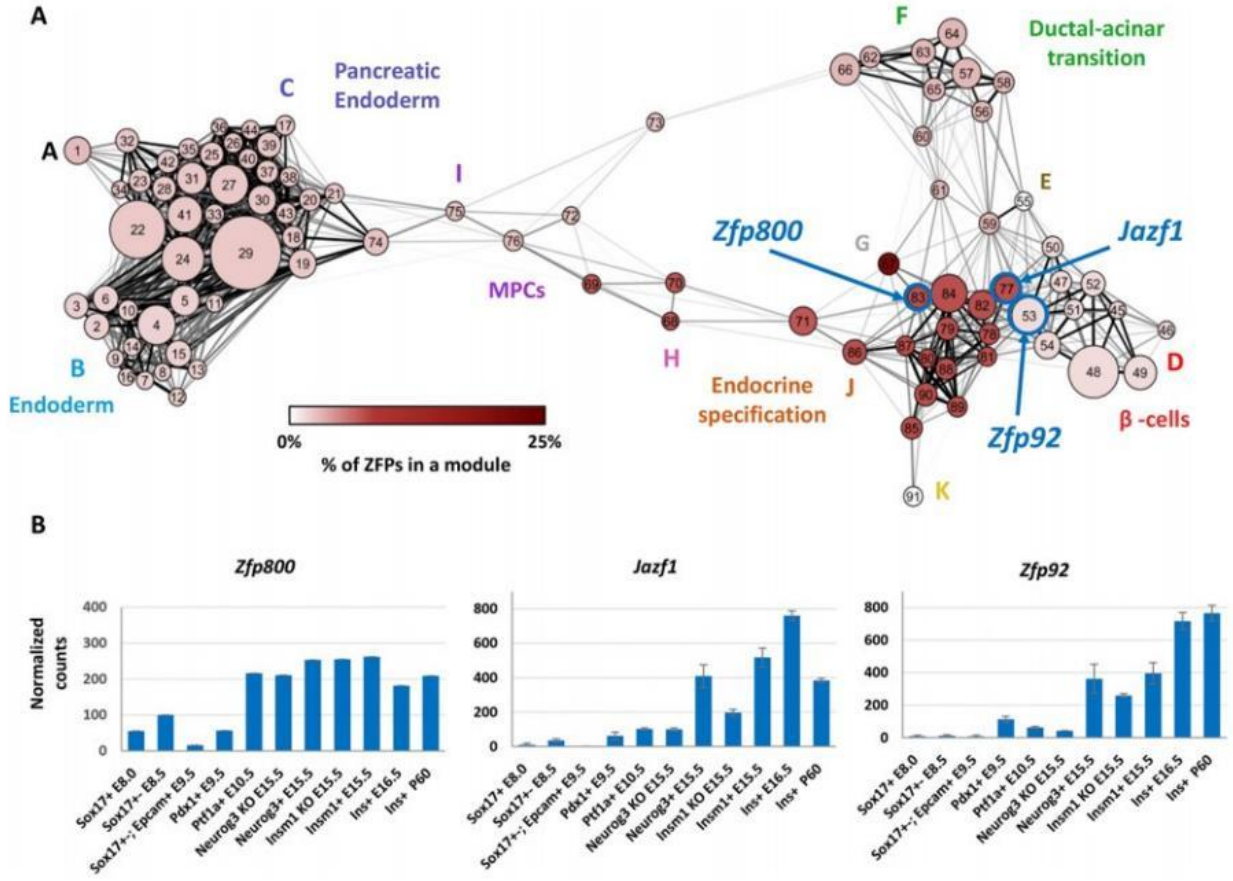


Figure 4.8. Relative proportion of C_2H_2-type zinc finger proteins in the β-cell GCN. (**A**) A meta-network view of GCN shows module nodes colored according to the proportion of ZFP-coding genes in the modules. Genes coding for ZFPs are enriched in meta-module J (endocrine specification). The C_2H_2-type ZFPs within each developmental module were identified via GO analysis. Quantified enrichments are represented as a percentage of C_2H_2 ZPF coding genes per total number of genes within a given module. Thick blue outlines indicate modules that contain the three ZFPs selected for further analysis: Zfp800, Jazf1 and Zfp92. (**B**) Expression levels of Zfp800, Jazf1 and Zfp92 in profiled developmental lineages show that Zfp800 is preferentially expressed in pancreatic multipotent and endocrine progenitor cells, Jazf1 in endocrine progenitors and nascent β-cells, and Zfp92 in nascent and mature β-cells. This figure was modified and reprinted with permission from Osipovich, et al., 2021.

Through the use of iterative WGCNA, co-expressed genes were grouped into 91 distinct modules. These modules cluster further into 11 different meta-modules by community-detection analysis and again into three strongly connected module groups (SCMGs) (**Fig. 4.7B**). Each of these modules correlates with lineage-specific factors, and the SCMGs, in particular, correspond to critical transcriptional shifts from pre-endocrine and endocrine cell types. The modules that lie outside of SCMG I and II correspond to intermediate cell stages or genes dysregulated in our knockout datasets (**Fig. 4.7B**) (Osipovich et al., 2021). While the details of information gained from these analyses are highly valuable, I will not be exploring those exhaustively for the purposes of this document. Instead, I will refer the reader to review Osipovich et al., 2021.

Importantly, the generation of this GCN revealed a previously unreported relationship of ZFP expression with specific developmental stages in the endocrine lineage. Given the enrichment of ZFPs within the GCN, as well as their known roles as developmental regulators and yet incomplete study as a group, I calculated the proportion of C_2H_2-ZFPs in each of the GRN meta-modules. I found that approximately 7% of the genes in each of meta-modules H and J, which are enriched for genes involved in endocrine cell specification, were C_2H_2-ZFPs. Nearby meta-module G contained 20.6% C_2H_2-ZFPs, compared to all the other meta-modules, which contained 0-2.8% (**Fig. 4.8**). Several previously characterized ZFPs in the network exhibited high centrality, suggesting a role in regulating transcription of many other genes, including *Casz1* and *Zhx1*, two ZFPs is known to interact with repressive transcriptional complexes (Kim et al., 2007b; Liu et al., 2015). ZFPs associated with Type 2 Diabetes (T2D) were also identified in the network, such as *Zfp174*, *Zfp445,* and *Jazf1* (Fadista et al., 2014; Osipovich et al., 2021).

Gene	Full gene name	Protein domains	Network expression profile	Module ID	Relative level of conservation
Zfp92	Zinc finger protein 92	1 KRAB, 9 C2H2		D53	High
Jazf1	Juxtaposed With Another Zinc Finger Protein 1	3 C2H2		J77	Very high
Zfp329	Zinc finger protein 329	12 C2H2		J77	Moderate
Zfp800	Zinc finger protein 800	7 C2H2		J83	Very high
Zfhx4	zinc finger homeodomain 4	23 C2H2, 4 Homeodomain, 7 U1		Unclassified	High

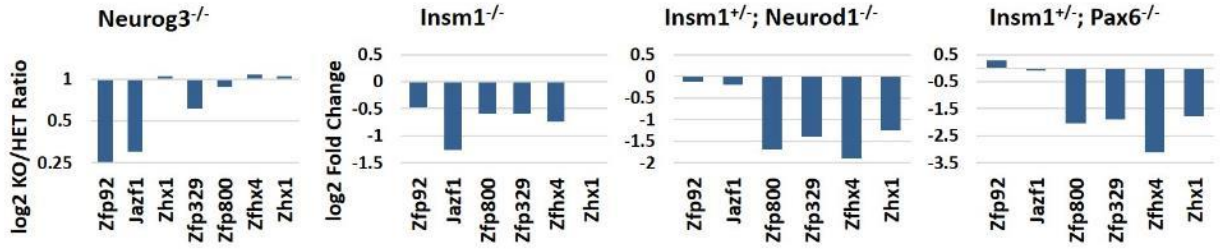


Figure 4.9. Developmental gene co-expression network traits of select zinc finger proteins and their expression changes in transcription factor knockout endocrine cells. (**A**) Table of characteristics of *Zfp92*, *Jazf1*, *Zfp800*, *Zfp329*, and *Zfhx4* within the developmental GCN. The network expression profile represents the expression level of each corresponding gene in each cell population included in the developmental network. (**B-E**) Expression changes (log_2 fold change) of each ZFP at E15.5 in animals lacking (**B**) *Neurog3*, (**C**) *Insm1*, (**D**) *Neurod1*, and (**E**) *Pax6* based on RNA-Seq data. This figure was modified and reprinted with permission from Osipovich, et al., 2021.

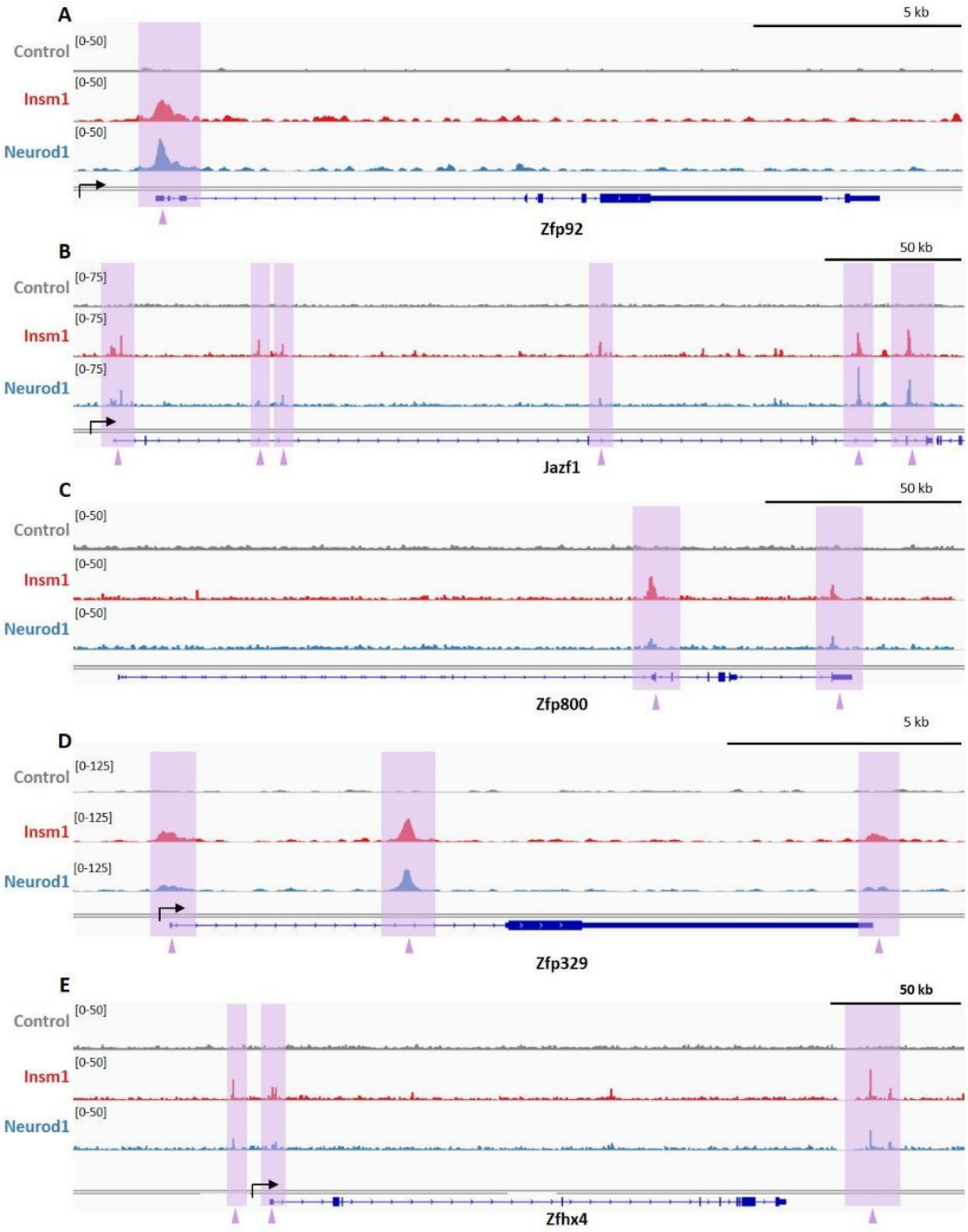


Figure 4.10. Direct binding of zinc finger proteins by Insm1 and Neurod1. Depicted are the ChIP-Seq coverage profiles for Insm1 and Neurod1 at (**A**) *Zfp92*, (**B**) *Jazf1*, (**C**) *Zfp800*, (**D**) *Zfp329*, and (**E**) *Zfhx4*. Strong transcription factor binding by both Insm1 and Neurod1 can be detected around the transcriptional start sites and some internal regions of each of the genes. This data comes from the same ChIP-Seq analysis performed in Chapter II. Insm1 binding profiles are in red, Neurod1 binding in blue. Regions with major peaks for both TFs are highlighted in purple and indicated by an arrow marker.

Perturbation of zinc finger proteins during development.

To determine the role of ZFPs in pancreas development and function, I selected five genes that encode zinc finger proteins (*Zfp92*, *Jazf1*, *Zfp800*, *Zfp329*, and *Zfhx4*) for further investigation. These genes were selected based on their expression patterns and centrality measures within the developmental network, regulation by *Neurog3*, regulation by *Insm1*, *Neurod1* and/or *Pax6*, and level of protein conservation across species (**Fig. 4.9**). Each of these genes is moderate to highly conserved across evolution, regulated by at least one of the previously mentioned TFs, and contains multiple C_2H_2 domains. Additionally, *Jazf1* and *Zfp800* have both been shown to be bound by, or near binding sites of, endocrine-specific TFs (i.e., Pdx1, Foxa2, Onecut1, Nkx2.2, Nkx6.1, Insm1, Neurod1) during development and in mature cells. Referring to the ChIP-Seq analysis described in Chapter II, I found that these ZFPs are specifically bound by Insm1 and Neurod1 (**Fig. 4.10**). These two genes also contain T2D-associated SNPs (**Table 4.3**). To study their function in mice, I utilized CRISPR-Cas9 technology to target each of these genes and induce INDEL formations and generate global knockout (KO) alleles for each ZFP (**Fig. 4.11**). Following the injection of target-specific guide RNAs into Cas9-expressing embryos, resulting founder animals were screened and their alleles characterized. Null alleles containing a frameshift mutation within the first protein-coding exon of the gene were successfully generated for each ZFP.

Zinc finger protein 92 (*Zfp92*) contains 9 C_2H_2 domains and one repressive KRAB domain and is highly expressed in mature β-cells (meta-module D). Mutagenesis of Zfp92 resulted in a 7 bp deletion of the third exon of Zfp92; a frameshift mutation predicted to impair protein function (**Fig. 4.12A**). Mice lacking Zfp92 have a mild growth phenotype, with male animals having a slightly reduced body weight. However, based on intraperitoneal glucose tolerance tests (IPGTT) at 12 weeks, the loss of *Zfp92* does not appear to have a measurable impact on β-cell function (data not shown). Similarly, *Zinc finger protein 329* (*Zfp329*) contains 12 C_2H_2 domains and is a member of meta-module J. Conversely, *Zfhx4* (*Zinc finger homeobox protein 4*), is a much more complex protein with 23 C_2H_2 domains, 4 homeodomains, and 7 U1 domains that was the only ZFP discussed here that remained unclassified in the developmental network (**Fig. 4.12E**). The null alleles for both genes contain an 8 bp deletion and result in a truncated, non-functional protein (**Fig. 4.12B-C**). Both *Zfp329* and *Zfhx4* KO animals were viable, fertile, and demonstrated no changes in body weight or impaired glucose tolerance (data not shown, SEM). Thus, I focused our efforts on further characterizing *Jazf1* and *Zfp800*.

Gene Loci	Condition Correlation	SNP	Reference: PMID (Year)
Jazf1	T2D	rs1708302	29358691 (2018)
	T2D	rs849134	27189021 (2016)
	T2D	rs864745	18372903 (2008)
	Height	rs1635852	18952825 (2008)
	BMI-adjusted waist circumference	rs849140	25673412 (2015)
Zfp800	T2D/CHD	rs17867832	24729826 (2014)
	Resting heart rate	rs11563648	27798624 (2016)
	Neuroticism, Intelligence	rs11766965	29942085 (2018)
	Heel bone mineral density	rs62621812, rs534049724	30048462 (2018)

Table 4.3 Disease associated single nucleotide polymorphisms located at or near *Jazf1* and *Zfp800*. Specific SNP identifiers, the associated condition, and corresponding references are provided.

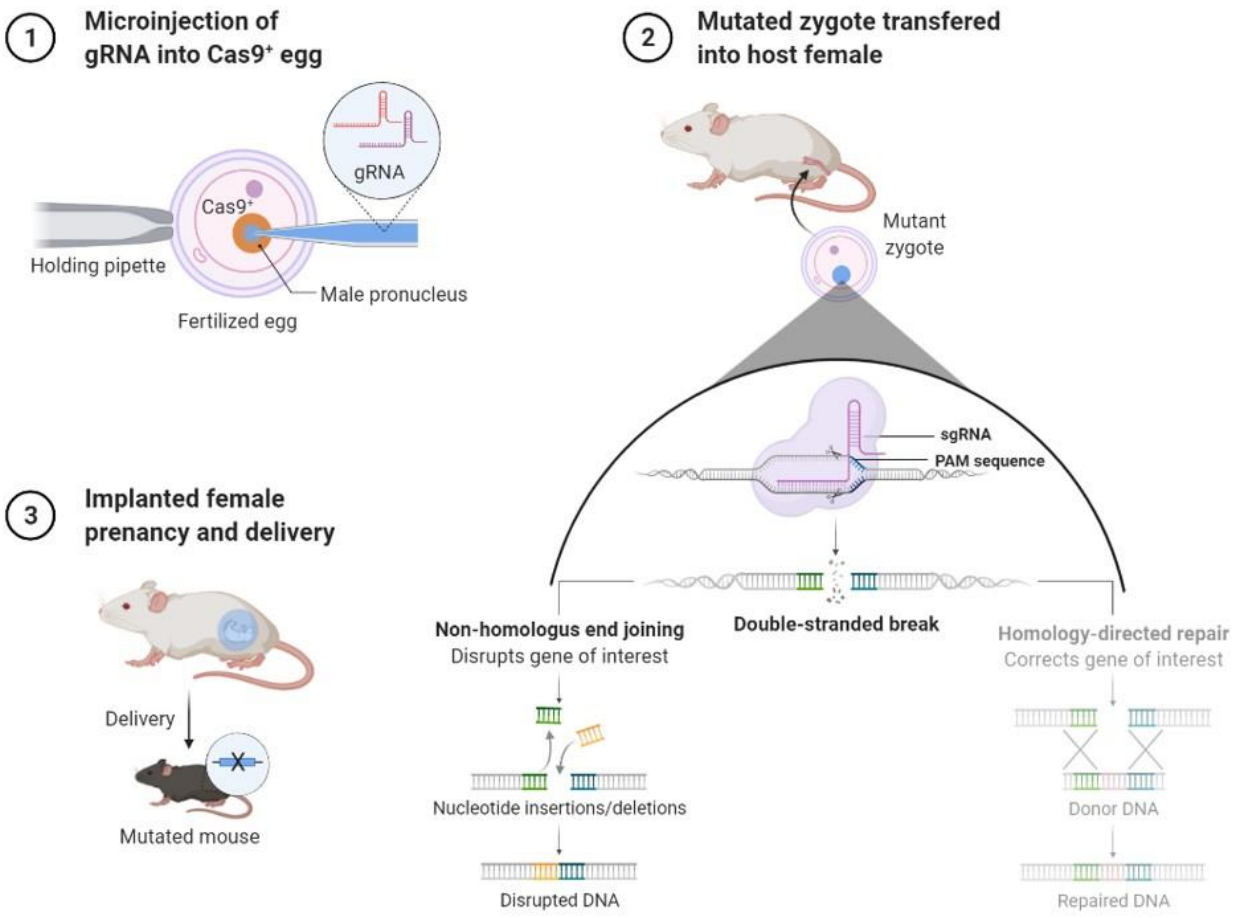


Figure 4.11. CRISPR/Cas9-mediated generation of zinc finger protein knockout mice. Schematic overview of the methods used for generating ZFP knockouts. Target-specific guide RNA (gRNA) complexes were micro-injected into Cas9-expressing blastocysts obtained from the mating of homozygous B6J/129(Cg)- Gt(ROSA)26Sor$^{tm1.1(CAG-cas9*,-EGFP)Fezh/J}$ mice (JAX 026179) with C57BL/6N mice. Mutations in the ZFP targets were the result of an induced double-stranded break by Cas9 and the subsequent repair by the innate cellular machinery via non-homologous end joining (NHEJ) mechanism. The imperfect repair process introduced INDEL mutations that generated a frame shift in the genetic code and a non-functioning protein. The injected zygotes were transferred to super-ovulated female hosts for implantation. The resulting pups were screened, and the individual mutations were characterized. This is an original work generated using BioRender.com.

Juxtaposed with another zinc finger protein 1, *Jazf1* (also known as *Tip27*), contains three C_2H_2 domains and has been implicated in many disease states. It contains several SNPs closely associated not only with T2D but also with height and BMI (**Table 4.3**). Interestingly, *Jazf1* is also known to translocate with *Suz12*, contributing to the development of cancer in humans (Hrzenjak, 2016). Past studies have suggested a protective role for *Jazf1* against diabetes, and levels are reduced in the pancreatic islet of T2D patients (Taneera et al., 2012). Our *Jazf1* knockout animals have an 8 bp deletion within the first exon, and while viable and fertile, they display a pronounced growth phenotype (**Fig. 4.13A and 4.14**). Animals lacking *Jazf1* are smaller at the time of weaning, and their growth remains stunted relative to their wild-type littermates (**Fig. 4.14A**). To determine if this growth retardation phenotype was correlated with changes in lean and fat body mass, I performed body composition measurements of adults at 12 weeks of age. While there was a trend for KO males to have a decrease in the percentage of body fat mass and an increase in lean body mass, those numbers did not reach statistical significance (**Fig. 4.14B-C**). Since insulin-like growth factor 1 (IGF-1) and growth hormone (GH) play important roles in growth, I also looked at plasma IGF-1 and GH levels using a Luminex assay. I did observe a decrease in IGF-1 levels; however, there was no difference in the level of plasma GH (**Fig. 4.14E-F**). To determine if Jazf1 is important for β-cell function, I performed IPGT tests and found only a slight impairment in their response to glucose at 30- and 60-minutes post-injection (**Fig. 4.14D**).

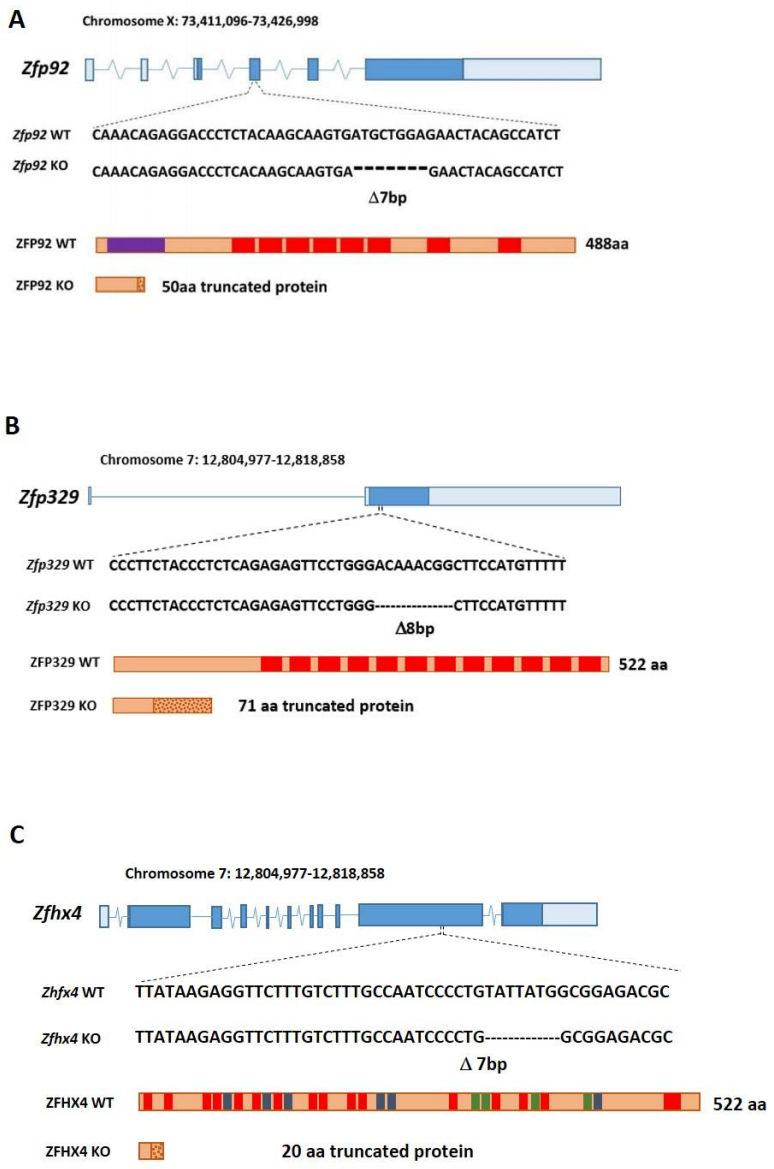


Figure 4.12. Characterization of *Zfp92*, *Zfp329* and *Zfhx4* knockout alleles. (**A**) Schematic representation of **Zfp92** gene and relative gene location and sequence of CRISPR/Cas9 generated 7 bp deletion. The deletion of 7bp leads to a frameshift that abrogates normal **ZFP92** 488 amino acid (aa) protein translation and leads to translation of a truncated 50 aa peptide with 8 missense amino acids at the end. (**B**) Representative image of a PCR genotyping gel for **Zfp92** wild type (WT), heterozygous (Het) and knockout (KO) mice. (**C**) Schematic representation of **Zfp329** gene and relative gene location and sequence of CRISPR/Cas9 generated 7bp deletion. The deletion of 7 bp leads to a frameshift that abrogates normal **ZFP329** 488 aa protein translation and leads to translation of a truncated 50 aa peptide with 8 missense aa at the end. (**D**) Representative image of a PCR genotyping gel for **Zfp329** wild type (WT), heterozygous (Het) and knockout (KO) mice. (**E**) Schematic representation of **Zfxh4** gene and relative gene location and sequence of CRISPR/Cas9 generated 7bp deletion. The deletion of 7 bp leads to a frameshift that abrogates normal **ZFhx4** 488 aa protein translation and leads to translation of a truncated 50 aa peptide with 8 missense aa at the end. (**F**) Representative image of a PCR genotyping gel for **Zfhx4** wild type (WT), heterozygous (Het) and knockout (KO) mice. This figure was modified and reprinted with permission from Osipovich, et al., 2021.

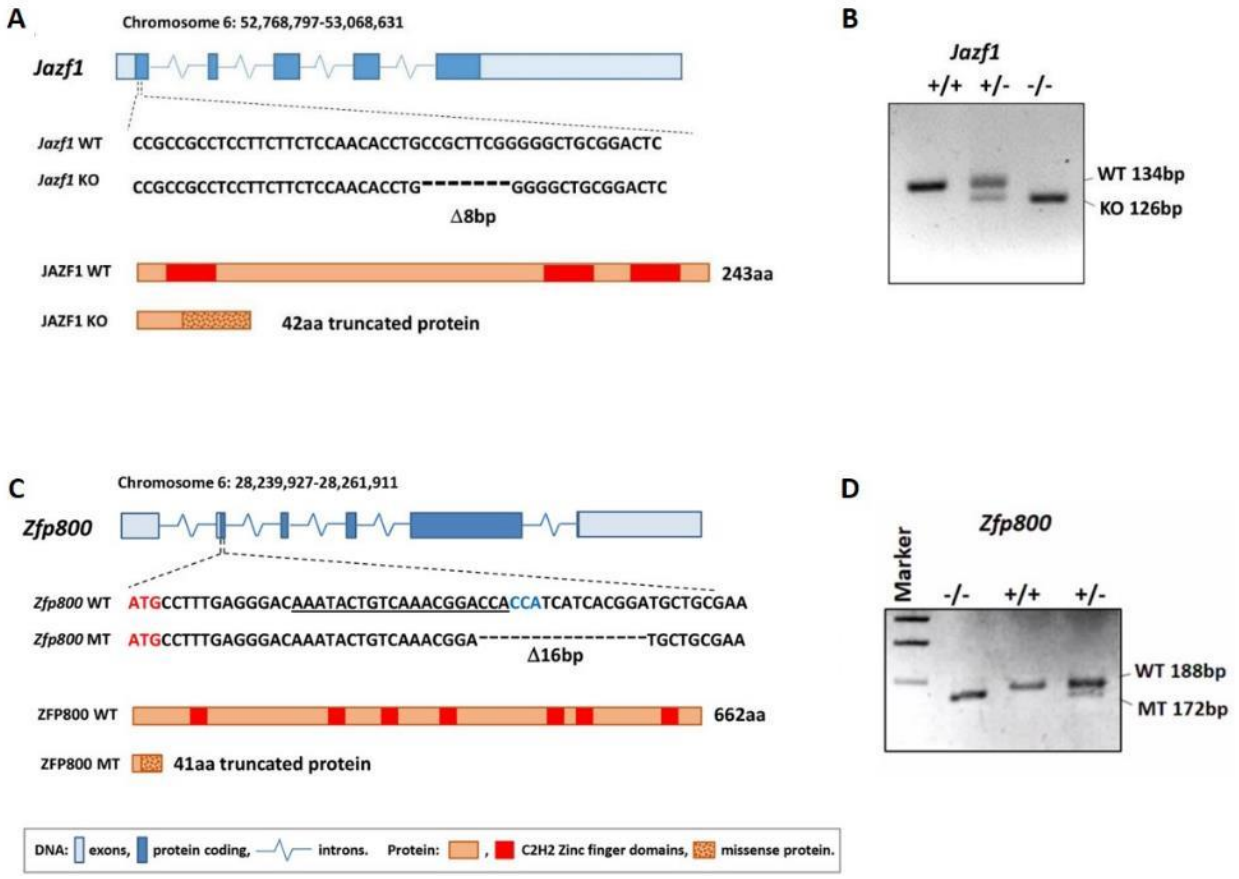
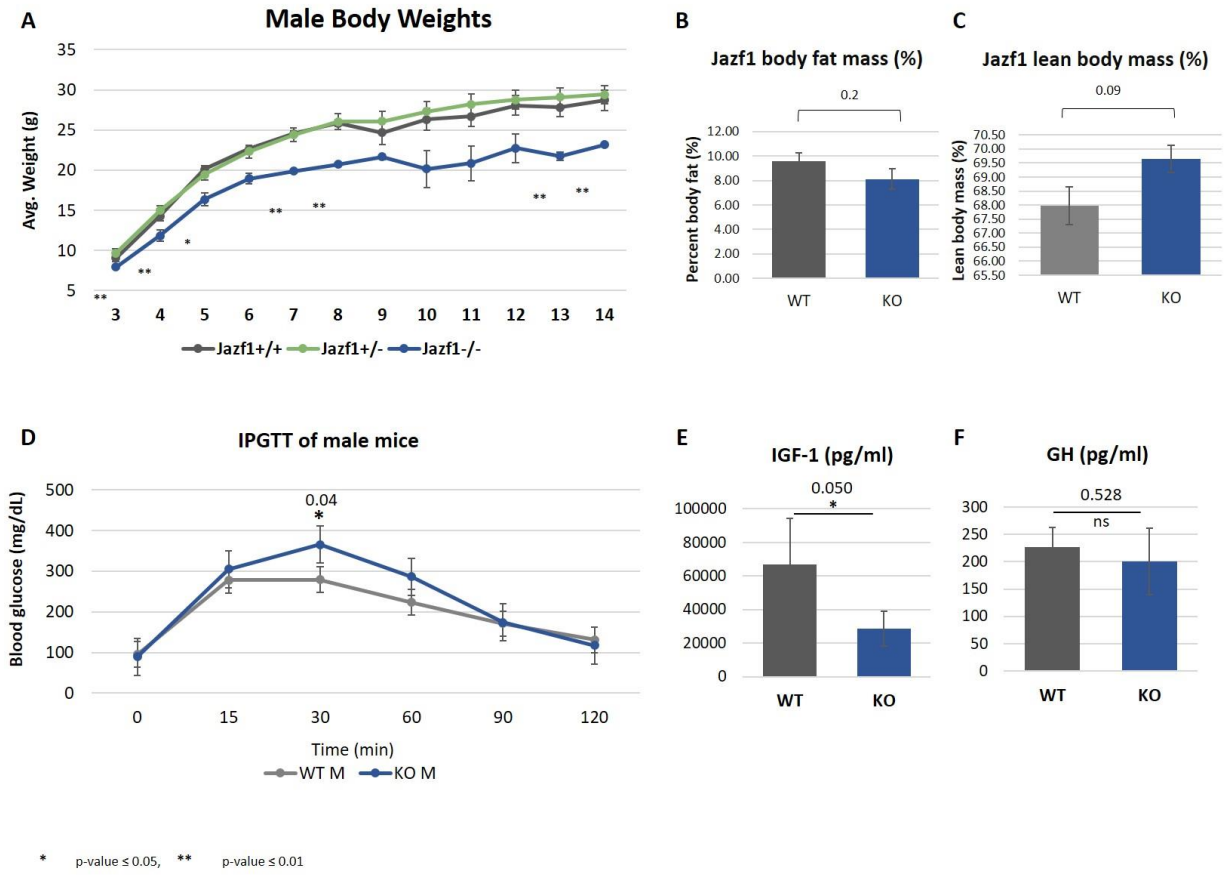


Figure 4.13. Characterization of *Jazf1* and *Zfp800* knockout alleles. (**A**) Schematic representation of ***Jazf1*** gene and relative gene location and sequence of CRISPR/Cas9 generated 8bp deletion. The deletion of 8 bp generates a frameshift that abrogates normal **JAZF1** 243 aa protein translation and leads to translation of a truncated 42aa peptide with 28 missense amino acids. (**B**) Representative image of a PCR genotyping gel for **Jazf1** wild type (WT), heterozygous (Het) and knockout (KO) mice. (**C**) Schematic representation of **Zfp800** gene and relative gene location and sequence of CRISPR/Cas9 generated 7bp deletion. The deletion of 7 bp leads to a frameshift that abrogates normal **ZFP800** 488 aa protein translation and leads to translation of a truncated 50 aa peptide with 8 missense amino acids at the end. (**D**) Representative image of a PCR genotyping gel for **Zfp800** wild type (WT), heterozygous (Het) and knockout (KO) mice. This figure was modified and reprinted with permission from Osipovich, et al., 2021.



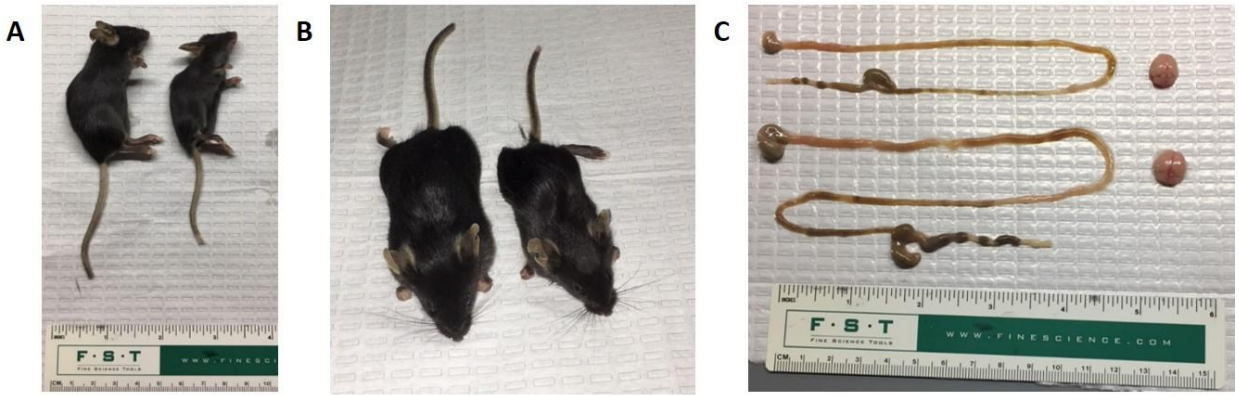
* p-value ≤ 0.05, ** p-value ≤ 0.01

Figure 4.14. *Jazf1* knockout animals exhibit growth defects. (**A**) Body weights of male wild type, heterozygous and null for *Jazf1* taken at weekly intervals. (**B-C**) Body composition measurements for wildtype and *Jazf1* KO animals at 14 weeks of age. (**B**) Body fat mass and (**C**) lean body mass are presented as percentages of total body weight. (**D**) Glucose tolerance test measurements of *Jazf1* wild type and knockout mice at 14 weeks of age. Plasma protein levels of (**E**) IGF-1 and (**F**) GH of *Jazf1* wild type and knockout animals determined by Luminex assays.

Zfp800 knockout mice exhibit perturbations in pancreas development.

Of the five zinc finger protein genes that I analyzed; *zinc finger protein 800* (*Zfp800*) KO animals displayed the most robust phenotype. Much like *Jazf1* or *Zfp329*, *Zfp800* contains an array of C_2H_2 domains, 7 in total (**Figs. 4.10 and 4.13C**). The null allele generated by CRISPR-Cas9 mutagenesis also results in the translation of a truncated protein due to a 16 bp frameshift deletion near the start codon (**Fig. 4.13C**). While heterozygous animals appear to be normal and had no impairments in weight or fertility, *Zfp800* KO animals were not viable and died within a few days after birth (**Fig. 4.15A-E**). To determine if perturbations in the pancreas contribute to the early death of KO pups, we performed immunohistochemical and bulk RNA-Seq analyses on whole pancreata at E18.5. Zfp800 KO embryos exhibited impaired endocrine cell development marked by a reduction in the number of *Ins*-, *Gcg*-, *Sst*-, and *Ppy*-expressing cells (**Fig. 4.16A-D**). Furthermore, acinar cells also showed developmental defects in the organization of rosettes and cellular polarity, as demonstrated by amylase and E-cadherin staining (**Fig. 4.16E-F**).

Likewise, bulk RNA-Seq analysis at E18.5 revealed the dysregulated expression of several genes, with 976 genes downregulated (DRGs) and 455 genes upregulated (URGs) (log_2FC |1|, p-adj < 0.05) (**Fig. 4.17A**). Functional enrichment analysis showed that downregulated genes are involved in oxidative phosphorylation, translation, rRNA processing, and pancreatic and hormone secretion, whereas upregulated genes are involved in cell projection morphogenesis, actin cytoskeleton, and heart, brain, and muscle development (**Fig. 4.17B**). Among the top downregulated genes are pancreatic hormones (*Ins1, Ins2, Gcg*), acinar genes (*Prss3, Cela1*), and TFs important for pancreas development (*Id3, Arx, Ptf1a, Isl1*) (**Fig. 4.17C**). At the same time, among the top upregulated genes were genes coding for other developmental TFs (*Hoxc8, Runx3, Nfia*), cytoskeletal components (*Xirp1, Map1b*), and developmental signaling proteins (*Amer2, Notch2*) (**Fig. 4.17C**). The decrease in expression of pancreatic hormones and TFs in *Zfp800* knockout pancreata was confirmed by RT-qPCR (**Fig. 4.17D**). Combined, our data show that *Zfp800* is required for pancreas development, and its disruption leads to multiple alterations in both tissue histology and gene expression. Although we cannot dismiss the possibility that premature death of KO animals is caused by disruptions in other tissues, it is clear that *Zfp800* plays a role in pancreas development.



D

Genotype	Expected	Observed
WT	8.5	9
Het	17	25
KO	8.5	0

X^2=12.29

p-value=0.0021

E

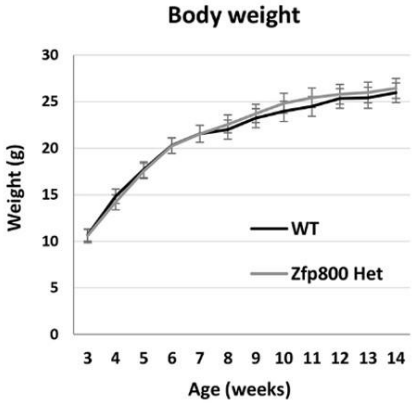


Figure 4.15. *Zfp800* knockout mice exhibit a failure to thrive and are lethal during post-natal stages. (**A-C and E**) *Zfp800* knockout mice have a stunted growth and reduced body size. Representative images comparing the size difference in (**A, B**) whole body, (**C**) intestinal tract, and brain of wildtype and *Zfp800* KO animals. (**D**) Chi-squared analysis of expected Mendelian ratios at weaning was significant. (**E**) Heterozygous animals show no differences in size or growth rates.

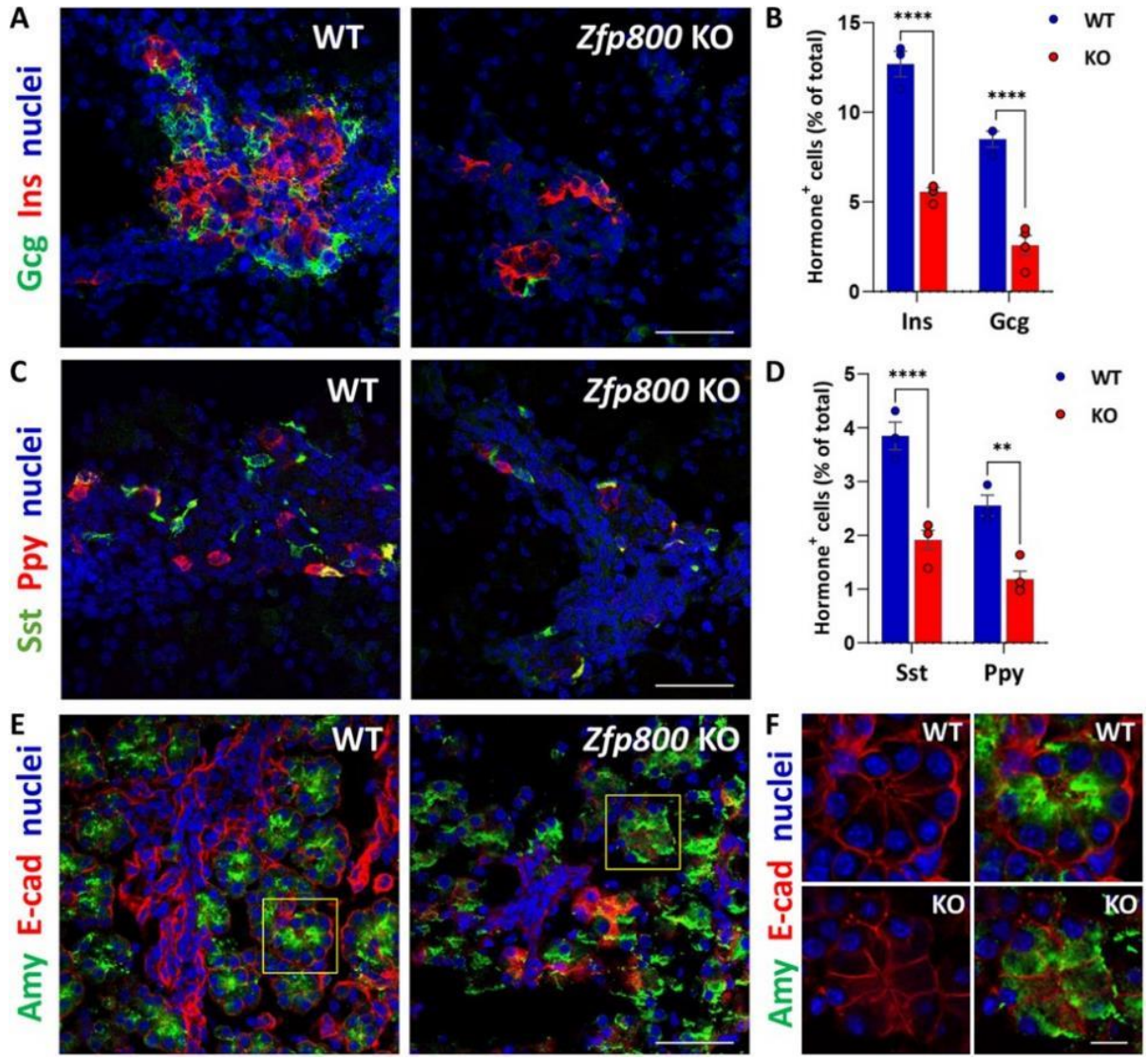


Figure 4.16. *Zfp800* knockout mice have defects in pancreatic endocrine and acinar tissue development. (**A-D**) Immunofluorescence staining (**A,C**) and quantification (**B,D**) of E18.5 pancreatic sections from *Zfp800* knockout (KO) and wild-type (WT) mice for pancreatic endocrine hormones insulin (Ins, red) and glucagon (Gcg, green) (**A,B**), and somatostatin (Sst, green) and pancreatic polypeptide (Ppy, red) expressing cells (**C,D**) show a reduction in the number of hormone-positive cells in the *Zfp800* KO. *n*=3, error bars represent s.e.m.; ****$P \leq 0.0001$; **$P \leq 0.01$ (one-way ANOVA). (**E,F**) Immunofluorescence staining with the acinar cell enzyme amylase (Amy, green) and cell membrane marker E-cadherin (E-Cad, red) of E18.5 pancreatic sections from KO and WT mice demonstrates defects in acinar tissue formation in the absence of *Zfp800*. Amylase is localized near the apical regions of the cells in the center of the acinus in WT tissues, whereas KO cells are disorganized, with amylase staining throughout. Yellow squares show areas that are enlarged in **F**. Cell nuclei are stained with To-Pro-3 (blue). Scale bars: 100 µm (A,C,E); 10 µm (**F**). This figure was modified and reprinted with permission from Osipovich, et al., 2021.

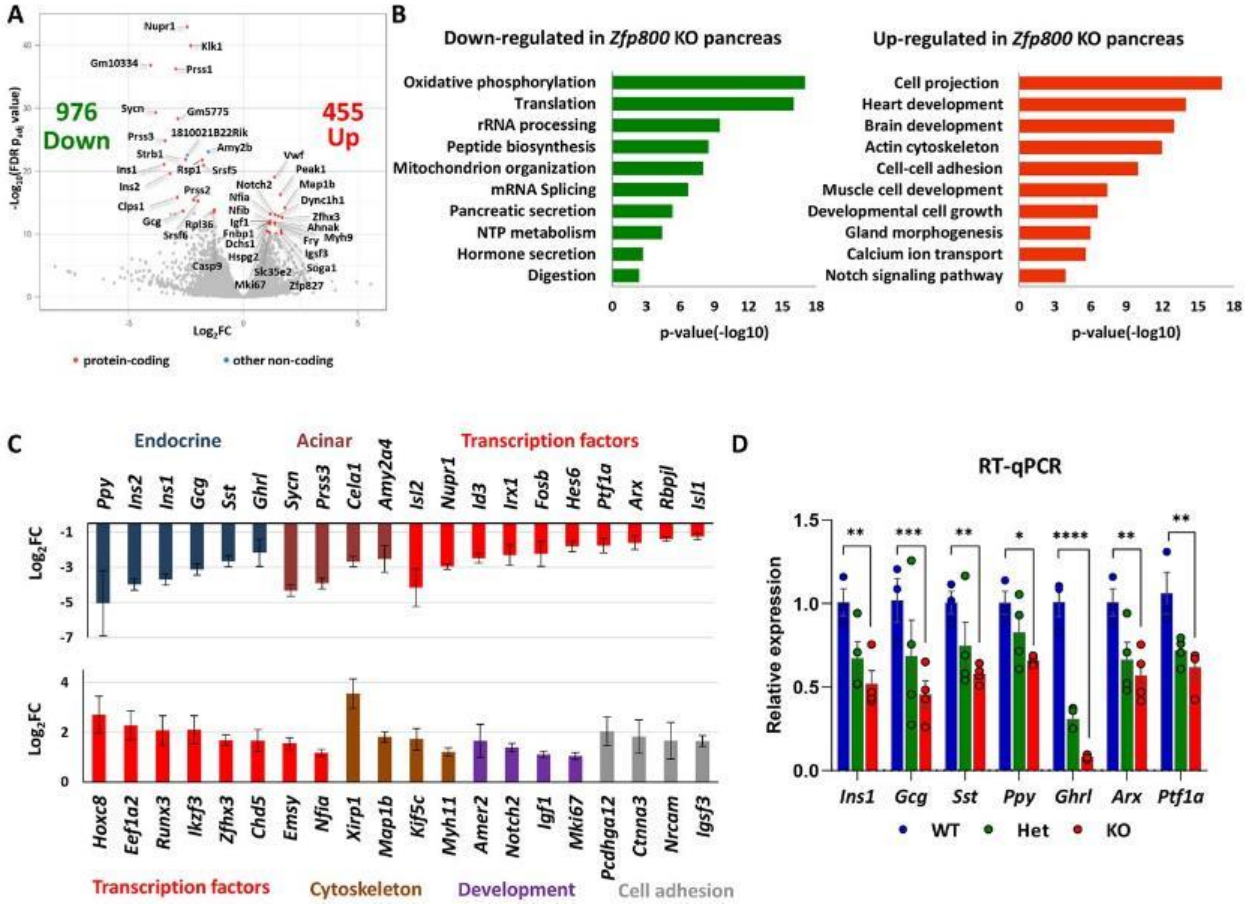


Figure 4.17. RNA-Seq analysis of pancreatic tissues from *Zfp800* knockout mice. (**A**) Volcano plot showing differentially expressed genes (log_2FC over *P*-value) in *Zfp800* KO versus WT pancreata from E18.5 embryos. The top ten differentially expressed genes (based on P_{adj} value) are indicated by names and total numbers of downregulated (Down) and upregulated (Up) genes are provided ($log_2FC<-1$ or >1, $P_{adj}<0.05$). (**B**) Functional enrichment analysis of downregulated and upregulated genes. Select top enriched pathways are shown. (**C**) Differential expression of select top downregulated (top graph) and upregulated (bottom graph) genes. Colors indicate gene functional associations. Log_2FC: log_2 fold change *Zfp800* KO versus WT ($P_{adj}<0.05$), error bars represent s.e.m. (Log_2FC) determined by DEseq. (**D**) RT-qPCR analysis of mRNA expression for select pancreatic genes in pancreatic RNA samples collected at E18.5. *Zfp800* wild-type (WT), heterozygous (Het) and knockout (KO) samples. *n*=3-4, error bars represent s.e.m. ****$P\leq0.0001$; ***$P\leq0.001$; **$P\leq0.01$; *$P\leq0.05$ (two-way ANOVA). This figure was modified and reprinted with permission from Osipovich, et al., 2021.

DISCUSSION

Elucidating the role of zinc finger proteins in pancreas development.

Based on their co-expression patterns in the developmental co-expression network, associations with T2D, and level of conservation across evolution, I investigated the roles of *Zfp92*, *Jazf1*, *Zfp800*, *Zfp329,* and *Zfhx4*. All these factors have varying degrees of protein structure complexity, but each of them contains an array of C_2H_2-type zinc finger domains. *Zfp329* and *Zfp800* both become expressed at ~E10.5 in multipotent progenitor cells and maintain expression in adult β-cells. *Zfp92* and *Jazf1*, on the other hand, are expressed later at E15.5 and are maintained through adult stages. Still, *Zfhx4* was in a class of its own, with its expression beginning primarily at E9.5, peaking at E15.5-E16.5, and diminishing in later developmental stages, as determined by our gene co-expression network data.

While the loss of *Zfp329* and *Zfhx4* did not reveal any immediate consequences, and the absence of *Zfp92* presented only a mild phenotype, phenotypes caused by the loss of *Jazf1* and *Zfp800* were apparent. Our observations of *Jazf1* KO mice had stunted growth and mild impairments in plasma IGF-1 levels and glucose tolerance, indicating a more nuanced role for this ZFP. Based on the expression profile of *Jazf1* and its linkages to T2D, it is likely that this factor is more important for later developmental stages and mature cells. Indeed, while these experiments were ongoing, Kobiita et al., 2020 reported the generation and characterization of a different *Jazf1* KO in β-cells. Based on genetic deletion, biochemical and physiological analyses in the context of the highly diabetogenic Akita mutation, they found that *Jazf1* functions to protect against ER stress-induced apoptosis of endocrine cells, consistent with earlier studies showing its protective qualities in stressed hepatocytes for preventing lipogenesis and systemic inflammation-related disease (Kobiita et al., 2020; Yang et al., 2014).

In contrast, *Zfp800* expression is turned on earlier in development (~E10.5) and is maintained in adult cells. Supporting a role for *Zfp800* in pancreas development, KO mice are non-viable and show robust defects in islet morphology, acinar cell polarity, and a dysregulated transcriptome. Consistent with impaired islet morphology and endocrine cell types, the expression of *Arx* and *Isl1* are perturbed in *Zfp800* KO pancreata. This is also true for *Ptf1a* expression, which is critical for the specification and maturation of acinar cells (Kawaguchi et al., 2002). Much

like *Jazf1*, several SNPs associated with T2D can be found near the human *Znf800* locus (Mahajan et al., 2018). Combined with the results presented here and in Osipovich et al., 2021, this could suggest a role not only for pancreas development but endocrine cell function and identity as well. Additional analysis is necessary for establishing a deeper understanding of the function of *Zfp800* in endocrine cells.

Taken together, further investigation of zinc finger proteins is necessary to move our understanding of pancreas development and function forward. Even in this small sampling of five select ZFPs, I have identified functional roles for three, each having varying degrees of phenotypic severity. It is certain that of the hundreds yet to be characterized, many will have a role in the formation and function of pancreatic endocrine cells.

ACKNOWLEDGEMENTS

I thank A. Osipovich for help in performing immunohistochemical and RNA-Seq analysis of *Zfp800* knockout animals, and A. Chapman and J. Coeur for assistance with data collection. I also thank the Vanderbilt Genomic Editing Resource Core (VGER) for their help in generating ZFP KO alleles via CRISPR-mediated mutagenesis.

CHAPTER V – SUMMARY AND FUTURE DIRECTIONS

In this study, I sought to further investigate the role of three important *Neurog3*-dependent TFs. Using generated knockout mice for *Insm1*, *Neurod1*, and *Pax6* alleles, I quantified the morphological defects on endocrine cell development during embryonic stages by immunohistochemistry. We saw that all three KO models had defects in islet architecture, a reduced number of endocrine cells, and altered hormone expressing cell ratios. We also performed bulk RNA-Seq and differential splicing analysis of FACS purified KO endocrine cells and a reanalysis of Insm1 and Neurod1 DNA binding data in beta cells. Our results indicated that each TF regulated the expression of a unique subset of genes, as well as a large proportion of overlapping genes. Alternative splicing (AS) analysis revealed a novel role for Insm1, Neurod1, and Pax6 in regulating RBP splicing factors and alternative splicing events during endocrine cell development. This included AS of endocrine-specific TFs (*Pax4, Pax6, Tcf7l1*) and important functional genes (*Ins1, Insr, Gck*). By exploring the biological role of a subset of zinc finger proteins (ZFPs) that are transcriptionally regulated by one or more TFs (*Neurog3, Insm1, Neurod1,* or *Pax6*), we found that *Zfp92* and *Jazf1* KO animals have mild phenotypes more pronounced in mature animals, while the *Zfp800* KO was embryonically lethal and marked by perturbations in islet architecture and hormone expression. These data indicate that ZFPs do play a role in pancreas development and require further characterization of their molecular mechanisms.

Experimental caveats

While my studies have generated a wealth of data and information, there are a few caveats related to my experimental designs that are worth reflecting on. One of the most obvious is in my use of the *Insm1*$^{GFP.Cre}$ reporter allele (Osipovich et al., 2014). This allele was instrumental in being able to FACS sort out endocrine cells during the developmental stages to have an enriched cell population for RNA-Seq analysis. This prevented contamination from other cell types in the pancreas (i.e., acinar and ductal cells) and provided cleaner, more specific RNA-Seq data. However, this also introduced a potential confounding effect that I was unable to fully account for. A recent report by Tao et al. demonstrated that adult β-cells heterozygous for Insm1 have a haplo-insufficient phenotype and impaired cellular function (Tao et al., 2018). While the authors do not report observations of defects at developmental time points, we cannot eliminate the possibility that the reduced concentrations of *Insm1* in the *Neurod1* or *Pax6* KO models led to synthetic effects. Our morphological data largely mirrors that of previously

published reports, but we do observe a robust increase in apoptosis and pre-mature embryonic death in the *Pax6* KO animals. It would be interesting to look more closely at animals heterozygous for *Insm1* relative to wild type at the transcriptional level to determine if there are subtle changes in gene expression patterns that were not captured in our data. Additionally, to avoid this possible confound effect altogether, it would be possible to make and use an *Insm1* BAC transgenic model to provide access to a purified cell population and simultaneously maintain proper levels of *Insm1* expression.

Another experimental caveat includes my use of the mm10 reference genome that is based on the C57BL/6J background in our differentially splicing analysis. Since our animals were maintained in a CD-1 background, this could have potentially introduced missed splicing events. This is because there are subtle differences in splicing patterns across different genetic strains of mice, as well as biases in reference genome alignment approaches (Raghupathy et al., 2018). In the future it would be ideal to use reference genomes from the exact mouse strain used for the alignment of sequencing reads and differential splicing analysis to limit any biases from the data.

Expanding our fundamental understanding of biological mechanisms driving development and disease.

The fundamental knowledge gained from the study of pancreas development guides our understanding of how individuals may be predisposed to disease, including adult-onset, and how we can apply the information gained from basic studies to treat those diseases. Future studies directed at elucidating the role of *Insm1*, *Neurod1*, and *Pax6* in finer detail will be of great benefit to the field. Specifically, it will be important to determine how these TFs function within the GRN to establish and maintain the identity and function of islet cells. Gathering this information can be achieved in a few ways. It would be possible to create additional, conditional knockout models for other transcription factors, such as those described here, and perform RNA-Seq experiments to determine transcriptional targets. It would provide new insights into how well-connected a given TF is within the network. For example, it would be expected that the loss of highly connected TFs would result in the dysregulation of hundreds, or possibly thousands, of genes, while the loss of less well-connected TFs may only disrupt the expression of dozens. This information can be manually or computationally integrated to form causal biochemical reactions that drive cellular behaviors (Emmert-Streib et al., 2014).

The triangulation of evidence outlining the directional relationships of the targets of these and other TFs during development through the integration of transcriptional data, ChIP-Seq data, alternative splicing analysis, proteomics, and other methodologies will be important. While we have begun to take that approach here, additional experiments (e.g., ChIP-Seq, ATAC-Seq, mass spectrometry, etc.) performed at the same time point are necessary to provide confidence to what is known and to add to the gaps that remain (**Fig. 6.1**) (Sonawane et al., 2019). Since TFs also function through interactions with other co-regulators, it will be interesting to identify the factors that directly interact with each of these TFs to fine-tune the regulation of gene expression. For example, previous studies have shown that *Insm1*, *Foxa2*, and *Neurod1* can cooperatively regulate subsets of genes in adult beta-cells (Jia et al., 2015). Additionally, learning how those or other regulatory interactions drive key developmental processes is critical. For example, we know that having an adequate number of endocrine cells is essential for proper development and for function in adult animals. The loss of *Insm1*, *Neurod1*, or *Pax6* results in an overall reduction of endocrine cells. It has also been shown that tight control of the cell cycle is key for the temporal regulation of proliferation in pro-endocrine cells. Are these factors critical for governing the spatiotemporal proliferation needs of pro-endocrine cells to produce enough cells? Do these factors carry out other *Neurog3*-dependent functions necessary for endocrine cell differentiation? Furthermore, the precise transcriptional signals established at defined developmental stages that are important for the competency of a given cell to commit to a given lineage remains elusive. Answering these questions and characterizing the lineage-specific GRN of endocrine cells, will enable scientists to continue making progress in successfully differentiating cells into mature, functioning beta cells.

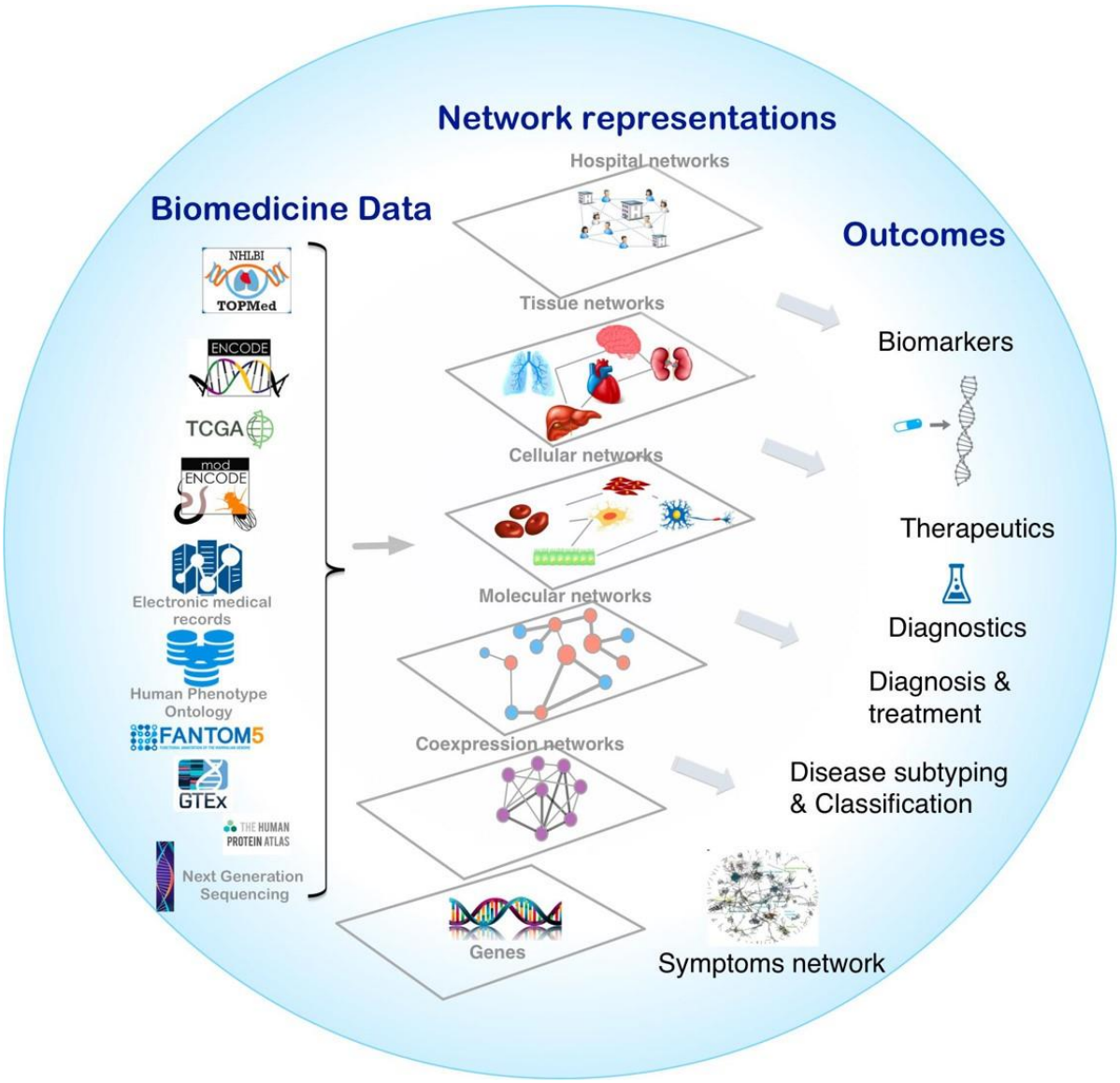


Figure 6.1. Regulatory networks inputs and outputs. Overview of network medicine approach depicting various biomedical data types discussed at length in the manuscript, along with network representations that simplify different components of multiple omics data from the genome, transcriptome, proteome, and metabolome as nodes that are connected by links (edges). Combining biomedical data with the appropriate network modeling approach allows derivation of disease associated information and outcomes like biomarkers, therapeutics targets, phenotype-specific genes and interactions, and disease subtypes. This figure was reprinted with permission from Sonawane, et. al., 2019.

Expansion and further exploration of the gene regulatory network.

As previously mentioned, the expansion and fine-tuning of the endocrine-specific GRN is of great importance in deepening our understanding of developmental biology. Beyond simply having the raw information to guide future experiments and aid in interpretations, having a more developed GRN would enable us to mathematically model biological phenomena like endocrine cell differentiation. The interdisciplinary efforts of biologists, mathematicians, and bioinformaticians have led to evidence-based models. These models are conceptually guided by the principles of Waddington's landscape (Waddington and Kacser, 1957). The idea of Waddington's landscape originated in 1957 from Conrad Waddington, who described mammalian development as being unidirectional, where embryonic stem cells develop only into more, more differentiated states. This concept is often represented by a pluripotent stem cell (often depicted as a marble) rolling down a hill, with natural valleys or grooves (differentiated cell states) forming along the way. As the stem cell moves down into more differentiated cell states, there are several instances where its path may change (cell fate decisions), putting it on course for a specific cell type (Waddington and Kacser, 1957).

A number of useful applications of gene regulatory networks exist. Perhaps the most obvious of these is in providing a visual causal map of genetic interactions driving cellular behavior. GRNs can be used as a blueprint for understanding key relationships among genes, and this blueprint can help guide future hypotheses and experimentation (Emmert-Streib et al., 2014). One could utilize the network to guide perturbation experiments that may not have been immediately obvious based on other existing data. For example, in my studies, I was able to identify novel ZFPs based on expression networks and on their regulation by Insm1, Neuord1, and Pax6 in the regulatory network. These specific ZFP targets would likely not have been discovered without this collection of data being compiled into a network. This same approach could be used to identify other uncharacterized genes, or genes without a known tissue-specific function, to generate knockout animals for study. Furthermore, it would be really interesting to perform comparative network analyses (Emmert-Streib et al., 2014). If we could generate model GRNs for both β- and α-cells, for example, it would be possible to compare the two to determine where key structural differences lie. These differences could be in the presence or absence of members in the network, directionality of regulatory interactions, or the general activity of a given node that happens to be cell-type dependent. Understanding the unique properties of the two networks could help understand key drivers of this cell fate choice and possibly improve protocols for generating β-cells specifically.

In the same vein, knowing the regulatory architecture of both a normal cell type and comparing that to the regulatory network of cells in a disease state, we could identify biomarkers of that disease state. This I due to the fact that many diseases are complex in nature, may have multiple genetic perturbations, and in turn, their hallmarks are driven by perturbations in the molecular pathways of the cell (Emmert-Streib et al., 2014). For example, T2D is a complex, polygenic disease that is progressive. If we were able to identify cells having a perturbed GRN architecture that mirrors a shift toward the T2D disease state, it would be possible to identify key pathways that are responsible for the loss of β-cell function or identity. Likewise, this information could help guide therapeutic treatments of the disease. We would not only be able to use the network to identify possible drug targets, but it could be used to model the effects of a drug. It would even be feasible to determine associations between regulatory network states and patient responses to treatments (Fang et al., 2020). These approaches would be one step closer to a healthcare system based on personalized medicine (Emmert-Streib et al., 2014).

The underpinning mechanism driving the cell fate choices that the stem cell makes is not chance but rather the regulatory interaction of the GRN. To best model cellular development from a GRN, it is not only important to know which factors are important but also to understand how the temporal sequence of activation affects choices and influences the end phenotype of the cells (Ladewig et al., 2013). By expanding our current understanding of the inherent regulatory interactions of the GRN through perturbations much like those described here, we can build a more accurate and useful model underlying cell fate decisions and cellular functions (**Fig. 6.2**). In short, GRNs can potentially be used for generating novel hypotheses, experimental validation, and for the discovery of novel gene functions (Gupta and Singh, 2019).

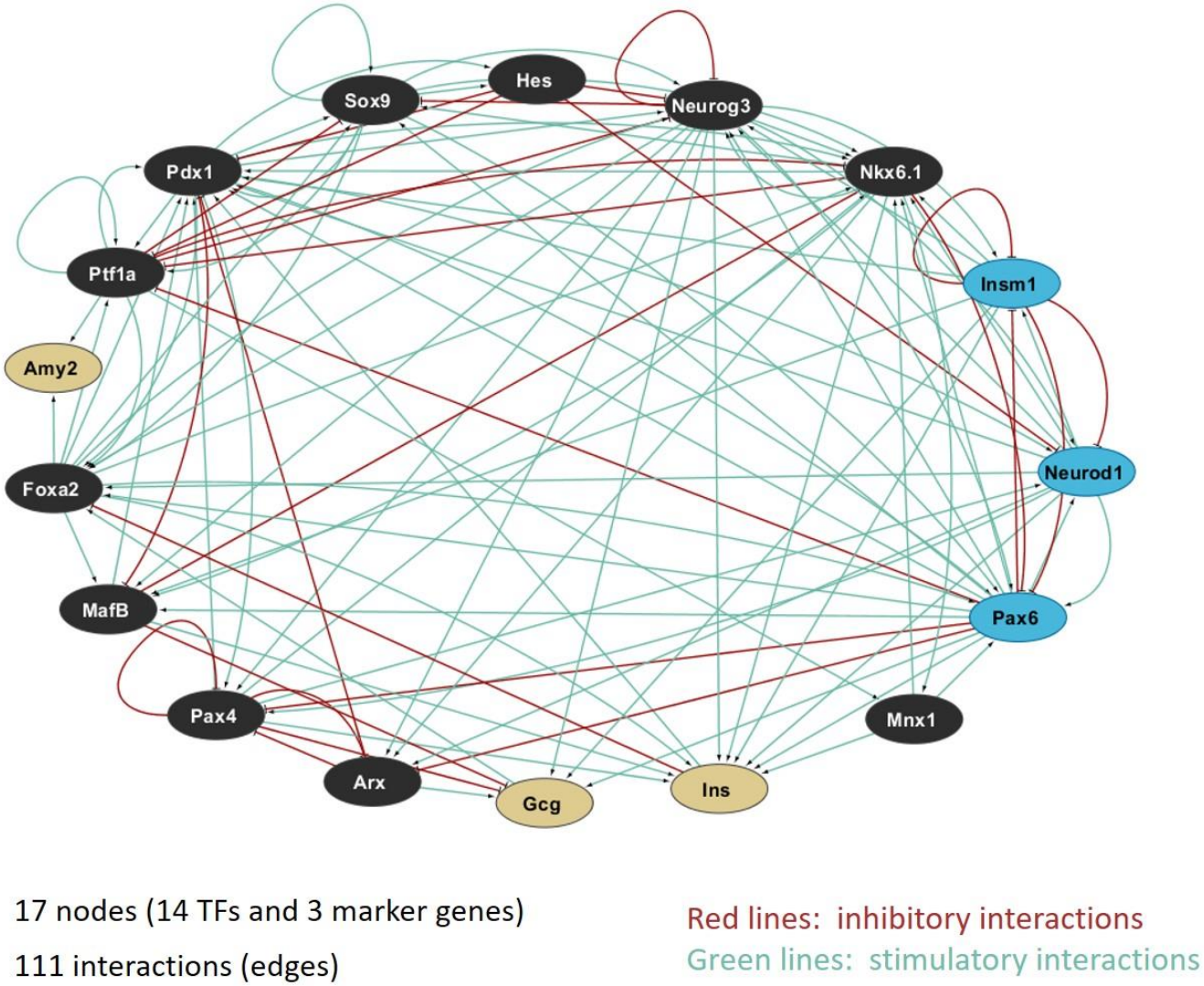


Figure 6.2. Inferred gene regulatory network of developing pancreatic endocrine cells. An expanded pancreatic endocrine cell GRN highlighting interactions among key transcription factors and hormones. Regulatory edges are based on previously reported findings and newly inferred relationships based on our RNA-Seq data described herein. This rendering contains 17 nodes (14 TFs and 3 marker genes) and 111 regulatory interactions (edges). Red lines represent inhibitory interactions, while green lines are activating. Insm1, Neurod1, and Pax6 (TFs discussed here) are highlighted in blue. This is an original work generated using Cytoscape.

Role of alternative splicing in endocrine cells.

The human proteome contains over 200,000 protein products, but the human genome consists of only ~22,000 genes (Lee and Rio, 2015). This is made possible only through the mechanism of alternative splicing (AS). However, our understanding of the biological functions of AS is greatly underdeveloped, particularly in the context of developing pancreatic endocrine cells (Moss and Sussel, 2020). Therefore, a great challenge for the field lies in not only determining the role of splicing regulators but also the splice variant products. How do these factors contribute to endocrine cell development and function? What consequences are observed in beta cells when splicing regulators are lost or the ratio of a transcript's isoforms is altered? There are many approaches available to answer these and other questions. For example, it would be possible to perform conditional knockdown or knockout of specific RBPs to determine the effects in endocrine cell development on morphology, gene expression, and differential splicing events. Specific gene targets could also be studied by creating mutations at the binding sites of RBPs or by selectively expressing the desired isoform. Preferential isoform expression can be achieved by genetic manipulations, such as through CRISPR-mediated mutagenesis, or by introducing siRNAs that can target and knock down the other isoforms, or by transfecting with an expression vector for the desired isoform (Kisielow et al., 2002). It would be interesting to determine how many TFs (*Foxo1*, *Rfx2*, *Foxa2,* etc.) or other genes' (*Insr, Tnfrsf12a, Bmpr2, Robo2*, etc.) functions are affected by alternative splicing and what the physiological output of those changes would be. As previously discussed, there are several instances of disrupted alternative splicing of critical genes having adverse effects, and many are linked to disease states, such as diabetes (*e.g., Gck, Insr, Hnf4a*). It is highly likely that other undiscovered perturbations in splicing dynamics increase disease risk and manifestation.

Additionally, understanding the functions of splicing regulators and alternatively spliced gene products opens a door for future therapies. For example, others have demonstrated that antisense oligonucleotides can be used to specific exons to alter splicing products (Spitali and Aartsma-Rus, 2012). One study has even shown that by treating mice with antisense oligonucleotides targeted to CTLA-4, they were able to reduce the occurrence of diabetes and insulitis (Dlamini et al., 2017; Mourich et al., 2014). These data provide support for the feasibility of treating diseases through alternative splicing mechanisms.

Translating basic science research into improved regenerative medicine approaches.

Diabetes, defined by the body's inability to maintain glucose homeostasis, remains a global problem despite decades of dedicated research efforts. The trend in the number of cases worldwide has also only been growing for the last several years, with future projections rising at steady rates. To date, insulin treatments remain the most common form of therapy. In the last decade, researchers have made great strides in establishing regenerative medicine approaches to create a more enduring treatment for diabetes (Pagliuca et al., 2014; Rezania et al., 2014). Many approaches have been explored, including guided differentiation of induced pluripotent stem cells (iPSC) or embryonic stem cells (ESCs) and direct reprogramming of one cell type to another, but bypassing the transient pluripotent stage (**Fig. 6.3**). These approaches have been performed in many different variations, both *in vitro* and *in vivo*, with great progress being made in both areas. However, the knowledge enabling our advances in this endeavor has been driven by basic research focused on the innate factors involved in determining endocrine cell lineage allocation and the cellular function of mature cells. This means that through continuing to study the fundamentals of endocrine cell development, the field will be able to apply those concepts to optimizing reprogramming protocols.

In our journey to successfully generate beta-cells from non-beta cells, there are key measures that must be tracked and focused on. Those are reprogramming efficiency, cellular identity, stability, functionality, and safety (Ladewig et al., 2013). Having a high level of reprogramming or differentiation efficiency is critical for being able to produce adequate numbers of cells with reasonable resources. This would be measured as the percentage of differentiated cells relative to the number of starting cells. Cellular identity and function are closely related factors and are indications of how closely the newly converted cells resemble the desired cell type (e.g., morphology, transcriptome, etc.) and how well they function (e.g., glucose-responsive insulin secretion). Stability is an important measure to keep in mind and represents how well a converted cell maintains its identity, particularly after extrinsic factors have been removed (i.e., following transplantation). Finally, but perhaps most important, is the factor of safety. Not only do the cells need to maintain their identity and functionality once transplanted to serve as a true treatment for diabetes, but they must also be void of tumorigenesis tendencies. This is directly related to the proliferation potential of cells. Many scientists are focusing on directly reprogramming cells to bypass the pluripotent cell stage of iPSCs and reduce the cells' tumorigenic potential.

Ultimately, efforts must be made on this front to identify regulators that are key for promoting or inhibiting specific cellular stages during development and dedifferentiation states. Gaining ground in our fundamental knowledge of these processes is essential for enhancing the scope of possibilities in restoring β-cell function by therapeutic means.

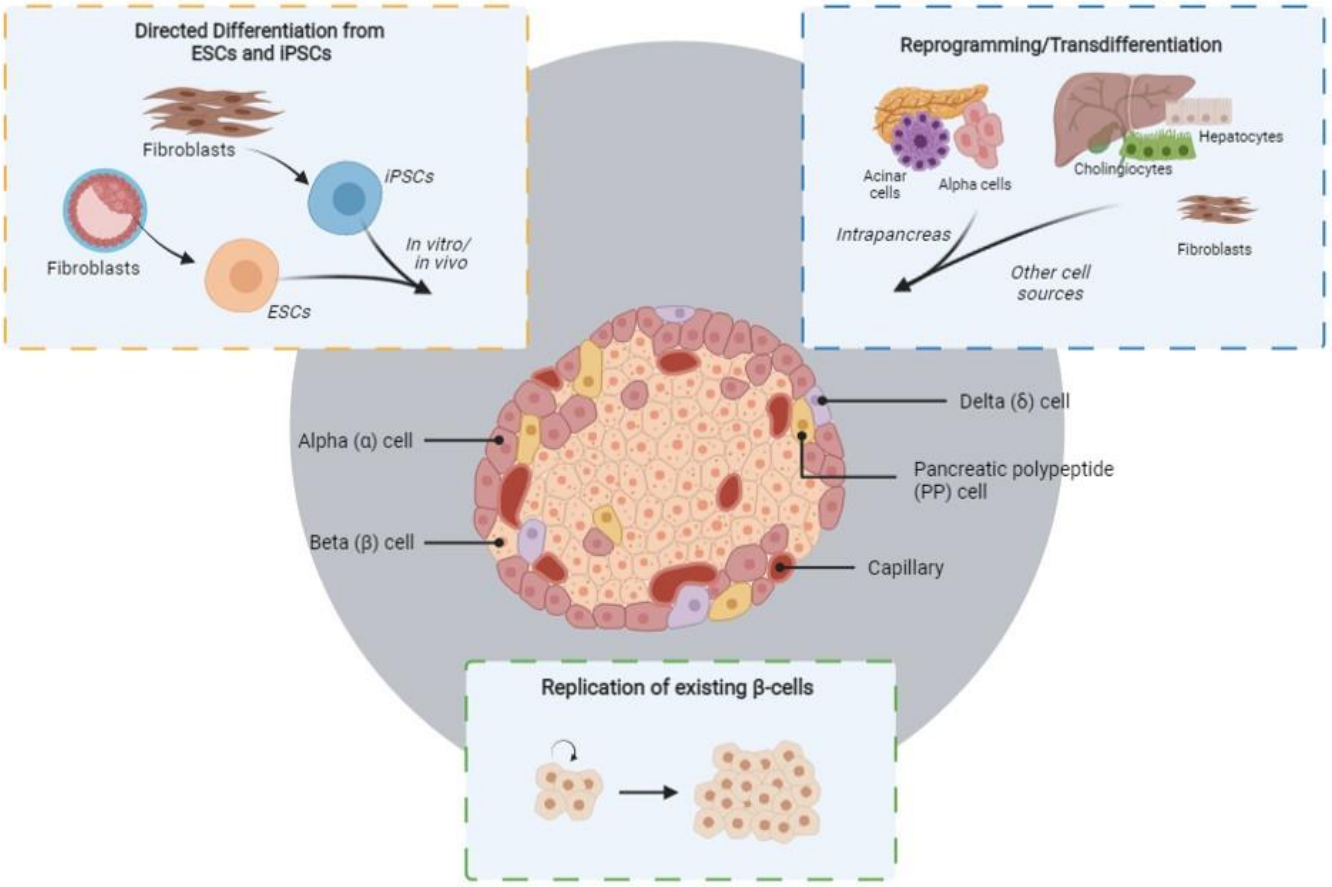


Figure 6.3. Strategies to generate new β-cells. Schematic representation of promising pathways for generating new functionally mature β-cells for treating diabetes. Approaches include directed differentiation of embryonic stem cells (ESCs) or induced pluripotent stem cells (iPSCs), direct reprogramming from alternate cell sources, and the proliferation of existing β-cells. This figure is an original work generated using BioRender.com.

Further characterization of zinc finger proteins

Given the fact that only a fraction of zinc finger proteins (ZFPs) has been characterized to date, there is a great opportunity to expand our understanding of this class of proteins. As previously mentioned, the more recent evolution of ZFPs in higher-order species combined with the fact that they constitute the largest class of proteins in humans makes them a key target for investigation. Furthermore, they have been shown to have tissue-specific and temporally regulated expression patterns, and play key roles in development (Kang et al., 2009; Ladomery and Dellaire, 2002; Osipovich et al., 2021; Watanabe et al., 2009). Even though they are abundant, we are still unsure of the total number of genes and their protein products, making the task of defining and characterizing ZFPs one of the most immediate challenges. The community must also work to determine the biological function of ZFPs in the context of the developing pancreas. While this is currently a difficult challenge due to the lack of adequate resources (e.g., mutant alleles, specific antibodies, etc.), those resources will also not be established or become more widely available if no effort is put forth now.

Many questions remain with respect to the role of ZFPs in endocrine cells. Are there ZFPs that are uniquely expressed in endocrine cells that have yet to be determined? Do they function primarily as regulators of gene expression or as terminal “worker” genes? If the former, how important are they in establishing and maintaining the GRN necessary to differentiation or mature, functioning cells? These are critical questions to begin to probe into. Our study of *Zfp800* revealed that gene expression is robustly dysregulated in null pancreata (Osipovich et al., 2021).

REFERENCES

(2019). IDF Diabetes Atlas. In IDF Diabetes Atlas (International Diabetes Federation).
Agrawal, P., and Rao, S. (2021). Super-Enhancers and CTCF in Early Embryonic Cell Fate Decisions. Front Cell Dev Biol *9*, 653669.
Ahlgren, U., Jonsson, J., Jonsson, L., Simu, K., and Edlund, H. (1998). beta-cell-specific inactivation of the mouse Ipf1/Pdx1 gene results in loss of the beta-cell phenotype and maturity onset diabetes. Genes Dev *12*, 1763-1768.
Ahmad, Z., Rafeeq, M., Collombat, P., and Mansouri, A. (2015). Pax6 Inactivation in the Adult Pancreas Reveals Ghrelin as Endocrine Cell Maturation Marker. PLoS One *10*, e0144597.
Alpsoy, A., Sood, S., and Dykhuizen, E.C. (2021). At the Crossroad of Gene Regulation and Genome Organization: Potential Roles for ATP-Dependent Chromatin Remodelers in the Regulation of CTCF-Mediated 3D Architecture. Biology (Basel) *10*.
Alvelos, M.I., Bruggemann, M., Sutandy, F.R., Juan-Mateu, J., Colli, M.L., Busch, A., Lopes, M., Castela, A., Aartsma-Rus, A., Konig, J.*, et al.* (2021). The RNA-binding profile of the splicing factor SRSF6 in immortalized human pancreatic beta-cells. Life Sci Alliance *4*.
Alvelos, M.I., Juan-Mateu, J., Colli, M.L., Turatsinze, J.V., and Eizirik, D.L. (2018). When one becomes many-Alternative splicing in beta-cell function and failure. Diabetes Obes Metab *20 Suppl 2*, 77-87.
Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169.
Anderson, K.R., White, P., Kaestner, K.H., and Sussel, L. (2009). Identification of known and novel pancreas genes expressed downstream of Nkx2.2 during development. BMC Dev Biol *9*, 65.
Anko, M.L. (2014). Regulation of gene expression programmes by serine-arginine rich splicing factors. Semin Cell Dev Biol *32*, 11-21.
Apelqvist, A., Li, H., Sommer, L., Beatus, P., Anderson, D.J., Honjo, T., Hrabe de Angelis, M., Lendahl, U., and Edlund, H. (1999). Notch signalling controls pancreatic cell differentiation. Nature *400*, 877-881.
Arda, H.E., Benitez, C.M., and Kim, S.K. (2013). Gene regulatory networks governing pancreas development. Dev Cell *25*, 5-13.
Artner, I., Hang, Y., Guo, M., Gu, G., and Stein, R. (2008). MafA is a dedicated activator of the insulin gene in vivo. J Endocrinol *198*, 271-279.
Artner, I., Hang, Y., Mazur, M., Yamamoto, T., Guo, M., Lindner, J., Magnuson, M.A., and Stein, R. (2010). MafA and MafB regulate genes critical to beta-cells in a unique temporal manner. Diabetes *59*, 2530-2539.
Arzate-Mejia, R.G., Recillas-Targa, F., and Corces, V.G. (2018). Developing in 3D: the role of CTCF in cell differentiation. Development *145*.
Ashery-Padan, R., Marquardt, T., Zhou, X., and Gruss, P. (2000). Pax6 activity in the lens primordium is required for lens formation and for correct placement of a single retina in the eye. Genes Dev *14*, 2701-2711.
Ashery-Padan, R., Zhou, X., Marquardt, T., Herrera, P., Toube, L., Berry, A., and Gruss, P. (2004). Conditional inactivation of Pax6 in the pancreas causes early onset of diabetes. Dev Biol *269*, 479-488.
Azzarelli, R., Hurley, C., Sznurkowska, M.K., Rulands, S., Hardwick, L., Gamper, I., Ali, F., McCracken, L., Hindley, C., McDuff, F.*, et al.* (2017). Multi-site Neurogenin3 Phosphorylation Controls Pancreatic Endocrine Differentiation. Dev Cell *41*, 274-286 e275.
Bader, E., Migliorini, A., Gegg, M., Moruzzi, N., Gerdes, J., Roscioni, S.S., Bakhti, M., Brandl, E., Irmler, M., Beckers, J.*, et al.* (2016). Identification of proliferative and mature beta-cells in the islets of Langerhans. Nature *535*, 430-434.
Barbosa-Morais, N.L., Irimia, M., Pan, Q., Xiong, H.Y., Gueroussov, S., Lee, L.J., Slobodeniuc, V., Kutter, C., Watt, S., Colak, R.*, et al.* (2012). The evolutionary landscape of alternative splicing in vertebrate species. Science *338*, 1587-1593.

Bark, C., Bellinger, F.P., Kaushal, A., Mathews, J.R., Partridge, L.D., and Wilson, M.C. (2004). Developmentally regulated switch in alternatively spliced SNAP-25 isoforms alters facilitation of synaptic transmission. J Neurosci *24*, 8796-8805.
Bark, I.C., Hahn, K.M., Ryabinin, A.E., and Wilson, M.C. (1995). Differential expression of SNAP-25 protein isoforms during divergent vesicle fusion events of neural development. Proc Natl Acad Sci U S A *92*, 1510-1514.
Bechard, M.E., Bankaitis, E.D., Hipkens, S.B., Ustione, A., Piston, D.W., Yang, Y.P., Magnuson, M.A., and Wright, C.V. (2016). Precommitment low-level Neurog3 expression defines a long-lived mitotic endocrine-biased progenitor pool that drives production of endocrine-committed cells. Genes Dev *30*, 1852-1865.
Bechard, M.E., Bankaitis, E.D., Ustione, A., Piston, D.W., Magnuson, M.A., and Wright, C.V.E. (2017). FUCCI tracking shows cell-cycle-dependent Neurog3 variation in pancreatic progenitors. Genesis *55*.
Bechard, M.E., and Wright, C.V. (2017). New ideas connecting the cell cycle and pancreatic endocrine-lineage specification. Cell Cycle *16*, 301-303.
Benoit Bouvrette, L.P., Bovaird, S., Blanchette, M., and Lecuyer, E. (2020). oRNAment: a database of putative RNA binding protein target sites in the transcriptomes of model species. Nucleic Acids Res *48*, D166-D173.
Bertrand, N., Castro, D.S., and Guillemot, F. (2002). Proneural genes and the specification of neural cell types. Nat Rev Neurosci *3*, 517-530.
Biamonti, G., Infantino, L., Gaglio, D., and Amato, A. (2019). An Intricate Connection between Alternative Splicing and Phenotypic Plasticity in Development and Cancer. Cells *9*.
Blake, J.A., and Ziman, M.R. (2014). Pax genes: regulators of lineage specification and progenitor cell maintenance. Development *141*, 737-751.
Bonnal, S.C., Lopez-Oreja, I., and Valcarcel, J. (2020). Roles and mechanisms of alternative splicing in cancer - implications for care. Nat Rev Clin Oncol *17*, 457-474.
Bonner-Weir, S., Inada, A., Yatoh, S., Li, W.C., Aye, T., Toschi, E., and Sharma, A. (2008). Transdifferentiation of pancreatic ductal cells to endocrine beta-cells. Biochem Soc Trans *36*, 353-356.
Boschert, U., O'Shaughnessy, C., Dickinson, R., Tessari, M., Bendotti, C., Catsicas, S., and Pich, E.M. (1996). Developmental and plasticity-related differential expression of two SNAP-25 isoforms in the rat brain. J Comp Neurol *367*, 177-193.
Boward, B., Wu, T., and Dalton, S. (2016). Concise Review: Control of Cell Fate Through Cell Cycle and Pluripotency Networks. Stem Cells *34*, 1427-1436.
Brayer, K.J., and Segal, D.J. (2008). Keep your fingers off my DNA: protein-protein interactions mediated by C2H2 zinc finger domains. Cell Biochem Biophys *50*, 111-131.
Breslin, M.B., Zhu, M., and Lan, M.S. (2003). NeuroD1/E47 regulates the E-box element of a novel zinc finger transcription factor, IA-1, in developing nervous system. J Biol Chem *278*, 38991-38997.
Breslin, M.B., Zhu, M., Notkins, A.L., and Lan, M.S. (2002). Neuroendocrine differentiation factor, IA-1, is a transcriptional repressor and contains a specific DNA-binding domain: identification of consensus IA-1 binding sequence. Nucleic Acids Res *30*, 1038-1045.
Bruno, M., Mahgoub, M., and Macfarlan, T.S. (2019). The Arms Race Between KRAB-Zinc Finger Proteins and Endogenous Retroelements and Its Impact on Mammals. Annu Rev Genet *53*, 393-416.
Burlison, J.S., Long, Q., Fujitani, Y., Wright, C.V., and Magnuson, M.A. (2008). Pdx-1 and Ptf1a concurrently determine fate specification of pancreatic multipotent progenitor cells. Dev Biol *316*, 74-86.
Capelli, P., Martignoni, G., Pedica, F., Falconi, M., Antonello, D., Malpeli, G., and Scarpa, A. (2009). Endocrine neoplasms of the pancreas: pathologic and genetic features. Arch Pathol Lab Med *133*, 350-364.
Carazo, F., Gimeno, M., Ferrer-Bonsoms, J.A., and Rubio, A. (2019a). Integration of CLIP experiments of RNA-binding proteins: a novel approach to predict context-dependent splicing factors from transcriptomic data. BMC Genomics *20*, 521.
Carazo, F., Romero, J.P., and Rubio, A. (2019b). Upstream analysis of alternative splicing: a review of computational approaches to predict context-dependent splicing factors. Brief Bioinform *20*, 1358-1375.

Carmona-Aldana, F., Zampedri, C., Suaste-Olmos, F., Murillo-de-Ozores, A., Guerrero, G., Arzate-Mejia, R., Maldonado, E., Navarro, R.E., Chimal-Monroy, J., and Recillas-Targa, F. (2018). CTCF knockout reveals an essential role for this protein during the zebrafish development. Mech Dev *154*, 51-59.
Carroll, D. (2017). Genome Editing: Past, Present, and Future. Yale J Biol Med *90*, 653-659.
Cassandri, M., Smirnov, A., Novelli, F., Pitolli, C., Agostini, M., Malewicz, M., Melino, G., and Raschella, G. (2017). Zinc-finger proteins in health and disease. Cell Death Discov *3*, 17071.
Castro-Diaz, N., Ecco, G., Coluccio, A., Kapopoulou, A., Yazdanpanah, B., Friedli, M., Duc, J., Jang, S.M., Turelli, P., and Trono, D. (2014). Evolutionally dynamic L1 regulation in embryonic stem cells. Genes Dev *28*, 1397-1409.
Chen, T., and Dent, S.Y. (2014). Chromatin modifiers and remodellers: regulators of cellular differentiation. Nat Rev Genet *15*, 93-106.
Chen, Y., Wu, H., Wang, S., Koito, H., Li, J., Ye, F., Hoang, J., Escobar, S.S., Gow, A., Arnett, H.A*., et al.* (2009). The oligodendrocyte-specific G protein-coupled receptor GPR17 is a cell-intrinsic timer of myelination. Nat Neurosci *12*, 1398-1406.
Churchill, A.J., Gutierrez, G.D., Singer, R.A., Lorberbaum, D.S., Fischer, K.A., and Sussel, L. (2017). Genetic evidence that Nkx2.2 acts primarily downstream of Neurog3 in pancreatic endocrine lineage development. Elife *6*.
Cieply, B., and Carstens, R.P. (2015). Functional roles of alternative splicing factors in human disease. Wiley Interdiscip Rev RNA *6*, 311-326.
Cinti, F., Bouchi, R., Kim-Muller, J.Y., Ohmura, Y., Sandoval, P.R., Masini, M., Marselli, L., Suleiman, M., Ratner, L.E., Marchetti, P*., et al.* (2016). Evidence of beta-Cell Dedifferentiation in Human Type 2 Diabetes. J Clin Endocrinol Metab *101*, 1044-1054.
Colli, M.L., Ramos-Rodriguez, M., Nakayasu, E.S., Alvelos, M.I., Lopes, M., Hill, J.L.E., Turatsinze, J.V., Coomans de Brachene, A., Russell, M.A., Raurell-Vila, H*., et al.* (2020). An integrated multi-omics approach identifies the landscape of interferon-alpha-mediated responses of human pancreatic beta cells. Nat Commun *11*, 2584.
Collombat, P., Hecksher-Sorensen, J., Broccoli, V., Krull, J., Ponte, I., Mundiger, T., Smith, J., Gruss, P., Serup, P., and Mansouri, A. (2005). The simultaneous loss of Arx and Pax4 genes promotes a somatostatin-producing cell fate specification at the expense of the alpha- and beta-cell lineages in the mouse endocrine pancreas. Development *132*, 2969-2980.
Collombat, P., and Mansouri, A. (2009). [Pax4 transdifferentiates glucagon-secreting alpha cells to insulin-secreting beta endocrine pancreatic cells]. Med Sci (Paris) *25*, 763-765.
Collombat, P., Xu, X., Ravassard, P., Sosa-Pineda, B., Dussaud, S., Billestrup, N., Madsen, O.D., Serup, P., Heimberg, H., and Mansouri, A. (2009). The ectopic expression of Pax4 in the mouse pancreas converts progenitor cells into alpha and subsequently beta cells. Cell *138*, 449-462.
Conrad, E., Dai, C., Spaeth, J., Guo, M., Cyphert, H.A., Scoville, D., Carroll, J., Yu, W.M., Goodrich, L.V., Harlan, D.M*., et al.* (2016). The MAFB transcription factor impacts islet alpha-cell function in rodents and represents a unique signature of primate islet beta-cells. Am J Physiol Endocrinol Metab *310*, E91-E102.
Costantini, S., Prandini, P., Corradi, M., Pasquali, A., Contreas, G., Pignatti, P.F., Pinelli, L., Trabetti, E., and Maffeis, C. (2011). A novel synonymous substitution in the GCK gene causes aberrant splicing in an Italian patient with GCK-MODY phenotype. Diabetes Res Clin Pract *92*, e23-26.
Courtney, M., Gjernes, E., Druelle, N., Ravaud, C., Vieira, A., Ben-Othman, N., Pfeifer, A., Avolio, F., Leuckx, G., Lacas-Gervais, S*., et al.* (2013). The inactivation of Arx in pancreatic alpha-cells triggers their neogenesis and conversion into functional beta-like cells. PLoS Genet *9*, e1003934.
Cramer, P. (2019). Organization and regulation of gene transcription. Nature *573*, 45-54.
Dalton, S. (2015). Linking the Cell Cycle to Cell Fate Decisions. Trends Cell Biol *25*, 592-600.
Daraio, T., Bombek, L.K., Gosak, M., Valladolid-Acebes, I., Klemen, M.S., Refai, E., Berggren, P.O., Brismar, K., Rupnik, M.S., and Bark, C. (2017). SNAP-25b-deficiency increases insulin secretion and changes spatiotemporal profile of Ca(2+)oscillations in beta cell networks. Sci Rep *7*, 7744.
Dassaye, R., Naidoo, S., and Cerf, M.E. (2016). Transcription factor regulation of pancreatic organogenesis, differentiation and maturation. Islets *8*, 13-34.

Davenport, C., Diekmann, U., Budde, I., Detering, N., and Naujok, O. (2016). Anterior-Posterior Patterning of Definitive Endoderm Generated from Human Embryonic Stem Cells Depends on the Differential Signaling of Retinoic Acid, Wnt-, and BMP-Signaling. Stem Cells *34*, 2635-2647.
Dennis, D.J., Han, S., and Schuurmans, C. (2019). bHLH transcription factors in neural development, disease, and reprogramming. Brain Res *1705*, 48-65.
Desai, C., Garriga, G., McIntire, S.L., and Horvitz, H.R. (1988). A genetic pathway for the development of the Caenorhabditis elegans HSN motor neurons. Nature *336*, 638-646.
Desai, C., and Horvitz, H.R. (1989). Caenorhabditis elegans mutants defective in the functioning of the motor neurons responsible for egg laying. Genetics *121*, 703-721.
Desgraz, R., and Herrera, P.L. (2009). Pancreatic neurogenin 3-expressing cells are unipotent islet precursors. Development *136*, 3567-3574.
Dessimoz, J., and Grapin-Botton, A. (2006). Pancreas development and cancer: Wnt/beta-catenin at issue. Cell Cycle *5*, 7-10.
Di Liegro, C.M., Schiera, G., and Di Liegro, I. (2014). Regulation of mRNA transport, localization and translation in the nervous system of mammals (Review). Int J Mol Med *33*, 747-762.
Dimitri, P., Warner, J.T., Minton, J.A., Patch, A.M., Ellard, S., Hattersley, A.T., Barr, S., Hawkes, D., Wales, J.K., and Gregory, J.W. (2011). Novel GLIS3 mutations demonstrate an extended multisystem phenotype. Eur J Endocrinol *164*, 437-443.
Dlamini, Z., Mokoena, F., and Hull, R. (2017). Abnormalities in alternative splicing in diabetes: therapeutic targets. J Mol Endocrinol *59*, R93-R107.
Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.
Doi, H., Yoshida, K., Yasuda, T., Fukuda, M., Fukuda, Y., Morita, H., Ikeda, S., Kato, R., Tsurusaki, Y., Miyake, N.*, et al.* (2011). Exome sequencing reveals a homozygous SYT14 mutation in adult-onset, autosomal-recessive spinocerebellar ataxia with psychomotor retardation. Am J Hum Genet *89*, 320-327.
Dor, Y., and Glaser, B. (2013). beta-cell dedifferentiation and type 2 diabetes. N Engl J Med *368*, 572-573.
Dowen, J.M., Fan, Z.P., Hnisz, D., Ren, G., Abraham, B.J., Zhang, L.N., Weintraub, A.S., Schujiers, J., Lee, T.I., Zhao, K.*, et al.* (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell *159*, 374-387.
Duggan, A., Madathany, T., de Castro, S.C., Gerrelli, D., Guddati, K., and Garcia-Anoveros, J. (2008). Transient expression of the conserved zinc finger gene INSM1 in progenitors and nascent neurons throughout embryonic and adult neurogenesis. J Comp Neurol *507*, 1497-1520.
Duverger, O., and Morasso, M.I. (2009). Epidermal patterning and induction of different hair types during mouse embryonic development. Birth Defects Res C Embryo Today *87*, 263-272.
Dvinge, H. (2018). Regulation of alternative mRNA splicing: old players and new perspectives. FEBS Lett *592*, 2987-3006.
Ecco, G., Imbeault, M., and Trono, D. (2017). KRAB zinc finger proteins. Development *144*, 2719-2729.
Ediger, B.N., Du, A., Liu, J., Hunter, C.S., Walp, E.R., Schug, J., Kaestner, K.H., Stein, R., Stoffers, D.A., and May, C.L. (2014). Islet-1 Is essential for pancreatic beta-cell function. Diabetes *63*, 4206-4217.
Eizirik, D.L., Sammeth, M., Bouckenooghe, T., Bottu, G., Sisino, G., Igoillo-Esteve, M., Ortis, F., Santin, I., Colli, M.L., Barthson, J.*, et al.* (2012). The human pancreatic islet transcriptome: expression of candidate genes for type 1 diabetes and the impact of pro-inflammatory cytokines. PLoS Genet *8*, e1002552.
Ejarque, M., Cervantes, S., Pujadas, G., Tutusaus, A., Sanchez, L., and Gasa, R. (2013). Neurogenin3 cooperates with Foxa2 to autoactivate its own expression. J Biol Chem *288*, 11705-11717.
Emerson, R.O., and Thomas, J.H. (2009). Adaptive evolution in zinc finger transcription factors. PLoS Genet *5*, e1000325.
Emmert-Streib, F., Dehmer, M., and Haibe-Kains, B. (2014). Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks. Front Cell Dev Biol *2*, 38.

Escribano, O., Beneit, N., Rubio-Longas, C., Lopez-Pastor, A.R., and Gomez-Hernandez, A. (2017). The Role of Insulin Receptor Isoforms in Diabetes and Its Metabolic and Vascular Complications. J Diabetes Res *2017*, 1403206.
Fadista, J., Vikman, P., Laakso, E.O., Mollet, I.G., Esguerra, J.L., Taneera, J., Storm, P., Osmark, P., Ladenvall, C., Prasad, R.B.*, et al.* (2014). Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. Proc Natl Acad Sci U S A *111*, 13924-13929.
Fang, J., Pan, Z., Yu, H., Yang, S., Hu, X., Lu, X., and Li, L. (2020). Regulatory Master Genes Identification and Drug Repositioning by Integrative mRNA-miRNA Network Analysis for Acute Type A Aortic Dissection. Front Pharmacol *11*, 575765.
Fang, Y., Gao, J., Qi, L., and Li, N. (2014). CTCF-regulating endocrine function of pancreatic islet cells in transgenic mice. Horm Metab Res *46*, 419-423.
Farkas, L.M., Haffner, C., Giger, T., Khaitovich, P., Nowick, K., Birchmeier, C., Paabo, S., and Huttner, W.B. (2008). Insulinoma-associated 1 has a panneurogenic role and promotes the generation and expansion of basal progenitors in the developing mouse neocortex. Neuron *60*, 40-55.
Feero, W.G., Guttmacher, A.E., and Collins, F.S. (2010). Genomic medicine--an updated primer. N Engl J Med *362*, 2001-2011.
Filippova, G.N., Fagerlie, S., Klenova, E.M., Myers, C., Dehner, Y., Goodwin, G., Neiman, P.E., Collins, S.J., and Lobanenkov, V.V. (1996). An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. Mol Cell Biol *16*, 2802-2813.
Fiszbein, A., and Kornblihtt, A.R. (2017). Alternative splicing switches: Important players in cell differentiation. Bioessays *39*.
Flannick, J., Johansson, S., and Njolstad, P.R. (2016). Common and rare forms of diabetes mellitus: towards a continuum of diabetes subtypes. Nat Rev Endocrinol *12*, 394-406.
Franco, M.M., Prickett, A.R., and Oakey, R.J. (2014). The role of CCCTC-binding factor (CTCF) in genomic imprinting, development, and reproduction. Biol Reprod *91*, 125.
Fujitani, Y., Fujitani, S., Boyer, D.F., Gannon, M., Kawaguchi, Y., Ray, M., Shiota, M., Stein, R.W., Magnuson, M.A., and Wright, C.V. (2006). Targeted deletion of a cis-regulatory region reveals differential gene dosage requirements for Pdx1 in foregut organ differentiation and pancreas formation. Genes Dev *20*, 253-266.
Gao, T., McKenna, B., Li, C., Reichert, M., Nguyen, J., Singh, T., Yang, C., Pannikar, A., Doliba, N., Zhang, T.*, et al.* (2014). Pdx1 maintains beta cell identity and function by repressing an alpha cell program. Cell Metab *19*, 259-271.
Garin, I., Rica, I., Estalella, I., Oyarzabal, M., Rodriguez-Rigual, M., San Pedro, J.I., Perez-Nanclares, G., Fernandez-Rebollo, E., Busturia, M.A., Castano, L.*, et al.* (2008). Haploinsufficiency at GCK gene is not a frequent event in MODY2 patients. Clin Endocrinol (Oxf) *68*, 873-878.
Gebauer, F., and Hentze, M.W. (2004). Molecular mechanisms of translational control. Nat Rev Mol Cell Biol *5*, 827-835.
Gehring, N.H., and Roignant, J.Y. (2021). Anything but Ordinary - Emerging Splicing Mechanisms in Eukaryotic Gene Regulation. Trends Genet *37*, 355-372.
Georgia, S., Soliz, R., Li, M., Zhang, P., and Bhushan, A. (2006). p57 and Hes1 coordinate cell cycle exit with self-renewal of pancreatic progenitors. Dev Biol *298*, 22-31.
Gierl, M.S., Karoulias, N., Wende, H., Strehle, M., and Birchmeier, C. (2006). The zinc-finger factor Insm1 (IA-1) is essential for the development of pancreatic beta cells and intestinal endocrine cells. Genes Dev *20*, 2465-2478.
Gilbert, J.M., and Blum, B. (2018). Synaptotagmins Tweak Functional beta Cell Maturation. Dev Cell *45*, 284-286.
Glaser, T., Lane, J., and Housman, D. (1990). A mouse model of the aniridia-Wilms tumor deletion syndrome. Science *250*, 823-827.
Gonelle-Gispert, C., Halban, P.A., Niemann, H., Palmer, M., Catsicas, S., and Sadoul, K. (1999). SNAP-25a and -25b isoforms are both expressed in insulin-secreting cells and can function in insulin secretion. Biochem J *339 (Pt 1)*, 159-165.

Gopel, S.O., Kanno, T., Barg, S., Weng, X.G., Gromada, J., and Rorsman, P. (2000). Regulation of glucagon release in mouse -cells by KATP channels and inactivation of TTX-sensitive Na+ channels. J Physiol *528*, 509-520.
Gosmain, Y., Marthinet, E., Cheyssac, C., Guerardel, A., Mamin, A., Katz, L.S., Bouzakri, K., and Philippe, J. (2010). Pax6 controls the expression of critical genes involved in pancreatic {alpha} cell differentiation and function. J Biol Chem *285*, 33381-33393.
Goto, Y., De Silva, M.G., Toscani, A., Prabhakar, B.S., Notkins, A.L., and Lan, M.S. (1992). A novel human insulinoma-associated cDNA, IA-1, encodes a protein with "zinc-finger" DNA-binding motifs. J Biol Chem *267*, 15252-15257.
Gradwohl, G., Dierich, A., LeMeur, M., and Guillemot, F. (2000). neurogenin3 is required for the development of the four endocrine cell lineages of the pancreas. Proc Natl Acad Sci U S A *97*, 1607-1611.
Grapin-Botton, A. (2005). Ductal cells of the pancreas. Int J Biochem Cell Biol *37*, 504-510.
Gu, C., Stein, G.H., Pan, N., Goebbels, S., Hornberg, H., Nave, K.A., Herrera, P., White, P., Kaestner, K.H., Sussel, L., *et al.* (2010a). Pancreatic beta cells require NeuroD to achieve and maintain functional maturity. Cell Metab *11*, 298-310.
Gu, G., Brown, J.R., and Melton, D.A. (2003). Direct lineage tracing reveals the ontogeny of pancreatic cell fates during mouse embryogenesis. Mech Dev *120*, 35-43.
Gu, G., Dubauskaite, J., and Melton, D.A. (2002). Direct evidence for the pancreatic lineage: NGN3+ cells are islet progenitors and are distinct from duct progenitors. Development *129*, 2447-2457.
Gu, W., Winters, K.A., Motani, A.S., Komorowski, R., Zhang, Y., Liu, Q., Wu, X., Rulifson, I.C., Sivits, G., Jr., Graham, M., *et al.* (2010b). Glucagon receptor antagonist-mediated improvements in glycemic control are dependent on functional pancreatic GLP-1 receptor. Am J Physiol Endocrinol Metab *299*, E624-632.
Guo, T., Wang, W., Zhang, H., Liu, Y., Chen, P., Ma, K., and Zhou, C. (2011). ISL1 promotes pancreatic islet cell proliferation. PLoS One *6*, e22387.
Gupta, P., and Singh, S.K. (2019). Gene Regulatory Networks: Current Updates and Applications in Plant Biology. In Molecular Approaches in Plant Biology and Environmental Challenges, S.P. Singh, S.K. Upadhyay, A. Pandey, and S. Kumar, eds. (Singapore: Springer Singapore), pp. 395-417.
Gupta, R.M., and Musunuru, K. (2014). Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. J Clin Invest *124*, 4154-4161.
Habeb, A.M., Al-Magamsi, M.S., Eid, I.M., Ali, M.I., Hattersley, A.T., Hussain, K., and Ellard, S. (2012). Incidence, genetics, and clinical phenotype of permanent neonatal diabetes mellitus in northwest Saudi Arabia. Pediatr Diabetes *13*, 499-505.
Hanas, J.S., Hazuda, D.J., Bogenhagen, D.F., Wu, F.Y., and Wu, C.W. (1983). Xenopus transcription factor A requires zinc for binding to the 5 S RNA gene. J Biol Chem *258*, 14120-14125.
Hanson, I.M. (2003). PAX6 and congenital eye malformations. Pediatr Res *54*, 791-796.
Harries, L.W., Ellard, S., Stride, A., Morgan, N.G., and Hattersley, A.T. (2006). Isomers of the TCF1 gene encoding hepatocyte nuclear factor-1 alpha show differential expression in the pancreas and define the relationship between mutation position and clinical phenotype in monogenic diabetes. Hum Mol Genet *15*, 2216-2224.
Harries, L.W., Sloman, M.J., Sellers, E.A., Hattersley, A.T., and Ellard, S. (2008). Diabetes susceptibility in the Canadian Oji-Cree population is moderated by abnormal mRNA processing of HNF1A G319S transcripts. Diabetes *57*, 1978-1982.
Heller, R.S., Stoffers, D.A., Liu, A., Schedl, A., Crenshaw, E.B., 3rd, Madsen, O.D., and Serup, P. (2004). The role of Brn4/Pou3f4 and Pax6 in forming the pancreatic glucagon cell identity. Dev Biol *268*, 123-134.
Heremans, Y., Van De Casteele, M., in't Veld, P., Gradwohl, G., Serup, P., Madsen, O., Pipeleers, D., and Heimberg, H. (2002). Recapitulation of embryonic neuroendocrine differentiation in adult human pancreatic duct cells expressing neurogenin 3. J Cell Biol *159*, 303-312.
Herrera, P.L. (2000). Adult insulin- and glucagon-producing cells differentiate from two independent cell lineages. Development *127*, 2317-2322.
Hogan, B.L., Hirst, E.M., Horsburgh, G., and Hetherington, C.M. (1988). Small eye (Sey): a mouse model for the genetic analysis of craniofacial abnormalities. Development *103 Suppl*, 115-119.

Hogan, B.L., Horsburgh, G., Cohen, J., Hetherington, C.M., Fisher, G., and Lyon, M.F. (1986). Small eyes (Sey): a homozygous lethal mutation on chromosome 2 which affects the differentiation of both lens and nasal placodes in the mouse. J Embryol Exp Morphol *97*, 95-110.
Holland, A.M., Hale, M.A., Kagami, H., Hammer, R.E., and MacDonald, R.J. (2002). Experimental control of pancreatic development and maintenance. Proc Natl Acad Sci U S A *99*, 12236-12241.
Hribal, M.L., Perego, L., Lovari, S., Andreozzi, F., Menghini, R., Perego, C., Finzi, G., Usellini, L., Placidi, C., Capella, C.*, et al.* (2003). Chronic hyperglycemia impairs insulin secretion by affecting insulin receptor expression, splicing, and signaling in RIN beta cell line and human islets of Langerhans. FASEB J *17*, 1340-1342.
Hrzenjak, A. (2016). JAZF1/SUZ12 gene fusion in endometrial stromal sarcomas. Orphanet J Rare Dis *11*, 15.
Hu, R., Walker, E., Huang, C., Xu, Y., Weng, C., Erickson, G.E., Coldren, A., Yang, X., Brissova, M., Kaverina, I.*, et al.* (2020). Myt Transcription Factors Prevent Stress-Response Gene Overactivation to Enable Postnatal Pancreatic beta Cell Proliferation, Function, and Survival. Dev Cell *53*, 754.
Huang, C., Walker, E.M., Dadi, P.K., Hu, R., Xu, Y., Zhang, W., Sanavia, T., Mun, J., Liu, J., Nair, G.G.*, et al.* (2018). Synaptotagmin 4 Regulates Pancreatic beta Cell Maturation by Modulating the Ca(2+) Sensitivity of Insulin Secretion Vesicles. Dev Cell *45*, 347-361 e345.
Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res *37*, 1-13.
Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44-57.
Huang, H.P., Liu, M., El-Hodiri, H.M., Chu, K., Jamrich, M., and Tsai, M.J. (2000). Regulation of the pancreatic islet-specific gene BETA2 (neuroD) by neurogenin 3. Mol Cell Biol *20*, 3292-3307.
Ikematsu, Y., Tanaka, K., Toyokawa, G., Ijichi, K., Ando, N., Yoneshima, Y., Iwama, E., Inoue, H., Tagawa, T., Nakanishi, Y.*, et al.* (2020). NEUROD1 is highly expressed in extensive-disease small cell lung cancer and promotes tumor cell migration. Lung Cancer *146*, 97-104.
Imbeault, M., Helleboid, P.Y., and Trono, D. (2017). KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. Nature *543*, 550-554.
Irfan, M., Gopaul, K.R., Miry, O., Hokfelt, T., Stanton, P.K., and Bark, C. (2019). SNAP-25 isoforms differentially regulate synaptic transmission and long-term synaptic plasticity at central synapses. Sci Rep *9*, 6403.
Irimia, M., Rukov, J.L., Penny, D., and Roy, S.W. (2007). Functional and evolutionary analysis of alternatively spliced genes is consistent with an early eukaryotic origin of alternative splicing. BMC Evol Biol *7*, 188.
Iype, T., Francis, J., Garmey, J.C., Schisler, J.C., Nesher, R., Weir, G.C., Becker, T.C., Newgard, C.B., Griffen, S.C., and Mirmira, R.G. (2005). Mechanism of insulin gene regulation by the pancreatic transcription factor Pdx-1: application of pre-mRNA analysis and chromatin immunoprecipitation to assess formation of functional transcriptional complexes. J Biol Chem *280*, 16798-16807.
Jacobs, F.M., Greenberg, D., Nguyen, N., Haeussler, M., Ewing, A.D., Katzman, S., Paten, B., Salama, S.R., and Haussler, D. (2014). An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. Nature *516*, 242-245.
Jacobson, D.A., and Philipson, L.H. (2007). Action potentials and insulin secretion: new insights into the role of Kv channels. Diabetes Obes Metab *9 Suppl 2*, 89-98.
Jacobsson, G., Bark, C., and Meister, B. (1999). Differential expression of SNAP-25a and SNAP-25b RNA transcripts in cranial nerve nuclei. J Comp Neurol *411*, 591-600.
Jacquemin, P., Durviaux, S.M., Jensen, J., Godfraind, C., Gradwohl, G., Guillemot, F., Madsen, O.D., Carmeliet, P., Dewerchin, M., Collen, D.*, et al.* (2000). Transcription factor hepatocyte nuclear factor 6 regulates pancreatic endocrine cell differentiation and controls expression of the proendocrine gene ngn3. Mol Cell Biol *20*, 4445-4454.
Jain, K., Fraser, C.S., Marunde, M.R., Parker, M.M., Sagum, C., Burg, J.M., Hall, N., Popova, I.K., Rodriguez, K.L., Vaidya, A.*, et al.* (2020). Characterization of the plant homeodomain (PHD) reader family for their histone tail interactions. Epigenetics Chromatin *13*, 3.

Jamieson, A.C., Miller, J.C., and Pabo, C.O. (2003). Drug discovery with engineered zinc-finger proteins. Nat Rev Drug Discov *2*, 361-368.
Jan, Y.N., and Jan, L.Y. (1993). HLH proteins, fly neurogenesis, and vertebrate myogenesis. Cell *75*, 827-830.
Jeans, A.F., Oliver, P.L., Johnson, R., Capogna, M., Vikman, J., Molnar, Z., Babbs, A., Partridge, C.J., Salehi, A., Bengtsson, M*., et al.* (2007). A dominant mutation in Snap25 causes impaired vesicle trafficking, sensorimotor gating, and ataxia in the blind-drunk mouse. Proc Natl Acad Sci U S A *104*, 2431-2436.
Jeffery, N., Richardson, S., Chambers, D., Morgan, N.G., and Harries, L.W. (2019). Cellular stressors may alter islet hormone cell proportions by moderation of alternative splicing patterns. Hum Mol Genet *28*, 2763-2774.
Jensen, J. (2004). Gene regulatory factors in pancreatic development. Dev Dyn *229*, 176-200.
Jensen, J., Heller, R.S., Funder-Nielsen, T., Pedersen, E.E., Lindsell, C., Weinmaster, G., Madsen, O.D., and Serup, P. (2000a). Independent development of pancreatic alpha- and beta-cells from neurogenin3-expressing precursors: a role for the notch pathway in repression of premature differentiation. Diabetes *49*, 163-176.
Jensen, J., Pedersen, E.E., Galante, P., Hald, J., Heller, R.S., Ishibashi, M., Kageyama, R., Guillemot, F., Serup, P., and Madsen, O.D. (2000b). Control of endodermal endocrine development by Hes-1. Nat Genet *24*, 36-44.
Jeong, S. (2017). SR Proteins: Binders, Regulators, and Connectors of RNA. Mol Cells *40*, 1-9.
Jetten, A.M. (2018). GLIS1-3 transcription factors: critical roles in the regulation of multiple physiological processes and diseases. Cell Mol Life Sci *75*, 3473-3494.
Jewell, J.L., Oh, E., and Thurmond, D.C. (2010). Exocytosis mechanisms underlying insulin release and glucose uptake: conserved roles for Munc18c and syntaxin 4. Am J Physiol Regul Integr Comp Physiol *298*, R517-531.
Jia, S., Ivanov, A., Blasevic, D., Muller, T., Purfurst, B., Sun, W., Chen, W., Poy, M.N., Rajewsky, N., and Birchmeier, C. (2015). Insm1 cooperates with Neurod1 and Foxa2 to maintain mature pancreatic beta-cell function. EMBO J *34*, 1417-1433.
Joazeiro, C.A., and Weissman, A.M. (2000). RING finger proteins: mediators of ubiquitin ligase activity. Cell *102*, 549-552.
Johansson, K.A., Dursun, U., Jordan, N., Gu, G., Beermann, F., Gradwohl, G., and Grapin-Botton, A. (2007). Temporal control of neurogenin3 activity in pancreas progenitors reveals competence windows for the generation of different endocrine cell types. Dev Cell *12*, 457-465.
Jones, S. (2004). An overview of the basic helix-loop-helix proteins. Genome Biol *5*, 226.
Jonsson, J., Carlsson, L., Edlund, T., and Edlund, H. (1994). Insulin-promoter-factor 1 is required for pancreas development in mice. Nature *371*, 606-609.
Jorgensen, M.C., Ahnfelt-Ronne, J., Hald, J., Madsen, O.D., Serup, P., and Hecksher-Sorensen, J. (2007). An illustrated review of early pancreas development in the mouse. Endocr Rev *28*, 685-705.
Juan-Mateu, J., Alvelos, M.I., Turatsinze, J.V., Villate, O., Lizarraga-Mollinedo, E., Grieco, F.A., Marroqui, L., Bugliani, M., Marchetti, P., and Eizirik, D.L. (2018). SRp55 Regulates a Splicing Network That Controls Human Pancreatic beta-Cell Function and Survival. Diabetes *67*, 423-436.
Kadkova, A., Radecke, J., and Sorensen, J.B. (2019). The SNAP-25 Protein Family. Neuroscience *420*, 50-71.
Kang, H.S., Kim, Y.S., ZeRuth, G., Beak, J.Y., Gerrish, K., Kilic, G., Sosa-Pineda, B., Jensen, J., Pierreux, C.E., Lemaigre, F.P*., et al.* (2009). Transcription factor Glis3, a novel critical player in the regulation of pancreatic beta-cell development and insulin gene expression. Mol Cell Biol *29*, 6366-6379.
Kang, H.S., Takeda, Y., Jeon, K., and Jetten, A.M. (2016). The Spatiotemporal Pattern of Glis3 Expression Indicates a Regulatory Function in Bipotent and Endocrine Progenitors during Early Pancreatic Development and in Beta, PP and Ductal Cells. PLoS One *11*, e0157138.
Kang, M.A., and Lee, J.S. (2021). A Newly Assigned Role of CTCF in Cellular Response to Broken DNAs. Biomolecules *11*.
Karthikeyan, A.S., Lai, Y.H., and Khetan, V. (2017). The Directions Are on the Box. J Pediatr Ophthalmol Strabismus *54*, 75-76.
Kawaguchi, Y., Cooper, B., Gannon, M., Ray, M., MacDonald, R.J., and Wright, C.V. (2002). The role of the transcriptional regulator Ptf1a in converting intestinal to pancreatic progenitors. Nat Genet *32*, 128-134.

Kelemen, O., Convertini, P., Zhang, Z., Wen, Y., Shen, M., Falaleeva, M., and Stamm, S. (2013). Function of alternative splicing. Gene *514*, 1-30.
Kim, E., Magen, A., and Ast, G. (2007a). Different levels of alternative splicing among eukaryotes. Nucleic Acids Res *35*, 125-131.
Kim, S.H., Park, J., Choi, M.C., Kim, H.P., Park, J.H., Jung, Y., Lee, J.H., Oh, D.Y., Im, S.A., Bang, Y.J.*, et al.* (2007b). Zinc-fingers and homeoboxes 1 (ZHX1) binds DNA methyltransferase (DNMT) 3B to enhance DNMT3B-mediated transcriptional repression. Biochem Biophys Res Commun *355*, 318-323.
Kim, W.Y. (2013). NeuroD regulates neuronal migration. Mol Cells *35*, 444-449.
Kim, Y.H., Larsen, H.L., Rue, P., Lemaire, L.A., Ferrer, J., and Grapin-Botton, A. (2015). Cell cycle-dependent differentiation dynamics balances growth and endocrine differentiation in the pancreas. PLoS Biol *13*, e1002111.
Kisielow, M., Kleiner, S., Nagasawa, M., Faisal, A., and Nagamine, Y. (2002). Isoform-specific knockdown and expression of adaptor protein ShcA using small interfering RNA. Biochem J *363*, 1-5.
Klenova, E.M., Nicolas, R.H., Paterson, H.F., Carne, A.F., Heath, C.M., Goodwin, G.H., Neiman, P.E., and Lobanenkov, V.V. (1993). CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms. Mol Cell Biol *13*, 7612-7624.
Kobiita, A., Godbersen, S., Araldi, E., Ghoshdastider, U., Schmid, M.W., Spinas, G., Moch, H., and Stoffel, M. (2020). The Diabetes Gene JAZF1 Is Essential for the Homeostatic Control of Ribosome Biogenesis and Function in Metabolic Stress. Cell Rep *32*, 107846.
Kopp, J.L., Dubois, C.L., Schaffer, A.E., Hao, E., Shih, H.P., Seymour, P.A., Ma, J., and Sander, M. (2011). Sox9+ ductal cells are multipotent progenitors throughout development but do not produce new endocrine cells in the normal or injured adult pancreas. Development *138*, 653-665.
Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.
Krentz, N.A.J., van Hoof, D., Li, Z., Watanabe, A., Tang, M., Nian, C., German, M.S., and Lynn, F.C. (2017). Phosphorylation of NEUROG3 Links Endocrine Differentiation to the Cell Cycle in Pancreatic Progenitors. Dev Cell *41*, 129-142 e126.
Kulzer, J.R., Stitzel, M.L., Morken, M.A., Huyghe, J.R., Fuchsberger, C., Kuusisto, J., Laakso, M., Boehnke, M., Collins, F.S., and Mohlke, K.L. (2014). A common functional regulatory variant at a type 2 diabetes locus upregulates ARAP1 expression in the pancreatic beta cell. Am J Hum Genet *94*, 186-197.
Kumar, M., Jordan, N., Melton, D., and Grapin-Botton, A. (2003). Signals from lateral plate mesoderm instruct endoderm toward a pancreatic fate. Dev Biol *259*, 109-122.
Kuroda, A., Kaneto, H., Fujitani, Y., Watada, H., Nakatani, Y., Kajimoto, Y., Matsuhisa, M., Yamasakai, Y., and Fujiwara, M. (2004). Mutation of the Pax6 gene causes impaired glucose-stimulated insulin secretion. Diabetologia *47*, 2039-2041.
Ladewig, J., Koch, P., and Brustle, O. (2013). Leveling Waddington: the emergence of direct programming and the loss of cell fate hierarchies. Nat Rev Mol Cell Biol *14*, 225-236.
Ladomery, M., and Dellaire, G. (2002). Multifunctional zinc finger proteins in development and disease. Ann Hum Genet *66*, 331-342.
Lan, M.S., and Breslin, M.B. (2009). Structure, expression, and biological function of INSM1 transcription factor in neuroendocrine differentiation. FASEB J *23*, 2024-2033.
Lan, M.S., Lu, J., Goto, Y., and Notkins, A.L. (1994). Molecular cloning and identification of a receptor-type protein tyrosine phosphatase, IA-2, from human insulinoma. DNA Cell Biol *13*, 505-514.
Le Bacquer, O., Shu, L., Marchand, M., Neve, B., Paroni, F., Kerr Conte, J., Pattou, F., Froguel, P., and Maedler, K. (2011). TCF7L2 splice variants have distinct effects on beta-cell turnover and function. Hum Mol Genet *20*, 1906-1915.
Lee, J.C., Smith, S.B., Watada, H., Lin, J., Scheel, D., Wang, J., Mirmira, R.G., and German, M.S. (2001). Regulation of the pancreatic pro-endocrine gene neurogenin3. Diabetes *50*, 928-936.

Lee, J.E., Hollenberg, S.M., Snider, L., Turner, D.L., Lipnick, N., and Weintraub, H. (1995). Conversion of Xenopus ectoderm into neurons by NeuroD, a basic helix-loop-helix protein. Science *268*, 836-844.
Lee, Y., and Rio, D.C. (2015). Mechanisms and Regulation of Alternative Pre-mRNA Splicing. Annu Rev Biochem *84*, 291-323.
Levine, M., and Davidson, E.H. (2005). Gene regulatory networks for development. Proc Natl Acad Sci U S A *102*, 4936-4942.
Lewis, S.L., and Tam, P.P. (2006). Definitive endoderm of the mouse embryo: formation, cell fates, and morphogenetic function. Dev Dyn *235*, 2315-2329.
Li, X.Y., Zhai, W.J., and Teng, C.B. (2015). Notch Signaling in Pancreatic Development. Int J Mol Sci *17*.
Liang, T., Qin, T., Kang, F., Kang, Y., Xie, L., Zhu, D., Dolai, S., Greitzer-Antes, D., Baker, R.K., Feng, D*., et al.* (2020). SNAP23 depletion enables more SNAP25/calcium channel excitosome formation to increase insulin exocytosis in type 2 diabetes. JCI Insight *5*.
Liang, Y., Jetton, T.L., Zimmerman, E.C., Najafi, H., Matschinsky, F.M., and Magnuson, M.A. (1991). Effects of alternate RNA splicing on glucokinase isoform activities in the pancreatic islet, liver, and pituitary. J Biol Chem *266*, 6999-7007.
Licatalosi, D.D., and Darnell, R.B. (2010). RNA processing and its regulation: global insights into biological networks. Nat Rev Genet *11*, 75-87.
Liu, M., Pleasure, S.J., Collins, A.E., Noebels, J.L., Naya, F.J., Tsai, M.J., and Lowenstein, D.H. (2000). Loss of BETA2/NeuroD leads to malformation of the dentate gyrus and epilepsy. Proc Natl Acad Sci U S A *97*, 865-870.
Liu, W.D., Wang, H.W., Muguira, M., Breslin, M.B., and Lan, M.S. (2006). INSM1 functions as a transcriptional repressor of the neuroD/beta2 gene through the recruitment of cyclin D1 and histone deacetylases. Biochem J *397*, 169-177.
Liu, Z., Lam, N., and Thiele, C.J. (2015). Zinc finger transcription factor CASZ1 interacts with histones, DNA repair proteins and recruits NuRD complex to regulate gene transcription. Oncotarget *6*, 27628-27640.
Lizio, M., Ishizu, Y., Itoh, M., Lassmann, T., Hasegawa, A., Kubosaki, A., Severin, J., Kawaji, H., Nakamura, Y., consortium, F*., et al.* (2015). Mapping Mammalian Cell-type-specific Transcriptional Regulatory Networks Using KD-CAGE and ChIP-seq Data in the TC-YIK Cell Line. Front Genet *6*, 331.
Long, J.C., and Caceres, J.F. (2009). The SR protein family of splicing factors: master regulators of gene expression. Biochem J *417*, 15-27.
Lorini, R., Klersy, C., d'Annunzio, G., Massa, O., Minuto, N., Iafusco, D., Bellanne-Chantelot, C., Frongia, A.P., Toni, S., Meschi, F*., et al.* (2009). Maturity-onset diabetes of the young in children with incidental hyperglycemia: a multicenter Italian study of 172 families. Diabetes Care *32*, 1864-1866.
Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.
Lupo, A., Cesaro, E., Montano, G., Zurlo, D., Izzo, P., and Costanzo, P. (2013). KRAB-Zinc Finger Proteins: A Repressor Family Displaying Multiple Biological Functions. Curr Genomics *14*, 268-278.
Lynn, F.C., Smith, S.B., Wilson, M.E., Yang, K.Y., Nekrep, N., and German, M.S. (2007). Sox9 coordinates a transcriptional network in pancreatic progenitor cells. Proc Natl Acad Sci U S A *104*, 10500-10505.
Lyttle, B.M., Li, J., Krishnamurthy, M., Fellows, F., Wheeler, M.B., Goodyer, C.G., and Wang, R. (2008). Transcription factor expression in the developing human fetal endocrine pancreas. Diabetologia *51*, 1169-1180.
Mackeh, R., Marr, A.K., Fadda, A., and Kino, T. (2018). C2H2-Type Zinc Finger Proteins: Evolutionarily Old and New Partners of the Nuclear Hormone Receptors. Nucl Recept Signal *15*, 1550762918801071.
Macneil, L.T., and Walhout, A.J. (2011). Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. Genome Res *21*, 645-657.
Mahajan, A., Taliun, D., Thurner, M., Robertson, N.R., Torres, J.M., Rayner, N.W., Payne, A.J., Steinthorsdottir, V., Scott, R.A., Grarup, N*., et al.* (2018). Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat Genet *50*, 1505-1513.

Mahalakshmi, B., Baskaran, R., Shanmugavadivu, M., Nguyen, N.T., and Velmurugan, B.K. (2020). Insulinoma-associated protein 1 (INSM1): a potential biomarker and therapeutic target for neuroendocrine tumors. Cell Oncol (Dordr) *43*, 367-376.
Manning, K.S., and Cooper, T.A. (2017). The roles of RNA processing in translating genotype to phenotype. Nat Rev Mol Cell Biol *18*, 102-114.
Martinez-Contreras, R., Cloutier, P., Shkreta, L., Fisette, J.F., Revil, T., and Chabot, B. (2007). hnRNP proteins and splicing control. Adv Exp Med Biol *623*, 123-147.
Mastracci, T.L., Anderson, K.R., Papizan, J.B., and Sussel, L. (2013). Regulation of Neurod1 contributes to the lineage potential of Neurogenin3+ endocrine precursor cells in the pancreas. PLoS Genet *9*, e1003278.
Matsuoka, T.A., Artner, I., Henderson, E., Means, A., Sander, M., and Stein, R. (2004). The MafA transcription factor appears to be responsible for tissue-specific expression of insulin. Proc Natl Acad Sci U S A *101*, 2930-2933.
Matsuoka, T.A., Kaneto, H., Stein, R., Miyatsuka, T., Kawamori, D., Henderson, E., Kojima, I., Matsuhisa, M., Hori, M., and Yamasaki, Y. (2007). MafA regulates expression of genes important to islet beta-cell function. Mol Endocrinol *21*, 2764-2774.
McDonald, E., Li, J., Krishnamurthy, M., Fellows, G.F., Goodyer, C.G., and Wang, R. (2012). SOX9 regulates endocrine cell differentiation during human fetal pancreas development. Int J Biochem Cell Biol *44*, 72-83.
Mellitzer, G., Bonne, S., Luco, R.F., Van De Casteele, M., Lenne-Samuel, N., Collombat, P., Mansouri, A., Lee, J., Lan, M., Pipeleers, D.*, et al.* (2006). IA1 is NGN3-dependent and essential for differentiation of the endocrine pancreas. EMBO J *25*, 1344-1352.
Merkin, J., Russell, C., Chen, P., and Burge, C.B. (2012). Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. Science *338*, 1593-1599.
Mihola, O., Trachtulec, Z., Vlcek, C., Schimenti, J.C., and Forejt, J. (2009). A mouse speciation gene encodes a meiotic histone H3 methyltransferase. Science *323*, 373-375.
Miller, J., McLachlan, A.D., and Klug, A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. EMBO J *4*, 1609-1614.
Miller, K., Kim, A., Kilimnik, G., Jo, J., Moka, U., Periwal, V., and Hara, M. (2009). Islet formation during the neonatal development in mice. PLoS One *4*, e7739.
Mitchell, R.K., Nguyen-Tu, M.S., Chabosseau, P., Callingham, R.M., Pullen, T.J., Cheung, R., Leclerc, I., Hodson, D.J., and Rutter, G.A. (2017). The transcription factor Pax6 is required for pancreatic beta cell identity, glucose-regulated ATP synthesis, and Ca(2+) dynamics in adult mice. J Biol Chem *292*, 8892-8906.
Miyata, T., Maeda, T., and Lee, J.E. (1999). NeuroD is required for differentiation of the granule cells in the cerebellum and hippocampus. Genes Dev *13*, 1647-1652.
Moin, A.S.M., and Butler, A.E. (2019). Alterations in Beta Cell Identity in Type 1 and Type 2 Diabetes. Curr Diab Rep *19*, 83.
Moller, D.E., Yokota, A., Caro, J.F., and Flier, J.S. (1989). Tissue-specific expression of two alternatively spliced insulin receptor mRNAs in man. Mol Endocrinol *3*, 1263-1269.
Moore, J.M., Rabaia, N.A., Smith, L.E., Fagerlie, S., Gurley, K., Loukinov, D., Disteche, C.M., Collins, S.J., Kemp, C.J., Lobanenkov, V.V.*, et al.* (2012). Loss of maternal CTCF is associated with peri-implantation lethality of Ctcf null embryos. PLoS One *7*, e34915.
Moore, K.S., and von Lindern, M. (2018). RNA Binding Proteins and Regulation of mRNA Translation in Erythropoiesis. Front Physiol *9*, 910.
Moss, N.D., and Sussel, L. (2020). mRNA Processing: An Emerging Frontier in the Regulation of Pancreatic beta Cell Function. Front Genet *11*, 983.
Mosser, R.E., Maulis, M.F., Moulle, V.S., Dunn, J.C., Carboneau, B.A., Arasi, K., Pappan, K., Poitout, V., and Gannon, M. (2015). High-fat diet-induced beta-cell proliferation occurs prior to insulin resistance in C57Bl/6J male mice. Am J Physiol Endocrinol Metab *308*, E573-582.

Mourich, D.V., Oda, S.K., Schnell, F.J., Crumley, S.L., Hauck, L.L., Moentenich, C.A., Marshall, N.B., Hinrichs, D.J., and Iversen, P.L. (2014). Alternative splice forms of CTLA-4 induced by antisense mediated splice-switching influences autoimmune diabetes susceptibility in NOD mice. Nucleic Acid Ther *24*, 114-126.
Mukhopadhyay, S., Dermawan, J.K., Lanigan, C.P., and Farver, C.F. (2019). Insulinoma-associated protein 1 (INSM1) is a sensitive and highly specific marker of neuroendocrine differentiation in primary lung neoplasms: an immunohistochemical study of 345 cases, including 292 whole-tissue sections. Mod Pathol *32*, 100-109.
Murtaugh, L.C. (2007). Pancreas and beta-cell development: from the actual to the possible. Development *134*, 427-438.
Nagy, G., Milosevic, I., Fasshauer, D., Muller, E.M., de Groot, B.L., Lang, T., Wilson, M.C., and Sorensen, J.B. (2005). Alternative splicing of SNAP-25 regulates secretion through nonconservative substitutions in the SNARE domain. Mol Biol Cell *16*, 5675-5685.
Naujok, O., Burns, C., Jones, P.M., and Lenzen, S. (2011). Insulin-producing surrogate beta-cells from embryonic stem cells: are we there yet? Mol Ther *19*, 1759-1768.
Naya, F.J., Huang, H.P., Qiu, Y., Mutoh, H., DeMayo, F.J., Leiter, A.B., and Tsai, M.J. (1997). Diabetes, defective pancreatic morphogenesis, and abnormal enteroendocrine differentiation in BETA2/neuroD-deficient mice. Genes Dev *11*, 2323-2334.
Naya, F.J., Stellrecht, C.M., and Tsai, M.J. (1995). Tissue-specific regulation of the insulin gene by a novel basic helix-loop-helix transcription factor. Genes Dev *9*, 1009-1019.
Nelson, L.B., Spaeth, G.L., Nowinski, T.S., Margo, C.E., and Jackson, L. (1984). Aniridia. A review. Surv Ophthalmol *28*, 621-642.
Nishimura, W., Kondo, T., Salameh, T., El Khattabi, I., Dodge, R., Bonner-Weir, S., and Sharma, A. (2006). A switch from MafB to MafA expression accompanies differentiation to pancreatic beta-cells. Dev Biol *293*, 526-539.
Norris, A.D., and Calarco, J.A. (2012). Emerging Roles of Alternative Pre-mRNA Splicing Regulation in Neuronal Development and Function. Front Neurosci *6*, 122.
Nusslein-Volhard, C., Lohs-Schardin, M., Sander, K., and Cremer, C. (1980). A dorso-ventral shift of embryonic primordia in a new maternal-effect mutant of Drosophila. Nature *283*, 474-476.
Nusslein-Volhard, C., and Wieschaus, E. (1980). Mutations affecting segment number and polarity in Drosophila. Nature *287*, 795-801.
Offield, M.F., Jetton, T.L., Labosky, P.A., Ray, M., Stein, R.W., Magnuson, M.A., Hogan, B.L., and Wright, C.V. (1996). PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. Development *122*, 983-995.
Oliver-Krasinski, J.M., Kasner, M.T., Yang, J., Crutchlow, M.F., Rustgi, A.K., Kaestner, K.H., and Stoffers, D.A. (2009). The diabetes gene Pdx1 regulates the transcriptional network of pancreatic endocrine progenitor cells in mice. J Clin Invest *119*, 1888-1898.
Osipovich, A.B., Dudek, K.D., Greenfest-Allen, E., Cartailler, J.P., Manduchi, E., Potter Case, L., Choi, E., Chapman, A.G., Clayton, H.W., Gu, G.*, et al.* (2021). A developmental lineage-based gene co-expression network for mouse pancreatic beta-cells reveals a role for Zfp800 in pancreas development. Development *148*.
Osipovich, A.B., Long, Q., Manduchi, E., Gangula, R., Hipkens, S.B., Schneider, J., Okubo, T., Stoeckert, C.J., Jr., Takada, S., and Magnuson, M.A. (2014). Insm1 promotes endocrine cell differentiation by modulating the expression of a network of genes that includes Neurog3 and Ripply3. Development *141*, 2939-2949.
Osmark, P., Hansson, O., Jonsson, A., Ronn, T., Groop, L., and Renstrom, E. (2009). Unique splicing pattern of the TCF7L2 gene in human pancreatic islets. Diabetologia *52*, 850-854.
Ostenson, C.G., Gaisano, H., Sheu, L., Tibell, A., and Bartfai, T. (2006). Impaired gene and protein expression of exocytotic soluble N-ethylmaleimide attachment protein receptor complex proteins in pancreatic islets of type 2 diabetic patients. Diabetes *55*, 435-440.
Pagliuca, F.W., Millman, J.R., Gurtler, M., Segel, M., Van Dervort, A., Ryu, J.H., Peterson, Q.P., Greiner, D., and Melton, D.A. (2014). Generation of functional human pancreatic beta cells in vitro. Cell *159*, 428-439.

Pan, F.C., Brissova, M., Powers, A.C., Pfaff, S., and Wright, C.V. (2015). Inactivating the permanent neonatal diabetes gene Mnx1 switches insulin-producing beta-cells to a delta-like fate and reveals a facultative proliferative capacity in aged beta-cells. Development *142*, 3637-3648.
Pan, F.C., and Wright, C. (2011). Pancreas organogenesis: from bud to plexus to gland. Dev Dyn *240*, 530-565.
Panneerselvam, A., Kannan, A., Mariajoseph-Antony, L.F., and Prahalathan, C. (2019). PAX proteins and their role in pancreas. Diabetes Res Clin Pract *155*, 107792.
Papizan, J.B., Singer, R.A., Tschen, S.I., Dhawan, S., Friel, J.M., Hipkens, S.B., Magnuson, M.A., Bhushan, A., and Sussel, L. (2011). Nkx2.2 repressor complex regulates islet beta-cell specification and prevents beta-to-alpha-cell reprogramming. Genes Dev *25*, 2291-2305.
Pauklin, S., and Vallier, L. (2013). The cell-cycle state of stem cells determines cell fate propensity. Cell *155*, 135-147.
Pedersen, N., Mortensen, S., Sorensen, S.B., Pedersen, M.W., Rieneck, K., Bovin, L.F., and Poulsen, H.S. (2003). Transcriptional gene expression profiling of small cell lung cancer cells. Cancer Res *63*, 1943-1953.
Petri, A., Ahnfelt-Ronne, J., Frederiksen, K.S., Edwards, D.G., Madsen, D., Serup, P., Fleckner, J., and Heller, R.S. (2006). The effect of neurogenin3 deficiency on pancreatic gene expression in embryonic mice. J Mol Endocrinol *37*, 301-316.
Pihlajamaki, J., Lerin, C., Itkonen, P., Boes, T., Floss, T., Schroeder, J., Dearie, F., Crunkhorn, S., Burak, F., Jimenez-Chillaron, J.C*., et al.* (2011). Expression of the splicing factor gene SFRS10 is reduced in human obesity and contributes to enhanced lipogenesis. Cell Metab *14*, 208-218.
Pongubala, J.M.R., and Murre, C. (2021). Spatial Organization of Chromatin: Transcriptional Control of Adaptive Immune Cell Development. Front Immunol *12*, 633825.
Pradella, D., Naro, C., Sette, C., and Ghigna, C. (2017). EMT and stemness: flexible processes tuned by alternative splicing in development and cancer progression. Mol Cancer *16*, 8.
Qu, X., Afelik, S., Jensen, J.N., Bukys, M.A., Kobberup, S., Schmerr, M., Xiao, F., Nyeng, P., Veronica Albertoni, M., Grapin-Botton, A*., et al.* (2013). Notch-mediated post-translational control of Ngn3 protein stability regulates pancreatic patterning and cell fate commitment. Dev Biol *376*, 1-12.
Rabiee, B., Anwar, K.N., Shen, X., Putra, I., Liu, M., Jung, R., Afsharkhamseh, N., Rosenblatt, M.I., Fishman, G.A., Liu, X*., et al.* (2020). Gene dosage manipulation alleviates manifestations of hereditary PAX6 haploinsufficiency in mice. Sci Transl Med *12*.
Raghupathy, N., Choi, K., Vincent, M.J., Beane, G.L., Sheppard, K.S., Munger, S.C., Korstanje, R., Pardo-Manual de Villena, F., and Churchill, G.A. (2018). Hierarchical analysis of RNA-seq reads improves the accuracy of allele-specific expression. Bioinformatics *34*, 2177-2184.
Raj, B., O'Hanlon, D., Vessey, J.P., Pan, Q., Ray, D., Buckley, N.J., Miller, F.D., and Blencowe, B.J. (2011). Cross-regulation between an alternative splicing activator and a transcription repressor controls neurogenesis. Mol Cell *43*, 843-850.
Rankin, S.A., McCracken, K.W., Luedeke, D.M., Han, L., Wells, J.M., Shannon, J.M., and Zorn, A.M. (2018). Timing is everything: Reiterative Wnt, BMP and RA signaling regulate developmental competence during endoderm organogenesis. Dev Biol *434*, 121-132.
Reichert, M., and Rustgi, A.K. (2011). Pancreatic ductal cells in development, regeneration, and neoplasia. J Clin Invest *121*, 4572-4578.
Rezania, A., Bruin, J.E., Arora, P., Rubin, A., Batushansky, I., Asadi, A., O'Dwyer, S., Quiskamp, N., Mojibian, M., Albrecht, T*., et al.* (2014). Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. Nat Biotechnol *32*, 1121-1133.
Rivera-Perez, J.A., and Hadjantonakis, A.K. (2014). The Dynamics of Morphogenesis in the Early Mouse Embryo. Cold Spring Harb Perspect Biol *7*.
Roberts, R. (1967). Small eyes—a new dominant eye mutant in the mouse. Genetical Research *9*, 121-122.
Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat Biotechnol *29*, 24-26.

Romer, A.I., Singer, R.A., Sui, L., Egli, D., and Sussel, L. (2019). Murine Perinatal beta-Cell Proliferation and the Differentiation of Human Stem Cell-Derived Insulin-Expressing Cells Require NEUROD1. Diabetes *68*, 2259-2271.
Rosenbaum, J.N., Duggan, A., and Garcia-Anoveros, J. (2011). Insm1 promotes the transition of olfactory progenitors from apical and proliferative to basal, terminally dividing and neuronogenic. Neural Dev *6*, 6.
Rukstalis, J.M., and Habener, J.F. (2009). Neurogenin3: a master regulator of pancreatic islet differentiation and regeneration. Islets *1*, 177-184.
Rutter, W.J., Kemp, J.D., Bradshaw, W.S., Clark, W.R., Ronzio, R.A., and Sanders, T.G. (1968). Regulation of specific protein synthesis in cytodifferentiation. J Cell Physiol *72*, Suppl 1:1-18.
Sachdeva, M.M., and Stoffers, D.A. (2009). Minireview: Meeting the demand for insulin: molecular mechanisms of adaptive postnatal beta-cell mass expansion. Mol Endocrinol *23*, 747-758.
Sadoul, K., Lang, J., Montecucco, C., Weller, U., Regazzi, R., Catsicas, S., Wollheim, C.B., and Halban, P.A. (1995). SNAP-25 is expressed in islets of Langerhans and is involved in insulin release. J Cell Biol *128*, 1019-1028.
Sanchez, R., and Zhou, M.M. (2011). The PHD finger: a versatile epigenome reader. Trends Biochem Sci *36*, 364-372.
Sander, M., Neubuser, A., Kalamaras, J., Ee, H.C., Martin, G.R., and German, M.S. (1997). Genetic analysis reveals that PAX6 is required for normal transcription of pancreatic hormone genes and islet development. Genes Dev *11*, 1662-1673.
Schaffer, A.E., Taylor, B.L., Benthuysen, J.R., Liu, J., Thorel, F., Yuan, W., Jiao, Y., Kaestner, K.H., Herrera, P.L., Magnuson, M.A.*, et al.* (2013). Nkx6.1 controls a gene regulatory network required for establishing and maintaining pancreatic Beta cell identity. PLoS Genet *9*, e1003274.
Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat Methods *9*, 671-675.
Schwitzgebel, V.M., Scheel, D.W., Conners, J.R., Kalamaras, J., Lee, J.E., Anderson, D.J., Sussel, L., Johnson, J.D., and German, M.S. (2000). Expression of neurogenin3 reveals an islet cell precursor population in the pancreas. Development *127*, 3533-3542.
Scoville, D., Lichti-Kaiser, K., Grimm, S., and Jetten, A. (2019). GLIS3 binds pancreatic beta cell regulatory regions alongside other islet transcription factors. J Endocrinol.
Seino, S., Seino, M., Nishi, S., and Bell, G.I. (1989). Structure of the human insulin receptor gene and characterization of its promoter. Proc Natl Acad Sci U S A *86*, 114-118.
Senee, V., Chelala, C., Duchatelet, S., Feng, D., Blanc, H., Cossec, J.C., Charon, C., Nicolino, M., Boileau, P., Cavener, D.R.*, et al.* (2006). Mutations in GLIS3 are responsible for a rare syndrome with neonatal diabetes mellitus and congenital hypothyroidism. Nat Genet *38*, 682-687.
Sever, D., Hershko-Moshe, A., Srivastava, R., Eldor, R., Hibsher, D., Keren-Shaul, H., Amit, I., Bertuzzi, F., Krogvold, L., Dahl-Jorgensen, K.*, et al.* (2021). NF-kappaB activity during pancreas development regulates adult beta-cell mass by modulating neonatal beta-cell proliferation and apoptosis. Cell Death Discov *7*, 2.
Seymour, P.A., and Serup, P. (2019). Mesodermal induction of pancreatic fate commitment. Semin Cell Dev Biol *92*, 77-88.
Sharma, A., Halu, A., Decano, J.L., Padi, M., Liu, Y.Y., Prasad, R.B., Fadista, J., Santolini, M., Menche, J., Weiss, S.T.*, et al.* (2018). Controllability in an islet specific regulatory network identifies the transcriptional factor NFATC4, which regulates Type 2 Diabetes associated genes. NPJ Syst Biol Appl *4*, 25.
Sharma, A., Zangen, D.H., Reitz, P., Taneja, M., Lissauer, M.E., Miller, C.P., Weir, G.C., Habener, J.F., and Bonner-Weir, S. (1999). The homeodomain protein IDX-1 increases after an early burst of proliferation during pancreatic regeneration. Diabetes *48*, 507-513.
Sharon, N., Vanderhooft, J., Straubhaar, J., Mueller, J., Chawla, R., Zhou, Q., Engquist, E.N., Trapnell, C., Gifford, D.K., and Melton, D.A. (2019). Wnt Signaling Separates the Progenitor and Endocrine Compartments during Pancreas Development. Cell Rep *27*, 2281-2291 e2285.
Sheaffer, K.L., and Kaestner, K.H. (2012). Transcriptional networks in liver and intestinal development. Cold Spring Harb Perspect Biol *4*, a008284.

Sheets, T.P., Park, K.E., Park, C.H., Swift, S.M., Powell, A., Donovan, D.M., and Telugu, B.P. (2018). Targeted Mutation of NGN3 Gene Disrupts Pancreatic Endocrine Cell Development in Pigs. Sci Rep *8*, 3582.
Shih, H.P., Wang, A., and Sander, M. (2013). Pancreas organogenesis: from lineage determination to morphogenesis. Annu Rev Cell Dev Biol *29*, 81-105.
Singer, R.A., Arnes, L., Cui, Y., Wang, J., Gao, Y., Guney, M.A., Burnum-Johnson, K.E., Rabadan, R., Ansong, C., Orr, G.*, et al.* (2019). The Long Noncoding RNA Paupar Modulates PAX6 Regulatory Activities to Promote Alpha Cell Development and Function. Cell Metab *30*, 1091-1106 e1098.
Singh, U., Malik, M.A., Goswami, S., Shukla, S., and Kaur, J. (2016). Epigenetic regulation of human retinoblastoma. Tumour Biol *37*, 14427-14441.
Slack, J.M. (1995). Developmental biology of the pancreas. Development *121*, 1569-1580.
Smith, S.B., Gasa, R., Watada, H., Wang, J., Griffen, S.C., and German, M.S. (2003). Neurogenin3 and hepatic nuclear factor 1 cooperate in activating pancreatic expression of Pax4. J Biol Chem *278*, 38254-38259.
Sommer, L., Ma, Q., and Anderson, D.J. (1996). neurogenins, a novel family of atonal-related bHLH transcription factors, are putative mammalian neuronal determination genes that reveal progenitor cell heterogeneity in the developing CNS and PNS. Mol Cell Neurosci *8*, 221-241.
Sonawane, A.R., Weiss, S.T., Glass, K., and Sharma, A. (2019). Network Medicine in the Age of Biomedical Big Data. Front Genet *10*, 294.
Sosa-Pineda, B., Chowdhury, K., Torres, M., Oliver, G., and Gruss, P. (1997). The Pax4 gene is essential for differentiation of insulin-producing beta cells in the mammalian pancreas. Nature *386*, 399-402.
Soufi, A., and Dalton, S. (2016). Cycling through developmental decisions: how cell cycle dynamics control pluripotency, differentiation and reprogramming. Development *143*, 4301-4311.
Soyer, J., Flasse, L., Raffelsberger, W., Beucher, A., Orvain, C., Peers, B., Ravassard, P., Vermot, J., Voz, M.L., Mellitzer, G.*, et al.* (2010). Rfx6 is an Ngn3-dependent winged helix transcription factor required for pancreatic islet cell development. Development *137*, 203-212.
Spitali, P., and Aartsma-Rus, A. (2012). Splice modulating therapies for human disease. Cell *148*, 1085-1088.
Spoel, S.H. (2018). Orchestrating the proteome with post-translational modifications. J Exp Bot *69*, 4499-4503.
St-Onge, L., Sosa-Pineda, B., Chowdhury, K., Mansouri, A., and Gruss, P. (1997). Pax6 is required for differentiation of glucagon-producing alpha-cells in mouse pancreas. Nature *387*, 406-409.
Stapleton, P., Weith, A., Urbanek, P., Kozmik, Z., and Busslinger, M. (1993). Chromosomal localization of seven PAX genes and cloning of a novel family member, PAX-9. Nat Genet *3*, 292-298.
Stemple, D.L. (2005). Structure and function of the notochord: an essential organ for chordate development. Development *132*, 2503-2512.
Sujjitjoon, J., Kooptiwut, S., Chongjaroen, N., Tangjittipokin, W., Plengvidhya, N., and Yenchitsomanus, P.T. (2016). Aberrant mRNA splicing of paired box 4 (PAX4) IVS7-1G>A mutation causing maturity-onset diabetes of the young, type 9. Acta Diabetol *53*, 205-216.
Sun, D.G., Yang, J.H., Tong, Y., Zhao, G.J., and Ma, X. (2008). [A novel PAX6 mutation (c.1286delC) in the patients with hereditary congenital aniridia.]. Yi Chuan *30*, 1301-1306.
Sund, N.J., Vatamaniuk, M.Z., Casey, M., Ang, S.L., Magnuson, M.A., Stoffers, D.A., Matschinsky, F.M., and Kaestner, K.H. (2001). Tissue-specific deletion of Foxa2 in pancreatic beta cells results in hyperinsulinemic hypoglycemia. Genes Dev *15*, 1706-1715.
Sussel, L., Kalamaras, J., Hartigan-O'Connor, D.J., Meneses, J.J., Pedersen, R.A., Rubenstein, J.L., and German, M.S. (1998). Mice lacking the homeodomain transcription factor Nkx2.2 have diabetes due to arrested differentiation of pancreatic beta cells. Development *125*, 2213-2221.
Swisa, A., Avrahami, D., Eden, N., Zhang, J., Feleke, E., Dahan, T., Cohen-Tayar, Y., Stolovich-Rain, M., Kaestner, K.H., Glaser, B.*, et al.* (2017). PAX6 maintains beta cell identity by repressing genes of alternative islet cell types. J Clin Invest *127*, 230-243.
Talchai, C., Xuan, S., Lin, H.V., Sussel, L., and Accili, D. (2012). Pancreatic beta cell dedifferentiation as a mechanism of diabetic beta cell failure. Cell *150*, 1223-1234.

Taneera, J., Lang, S., Sharma, A., Fadista, J., Zhou, Y., Ahlqvist, E., Jonsson, A., Lyssenko, V., Vikman, P., Hansson, O*., et al.* (2012). A systems genetics approach identifies genes and pathways for type 2 diabetes in human islets. Cell Metab *16*, 122-134.
Tao, W., Zhang, Y., Ma, L., Deng, C., Duan, H., Liang, X., Liao, R., Lin, S., Nie, T., Chen, W*., et al.* (2018). Haploinsufficiency of Insm1 Impairs Postnatal Baseline beta-Cell Mass. Diabetes *67*, 2615-2625.
Thomas, J.H., and Emerson, R.O. (2009). Evolution of C2H2-zinc finger genes revisited. BMC Evol Biol *9*, 51.
Thomas, J.H., and Schneider, S. (2011). Coevolution of retroelements and tandem zinc finger genes. Genome Res *21*, 1800-1812.
Thompson, D., Regev, A., and Roy, S. (2015). Comparative analysis of gene regulatory networks: from network reconstruction to evolution. Annu Rev Cell Dev Biol *31*, 399-428.
Tian, W., Zhu, X.R., Qiao, C.Y., Ma, Y.N., Yang, F.Y., Zhou, Z., Feng, J.P., Sun, R., Xie, R.R., Lu, J*., et al.* (2021). Heterozygous PAX6 mutations may lead to hyper-proinsulinaemia and glucose intolerance: A case-control study in families with congenital aniridia. Diabet Med *38*, e14456.
Ton, C.C., Hirvonen, H., Miwa, H., Weil, M.M., Monaghan, P., Jordan, T., van Heyningen, V., Hastie, N.D., Meijers-Heijboer, H., Drechsler, M*., et al.* (1991). Positional cloning and characterization of a paired box- and homeobox-containing gene from the aniridia region. Cell *67*, 1059-1074.
Trujillo, C.A., Rice, E.S., Schaefer, N.K., Chaim, I.A., Wheeler, E.C., Madrigal, A.A., Buchanan, J., Preissl, S., Wang, A., Negraes, P.D*., et al.* (2021). Reintroduction of the archaic variant of NOVA1 in cortical organoids alters neurodevelopment. Science *371*.
Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S., and Gregory, P.D. (2010). Genome editing with engineered zinc finger nucleases. Nat Rev Genet *11*, 636-646.
Valladolid-Acebes, I., Daraio, T., Brismar, K., Harkany, T., Ogren, S.O., Hokfelt, T.G., and Bark, C. (2015). Replacing SNAP-25b with SNAP-25a expression results in metabolic disease. Proc Natl Acad Sci U S A *112*, E4326-4335.
van der Meer-de Jong, R., Dickinson, M.E., Woychik, R.P., Stubbs, L., Hetherington, C., and Hogan, B.L. (1990). Location of the gene involving the small eye mutation on mouse chromosome 2 suggests homology with human aniridia 2 (AN2). Genomics *7*, 270-275.
Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: function, expression and evolution. Nat Rev Genet *10*, 252-263.
Vilas, C.K., Emery, L.E., Denchi, E.L., and Miller, K.M. (2018). Caught with One's Zinc Fingers in the Genome Integrity Cookie Jar. Trends Genet *34*, 313-325.
Villate, O., Turatsinze, J.V., Mascali, L.G., Grieco, F.A., Nogueira, T.C., Cunha, D.A., Nardelli, T.R., Sammeth, M., Salunkhe, V.A., Esguerra, J.L*., et al.* (2014). Nova1 is a master regulator of alternative splicing in pancreatic beta cells. Nucleic Acids Res *42*, 11818-11830.
Waddington, C.H., and Kacser, H. (1957). The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology (Allen & Unwin).
Walther, C., Guenet, J.L., Simon, D., Deutsch, U., Jostes, B., Goulding, M.D., Plachov, D., Balling, R., and Gruss, P. (1991). Pax: a murine multigene family of paired box-containing genes. Genomics *11*, 424-434.
Wang, H., Gauthier, B.R., Hagenfeldt-Johansson, K.A., Iezzi, M., and Wollheim, C.B. (2002). Foxa2 (HNF3beta) controls multiple genes implicated in metabolism-secretion coupling of glucose-induced insulin release. J Biol Chem *277*, 17564-17570.
Wang, H., Maechler, P., Antinozzi, P.A., Hagenfeldt, K.A., and Wollheim, C.B. (2000). Hepatocyte nuclear factor 4alpha regulates the expression of pancreatic beta -cell genes implicated in glucose metabolism and nutrient-induced insulin secretion. J Biol Chem *275*, 35953-35959.
Wang, S., Jensen, J.N., Seymour, P.A., Hsu, W., Dor, Y., Sander, M., Magnuson, M.A., Serup, P., and Gu, G. (2009). Sustained Neurog3 expression in hormone-expressing islet cells is required for endocrine maturation and function. Proc Natl Acad Sci U S A *106*, 9715-9720.
Wang, S., Yan, J., Anderson, D.A., Xu, Y., Kanal, M.C., Cao, Z., Wright, C.V., and Gu, G. (2010). Neurog3 gene dosage regulates allocation of endocrine and exocrine cell fates in the developing mouse pancreas. Dev Biol *339*, 26-37.

Wang, Y., and Wang, Z. (2014). Systematical identification of splicing regulatory cis-elements and cognate trans-factors. Methods *65*, 350-358.
Wang, Z., York, N.W., Nichols, C.G., and Remedi, M.S. (2014). Pancreatic beta cell dedifferentiation in diabetes and redifferentiation following insulin therapy. Cell Metab *19*, 872-882.
Watanabe, N., Hiramatsu, K., Miyamoto, R., Yasuda, K., Suzuki, N., Oshima, N., Kiyonari, H., Shiba, D., Nishio, S., Mochizuki, T*., et al.* (2009). A murine model of neonatal diabetes mellitus in Glis3-deficient mice. FEBS Lett *583*, 2108-2113.
Weinberg, N., Ouziel-Yahalom, L., Knoller, S., Efrat, S., and Dor, Y. (2007). Lineage tracing evidence for in vitro dedifferentiation but rare proliferation of mouse pancreatic beta-cells. Diabetes *56*, 1299-1304.
Weintraub, H. (1993). The MyoD family and myogenesis: redundancy, networks, and thresholds. Cell *75*, 1241-1244.
Wells, J.M., and Melton, D.A. (1999). Vertebrate endoderm development. Annu Rev Cell Dev Biol *15*, 393-410.
Wen, J.H., Chen, Y.Y., Song, S.J., Ding, J., Gao, Y., Hu, Q.K., Feng, R.P., Liu, Y.Z., Ren, G.C., Zhang, C.Y*., et al.* (2009). Paired box 6 (PAX6) regulates glucose metabolism via proinsulin processing mediated by prohormone convertase 1/3 (PC1/3). Diabetologia *52*, 504-513.
Whittaker, J., Sorensen, H., Gadsboll, V.L., and Hinrichsen, J. (2002). Comparison of the functional insulin binding epitopes of the A and B isoforms of the insulin receptor. J Biol Chem *277*, 47380-47384.
Wickramasinghe, V.O., and Venkitaraman, A.R. (2016). RNA Processing and Genome Stability: Cause and Consequence. Mol Cell *61*, 496-505.
Wildner, H., Gierl, M.S., Strehle, M., Pla, P., and Birchmeier, C. (2008). Insm1 (IA-1) is a crucial component of the transcriptional network that controls differentiation of the sympatho-adrenal lineage. Development *135*, 473-481.
Witten, J.T., and Ule, J. (2011). Understanding splicing regulation through RNA splicing maps. Trends Genet *27*, 89-97.
Xu, E.E., Krentz, N.A., Tan, S., Chow, S.Z., Tang, M., Nian, C., and Lynn, F.C. (2015). SOX4 cooperates with neurogenin 3 to regulate endocrine pancreas formation in mouse models. Diabetologia *58*, 1013-1023.
Yang, C., Coker, K.J., Kim, J.K., Mora, S., Thurmond, D.C., Davis, A.C., Yang, B., Williamson, R.A., Shulman, G.I., and Pessin, J.E. (2001). Syntaxin 4 heterozygous knockout mice develop muscle insulin resistance. J Clin Invest *107*, 1311-1318.
Yang, M., Dai, J., Jia, Y., Suo, L., Li, S., Guo, Y., Liu, H., Li, L., and Yang, G. (2014). Overexpression of juxtaposed with another zinc finger gene 1 reduces proinflammatory cytokine release via inhibition of stress-activated protein kinases and nuclear factor-kappaB. FEBS J *281*, 3193-3205.
Yasuda, T., Kajimoto, Y., Fujitani, Y., Watada, H., Yamamoto, S., Watarai, T., Umayahara, Y., Matsuhisa, M., Gorogawa, S., Kuwayama, Y*., et al.* (2002). PAX6 mutation as a genetic factor common to aniridia and glucose intolerance. Diabetes *51*, 224-230.
Zhang, C., Moriguchi, T., Kajihara, M., Esaki, R., Harada, A., Shimohata, H., Oishi, H., Hamada, M., Morito, N., Hasegawa, K*., et al.* (2005). MafA is a key regulator of glucose-stimulated insulin secretion. Mol Cell Biol *25*, 4969-4976.
Zhang, T., Liu, W.D., Saunee, N.A., Breslin, M.B., and Lan, M.S. (2009). Zinc finger transcription factor INSM1 interrupts cyclin D1 and CDK4 binding and induces cell cycle arrest. J Biol Chem *284*, 5574-5581.
Zhang, X., Ye, P., Huang, H., Wang, B., Dong, F., and Ling, Q. (2020). TCF7L2 rs290487 C allele aberrantly enhances hepatic gluconeogenesis through allele-specific changes in transcription and chromatin binding. Aging (Albany NY) *12*, 13365-13387.
Zhao, B.S., Roundtree, I.A., and He, C. (2017). Post-transcriptional gene regulation by mRNA modifications. Nat Rev Mol Cell Biol *18*, 31-42.
Zhou, Q., Law, A.C., Rajagopal, J., Anderson, W.J., Gray, P.A., and Melton, D.A. (2007). A multipotent progenitor domain guides pancreatic organogenesis. Dev Cell *13*, 103-114.
Zhou, Z., and Fu, X.D. (2013). Regulation of splicing by SR proteins and SR protein-specific kinases. Chromosoma *122*, 191-207.

Zhu, M., Breslin, M.B., and Lan, M.S. (2002). Expression of a novel zinc-finger cDNA, IA-1, is associated with rat AR42J cells differentiation into insulin-positive cells. Pancreas *24*, 139-145.
Zorn, A.M., and Wells, J.M. (2009). Vertebrate endoderm development and organ formation. Annu Rev Cell Dev Biol *25*, 221-251.