NGN3-EXPRESSING PROGENITOR HETEROGENEITY DRIVES ENDOCRINE LINEAGE ALLOCATION IN PANCREAS DEVELOPMENT

By

Jing Liu

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

December 2015

Nashville, Tennessee

Approved:

Guoqiang Gu, PhD

Chris Wright, PhD

Mark Magnuson, MD

David Piston, PhD

Andrea Page-McCaw, PhD

ACKNOWLEDGEMENT

This work would not have been possible without the help from current and previous Gu lab members. Yanwen Xu and Ruiying Hu are two seasoned research assistants who always provide valuable technical support for my experiments. Chen is constantly exchanging ideas with me and providing suggestions for my project. Previous Gu lab member Jingbo Yan and Heather Funderburg helped me when I first joined the Gu lab by teaching me various experimental techniques and made my transition to a graduate student a pleasant process. I am blessed to have these wonderful colleagues to work with in the past five years. Above all, I would especially like to thank Dr. Guoqiang Gu for his invaluable discussions of experimental design, endless troubleshooting experiments with me, as well as mentoring my writing and presentation skills. Dr. Gu is an intelligent and industrious researcher who still does benchwork himself and I respect him as a real scientist.

I am full of gratitude to my committee members: Dr. Chris Wright, Dr. Mark Magnuson, Dr. David Piston, Dr. Andrea Page-McCaw, and a previous committee member Dr. Lance Prince for their insightful inputs to my project and helping me navigate through the graduate school. Many times I bothered them with committee meetings and solicited recommendation letters and they have always been supportive. My committee chair, Dr. Wright, is one of the most brilliant people I have ever met. He has the gift of explaining things clearly and concisely and has been a fantastic mentor over the years.

I would also like to thank the Vanderbilt diabetes and beta cell research community, especially the Wright Lab members, who are wonderful colleagues that create such a collaborative research environment. Dr. Fong-Cheng Pan is both an instructor who helped me with experiments and a friend with whom I shared many

memorable moments. Dr. Matthew Bechard and Eric Bankaitis provided helpful discussions and proofread my writings many times. Dr. Yu-Ping Yang helped me to set up my FACS assay and shared her protocol and tricks with me. The Vanderbilt Cell Imaging Shared Resource, Flow Cytometry Core, Vantage Sequencing Center, and ES/Transgenic Core provide excellent technical support.

This work was funded by the National Institute of Diabetes and Digestive and Kidney Diseases, the Juvenile Diabetes Research Foundation, and the Beta Cell Biology Consortium. My research would not have been possible without their financial support.

My thesis was proofread by Dr. Guoqiang Gu. I appreciate his time and bearing with my writing. I received more revise comments than I expected, which made the final version much better than the first draft. Were there any errors or inaccuracies, they are all mine.

And lastly, I am grateful to Han Zuo, Longlei Zhang, Eason Chan, Karen Mok, and many others for their unending inspiration. I have met many amazing people in Nashville and beyond in the last five years. Many of them become lifelong friends. Their encouragement and support is the source of my persistence and achievement today.

Many thanks to my parents.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Abstract

This chapter presents the background knowledge of this thesis research. First, pancreas anatomy and physiology are presented, with the function and regulation of endocrine hormones highlighted. Then, pancreas development is discussed. A few key transcription factors and their roles in pancreas morphogenesis and lineage allocation are covered in detail. Next, epidemiology and pathogenesis of diabetes mellitus, as well as its management, are briefly talked about. As gene and cell therapy is a promising route to the replenishment of lost β cells and restoration of euglycemia in diabetic patients, a subsequent section is devoted to β-cell regeneration and reprogramming, summarizing the most recent discoveries on this topic. This thesis research will contribute to our understanding of the molecular pathway of endocrine differentiation and shed light on new therapy development in the future, albeit far from clinical application at the moment. Lastly, because a large volume of this thesis research centers on lineage tracing and the Cre/loxP methodology, I also wrote a section to introduce cell lineage tracing, the Cre/loxP system, its variants and development in the last decade.

1.2 Overview of Pancreas Physiology

The human pancreas is a compound organ with both endocrine and exocrine functions. Anatomically, the pancreas resides in the abdominal cavity, behind the

stomach and underneath the liver (Fig 1.1 a). The exocrine pancreas is composed of acinar cells and duct cells. Acinar cells produce and secrete various digestive enzymes, such as lipase, protease, amylase and nuclease, *etc*. Duct cells form a highly branched transportation network, which ultimately merges into the main pancreatic duct. The main pancreatic duct in turn merges with the common bile duct from the gallbladder, and then enters the duodenum via the ampulla of Vater. Digestion enzymes and bicarbonate ions are thus delivered into the duodenum to aid food digestion (Shih et al., 2013; Slack, 1995).

The endocrine pancreas plays a more important role in maintaining metabolic homeostasis and is the main focus of my thesis. The endocrine pancreas is composed of endocrine cells, including α, β, δ, PP, and ε cells (Fig 1.1 c). These cells secrete glucagon, insulin, somatostatin, pancreatic polypeptide (PP) and ghrelin, respectively. The endocrine cells aggregate and form islets of Langerhans, which are the functional units of endocrine pancreas. Islets of Langerhans are scattered in the exocrine tissue and account for only 1-2% of total pancreas mass. The mouse islets have a well-defined spatial distribution of endocrine cells, with β cells in the center and other endocrine cells in the periphery. However, such a spatial architecture is not so obvious in human islets. In human islets, endocrine cells are mixed together with no obvious central vs marginal distinction (Bosco et al., 2010; Dai et al., 2012). In the adult mouse pancreas, β cells make up 60-80% of total endocrine cells; α cells represent 15-20%; δ cells take 5-10% and the rest of the endocrine cells take ~2% (Edlund, 2002). It should be noted that endocrine-cell percentages change over development and under different physiological conditions, as well as vary from species to species (Steiner et al., 2010). Islets are infiltrated by blood vessels and nerves. Secreted pancreatic endocrine hormones enter the circulation system through the capillary vessels and are transported to various target organs or tissues. On the other hand, the secretion activity of endocrine cells is subject

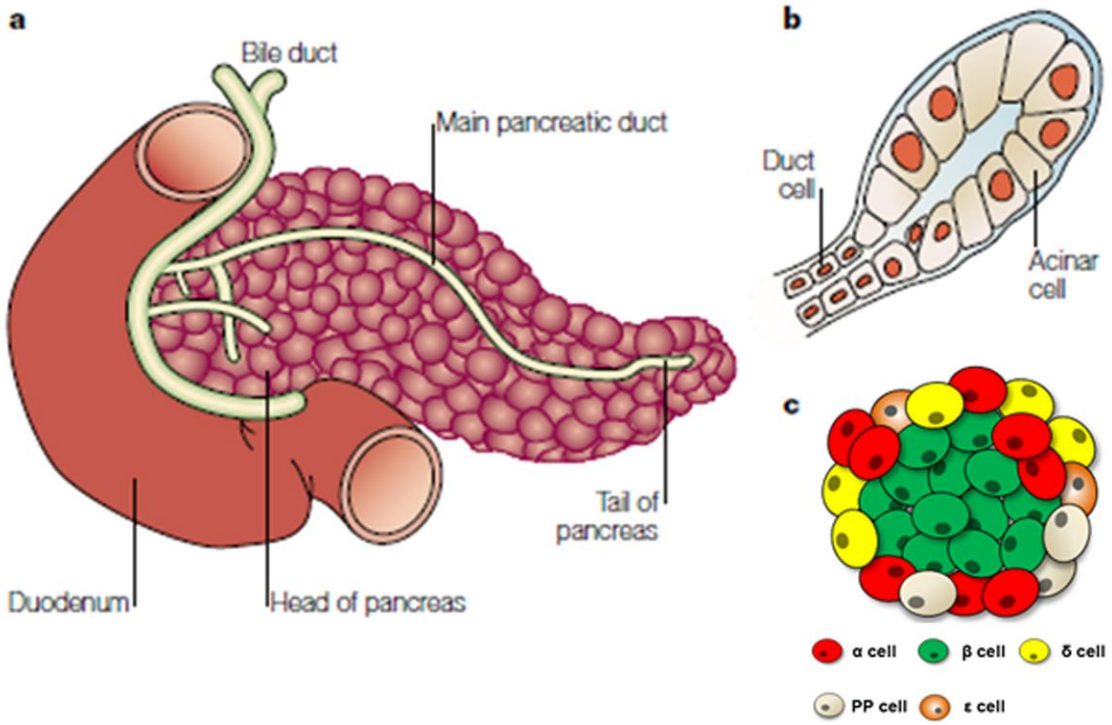


Fig 1.1 Anatomy of the mouse pancreas. (a) The pancreas resides right next to the duodenum. Main pancreatic duct and bile duct from the gallbladder merge and enter the duodenum. The pancreas comprises roughly two parts by their proximity to the duodenum, the head and the tail. (b) Exocrine pancreas, including duct cells and acinar cells. (c) An islet, showing α, β, δ, PP, and ε cells. Figure is adapted from Edlund, 2002.

to the regulation by circulating hormones from the bloodstream, as well as the nervous system (Chandra and Liddle, 2009).

The insulin-secreting β cells receive the most attention from researchers. The main function of β cells is to maintain blood glucose homeostasis through the secretion of insulin. The insulin mRNA is translated as a preproinsulin peptide and its maturation into insulin requires C-peptide removal and disulfide-bond formation (Davidson, 2004; Fu et al., 2013). Groups of six insulin molecules then assemble into a stable hexamer and are stored in vesicles readily to be released upon secretogog stimulation (Dunn, 2005). The primary insulin secretagog is glucose. When blood glucose level rises, the glucose transporter 2 (Glut2) on β-cell membrane takes up glucose, which undergoes glycolysis and mitochondrial oxidation, leading to an increased ATP/ADP ratio. Increased ATP/ADP ratio shuts down ATP-sensitive K_{ATP} channels and results in plasma membrane depolarization, which in turn opens voltage-gated Ca^{2+} channels and causes Ca^{2+} influx. Elevated cytosolic Ca^{2+} concentration triggers the fusion of insulin vesicles with plasma membrane and eventually insulin release. This vesicle fusion and secretion is a SNARE (*SNA*P (soluble NSF attachment protein) *RE*ceptor)-dependent process and possibly regulated by Ca^{2+}-sensing synaptotagmins (Ahren, 2009; Fu et al., 2013; Rorsman et al., 2000; Wang and Thurmond, 2009). In addition to this pathway, G protein-coupled receptors (GPCRs) transduce hormone (e.g., glucagon-like peptide 1, or GLP1) and metabolic signals (e.g., free fatty acids) into elevated second messengers, such as cyclic AMP (cAMP), diacylglycerol (DAG), and inositol-1,4,5-triphosphate (IP_3), which eventually activate protein kinase A (PKA) and protein kinase C (PKC) signaling pathways as well as Ca^{2+} release from the endoplasmic reticulum (ER) reservoir to modulate insulin secretion (Ahren, 2009; Blad et al., 2012).

Secreted insulin circulates to other body parts and exerts its anabolic function on target organs or tissues, including liver, skeletal muscle, and fat tissue (Saltiel and Kahn,

2001). The action of insulin involves many molecules and different signaling pathways. In brief, binding of insulin to insulin receptor, a tyrosine receptor kinase, leads to the receptor's autophosphorylation and activation. Activated insulin receptor initiates a cascade of phosphorylation events, leading to the activation of insulin receptor substrates (IRS), MAP kinase, PI3K/Akt, mTOR, and PKC signaling pathways, *etc*. These pathways act concertedly to enhance anabolic metabolism (glucose uptake, glycogen synthesis, lipid synthesis, protein synthesis, *etc*.), inhibit catabolic metabolism (gluconeogenesis, glycogenolysis, *etc*.), as well as regulate cell growth and differentiation (Saltiel and Kahn, 2001; Taniguchi et al., 2006).

Glucagon, on the other hand, antagonizes the function of insulin by promoting hepatic glucose production to prevent hypoglycemia. The coordination between glucagon and insulin maintains blood glucose level under tight control. In type II diabetic patients, the glucagon level is unexpectedly elevated and exacerbates the hyperglycemia resulting from insulin insufficiency and resistance (D'Alessio, 2011). It has become more evident in recent years that glucagon and α cells also play a role in diabetes pathology and they start to come into the center of research that has long been dominated by insulin and β bells (Burcelin et al., 2008; D'Alessio, 2011; Del Prato and Marchetti, 2004; Quesada et al., 2008). Besides its catabolic function as an endocrine hormone, glucagon also exerts a function directly on β cells to promote insulin secretion in a paracrine fashion, complementary to the regulation of insulin secretion by the blood glucose level, circulating hormones, and the nervous system. Concomitantly, insulin inhibits glucagon secretion, forming a feedback loop to keep blood glucose level under tight control (Elliott et al., 2015). Somatostatin, on the other hand, inhibits both glucagon and insulin secretion through binding to somatostatin receptors on α and β cells (Schwetz et al., 2013; Strowski and Blake, 2008). It is recently reported that β cell-secreted urocortin3 augments somatostatin secretion from δ cells (van der Meulen et al.,

2015). Such bidirectional regulation among endocrine cells establishes a fine-tuned feedback system, which enables islets to maintain metabolic homeostasis (Caicedo, 2013).

1.3 Overview of Pancreas Development

The pancreas develops from a region in the foregut endoderm epithelium (Fig 1.2) (Collombat et al., 2006; Edlund, 2002; Pan and Wright, 2011; Rieck et al., 2012; Romer and Sussel, 2015). The first observable morphological structure is the two pancreatic bud evaginations, which appear at around E9.5 and are marked by the expression of a homeobox transcription factor *Pdx1* (pancreatic and duodenal homeobox 1, also known as insulin-promoter-factor 1 or *Ipf1*) (Hale et al., 2005; Offield et al., 1996) (Fig 1.2 A, A', A''). These two $Pdx1^+$ pancreatic buds further proliferate and invade into the surrounding mesenchyme, forming ventral pancreas and dorsal pancreas. The two buds rotate towards each other to form a single organ as pancreatic development progresses (Fig 1.2 B'). In mouse, disruption of the *Pdx1* gene causes pancreatic agenesis (Jonsson et al., 1994; Offield et al., 1996). Cell lineage tracing showed that $Pdx1^+$ cells are the early multipotent progenitor cells (MPCs) that give birth to all pancreatic cell types, including duct cells, acinar cells, and endocrine cells (Gu et al., 2002). The expression of *Sox9* colocalizes with *Pdx1* expression before E12.5 and it is thus considered as an early MPC maker, which is supported by lineage tracing results with $Sox9^{Cre}$ and $Sox9^{CreER}$ mouse models (Akiyama et al., 2005; Furuyama et al., 2011; Kopp et al., 2011). Consequently, deletion of *Sox9* in the developing pancreas leads to the arrest of MPC expansion and pancreatic hypoplasia (Seymour et al., 2007). *Ptf1a*, another transcription factor that labels the MPCs, is critical in committing the pancreatic fate from the gut endoderm. Lineage tracing demonstrates that *Ptf1a* is expressed in the early

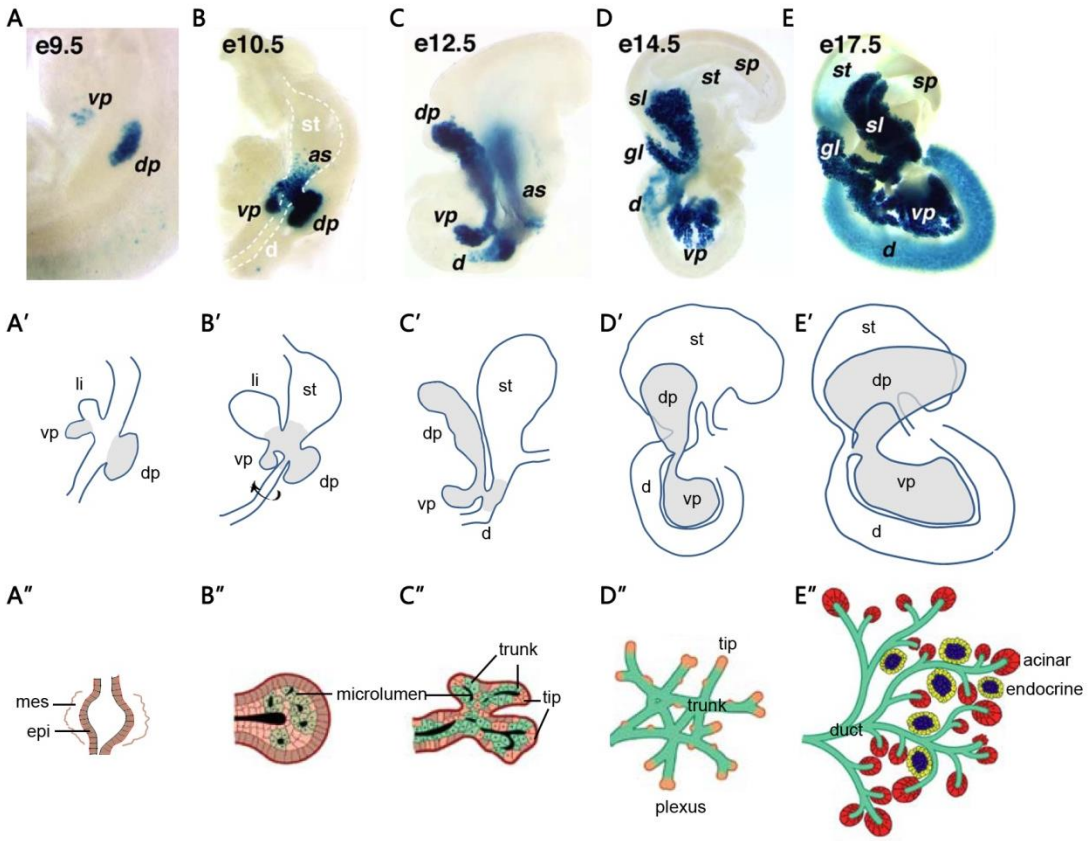


Fig 1.2 Overview of pancreas development. The pancreas develops from a pre-patterned region in the primitive gut endoderm epithelium. At E9.5, two observable pancreatic buds grow and invade into surrounding mesenchyme, forming ventral pancreas (vp) and dorsal pancreas (dp). Ventral pancreas and dorsal pancreas rotate towards each other and eventually form a single organ. As development progresses, $Ngn3^{+}$ progenitors delaminate from the duct epithelium and differentiate into endocrine cells, which further aggregate into islets of Langerhans. (A-E): whole mount staining of beta-galactosidase from the $Pdx1^{tTA/+}; Tg^{tetO\text{-}Pdx1\text{-}lacZ}$ mouse embryos, showing the expression pattern of Pdx1. Figures are adapted from Hale et al. 2005. (A'-E'): schematic representation of pancreas development. Liver, gallbladder and spleen are omitted from E12.5 on for visual clarity. li: liver. vp: ventral pancreas. dp: dorsal pancreas. st: stomach. d: duodenum. Shaded area represents $Pdx1^{+}$ area. (A''-E''): schematic drawing showing key cellular events. mes: mesenchyme. epi: epithelium. Figures are adapted from Pan and Wright, 2011.

pancreatic progenitor cells that eventually differentiate into duct, acinar and endocrine cells (Kawaguchi et al., 2002). *Ptf1a*-deficient mice manifest pancreas agenesis and the expansion of the duodenal epithelium domain, suggesting that *Ptf1a* regulates the adoption between pancreatic versus gut endoderm fate (Kawaguchi et al., 2002). Zhou et al. used a *Cpa1*^{CreER} mouse model to show that carboxypeptidase A1 (Cpa1)-positive cells are multipotent before E14.5. They proposed a "tip-trunk" model in which the pancreatic MPC population is located at the tips of the branching pancreatic epithelium and is characterized by Pdx1^{+}Ptf1a^{+}Cpa1^{+} expression (Zhou et al., 2007) (Fig 1.2 C'',D''; Fig 1.3). Starting from E14.5, the expression of *Ptf1a* and *Cpa1* becomes restricted to the tip compartment, which produces the eventual acinar cells, and the differentiation potential of the Ptf1a^{+} or Cpa1^{+} cells is progressively constrained to the acinar cell fate. The trunk compartment, on the other hand, harbors the Pdx1^{low}Sox9^{hi}Ptf1a^{-}Cpa1^{-} bipotent cells that give birth to duct cells and endocrine cells (Schaffer et al., 2010) (Fig 1.3). The downregulation of *Ptf1a* and *Cpa1* expression is necessary for the acquisition of the bipotent progenitor cell fate (Pan et al., 2013). Hnf1β is another marker for the MPCs before E13.5. From E13.5 to E15.5, Hnf1β is highly expressed in the trunk domain and marks the bipotent progneitors, in contrast to Ptf1a and Cpa1 (Solar et al., 2009).

A portion of the bipotent progenitors turn on the expression of a basic helix-loop-helix (bHLH) transcription factor *Neurogenin 3* (*Ngn3*). These Ngn3^{+} cells mark the earliest endocrine progenitors. Ngn3^{+} progenitors delaminate from the duct epithelium and initiate a cascade of gene activation/inactivation events, leading to their differentiation into different endocrine cell lineages (Fig 1.3). Much effort has been dedicated to the study of *Ngn3* and demonstrated its central role in endocrine pancreas development. *Ngn3*-null mice produce almost no endocrine cells (Gradwohl et al., 2000). Conversely, ectopic expression of *Ngn3* leads to the precocious differentiation of

pancreatic precursors into endocrine cells at the expense of exocrine lineages (Apelqvist et al., 1999; Schwitzgebel et al., 2000). Lineage tracing proves that *Ngn3*-expressing cells give rise to all endocrine cells (Gu et al., 2002; Schonhoff et al., 2004). It is notable, however, that *Ngn3* expressing level is also crucial in determining endocrine versus exocrine cell fate. In the *Ngn3*-haploinsufficient or null conditions, the proportion of acinar and duct cells increases at the expense of endocrine cells (Wang et al., 2010). Immunostaining reveals two $Ngn3^{+}$ populations, $Ngn3^{low}$ and $Ngn3^{high}$ progenitors, by us and other researchers (Seymour et al., 2008; Shih et al., 2012). A recent study of inducing human embryonic stem cells (hESCs) to differentiate into endocrine cells discovers that hESCs with *Ngn3* being knocked-out by CRISPR/Cas9 (CRISPR: clustered regularly interspaced short palindromic repeats; Cas9: CRISPR-associated protein 9) fail to differentiate into endocrine cells but hESCs with shRNA-mediated *Ngn3* knockdown can form endocrine cells, though fewer than unmanipulated hESCs (McGrath et al., 2015). The different results from these two experiments can be explained by the fact that shRNA-mediated gene knockdown does not fully erase the expression of *Ngn3* as compared to the CRISPR/Case9-mediated gene knockout method. Thus, the shRNA-mediated *Ngn3* knockdown experiment phenocopies an *Ngn3*-haploinsufficient situation and produces an intermediate amount of endocrine cells (McGrath et al., 2015). Together, these studies suggest the importance of *Ngn3* expression level in endocrine specification in both mice and humans.

Because of the importance of *Ngn3* and its expression level in endocrine pancreas development, *Ngn3* expression is subject to the regulation of many factors and signaling pathways. For instance, Sox9 (Lynn et al., 2007b), Pdx1 (Oliver-Krasinski et al., 2009) and Hnf6 (Jacquemin et al., 2000) activate *Ngn3* expression, while Notch signaling inhibits *Ngn3* expression (Apelqvist et al., 1999; Lee et al., 2001). Notch inhibition limits the $Ngn3^{+}$ progenitor pool size and thus keeps the endocrine and

exocrine lineages at balance, which is exemplified by many experiments perturbing Notch signaling (Apelqvist et al., 1999; Jensen et al., 2000). *Ngn3* is expressed in the endocrine progenitors and diminishes in endocrine cells soon after birth, although it is also suggested that *Ngn3* is expressed at a very low level in the adult islet cells and this sustained low level expression of *Ngn3* is required for maintaining islet function (Wang et al., 2009a). An *in vitro* experiment demonstrates the self-inhibitory ability of *Ngn3* (Smith et al., 2004). Once *Ngn3* reaches a high expression level, it binds to its own promoter and prevents itself from further expression. This phenomenon is used to explain the reduced expression of *Ngn3* after lineage commitment. Nonetheless, this result comes from an *in vitro* assay and may not represent the *in vivo* situation. We instead found that *Ngn3* can augment its own expression, possibly by inducing the expression of certain miRNAs that can tune down Notch signaling components Hes1 and Psen1 and thus remove Notch inhibition on *Ngn3* expression. The regulation of *Ngn3* expression will be discussed in detail in Chapter V.

Although $Ngn3^{+}$ progenitors as a population produce all types of endocrine cells, it is unclear how an individual $Ngn3^{+}$ progenitor makes its decision to become one endocrine cell type versus another. To this end, it is found that different endocrine cell types are not born randomly during pancreatic morphogenesis; rather, each endocrine cell type has its time window of production (Johansson et al., 2007). Using an $Ngn3^{ERTM}$ "add-back" mouse model, Johansson et al. was able to activate the exogenous $Ngn3^{ERTM}$ within defined time windows in the $Ngn3^{-/-}$ mouse pancreas and thus examine the differentiation potential of stage-specific $Ngn3^{+}$ progenitors. They found that $Ngn3^{+}$ progenitors give birth to α cells first, starting at as early as E9.5 and lasting to E14.5. A major wave of β-cell differentiation happens during the secondary transition, between E12.5-E16.5. δ and PP cells emerge in late gestational stages (Johansson et al., 2007). A competence window model based on this observation of successive but partially

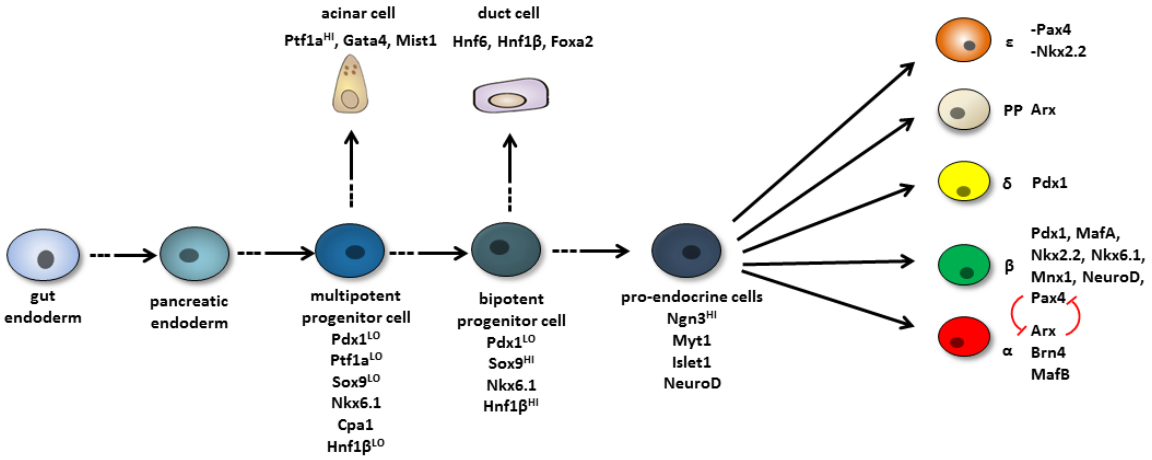


Fig 1.3 Gene transcription cascade in pancreas organogenesis. Schematic representation of stepwise pancreatic lineage specification. Selected important transcription factors at each stage are indicated. The antagonizing *Pax4* and *Arx* are emphasized with red lines. Modified from Pan and Wright, 2011.

overlapping differentiation of endocrine cells is proposed, stating that $Ngn3^+$ endocrine progenitors pass through consecutive temporal windows during which their differentiation competence changes (Johansson et al., 2007). Johansson et al. also found that the competence of $Ngn3^+$ progenitors is intrinsic to the epithelium but not influenced by signaling from the surrounding mesenchyme. When they recombined embryonic pancreatic epithelium and mesenchyme in a stage-matched or heterochronic manner for an *in vitro* culture assay, they discovered that the distribution of endocrine cells is similar between groups if the pancreatic epithelia used are of the same stage, regardless of the mesenchyme used (Johansson et al., 2007). This suggests that the competence shift is intrinsic to the pancreas epithelium rather than due to instructive signals from the surrounding mesenchyme. However, it should not be overlooked that Ffg10 signaling from the mesenchyme activates Notch signaling in the epithelium, which in turn inhibits the acquisition of the pro-endocrine marker Ngn3. Thus, Fgf10 signaling maintains the proliferation of $Pdx1^+$ progenitors to ensure an ample source of progenitors in the very beginning of endocrine differentiation (Bhushan et al., 2001; Norgaard et al., 2003).

The detailed mechanism of how the competence windows are established, maintained, and changed is largely unknown. It is well possible that upstream regulators, including Pdx1, Sox9, Hnf6, as well as Notch signaling and Fgf10 signaling from the mesenchyme concertedly modulate the expression of Ngn3 and subsequently its downstream targets, which form a transcription factor network that defines the status of the $Ngn3^+$ progenitors and the differentiation pathways they can adopt (Jensen, 2004; Wilson et al., 2003). Many transcription factors have been validated as *Ngn3* downstream targets, including *Arx* (Collombat et al., 2003), *Pax4* (Smith et al., 2003), *NeuroD* (Huang et al., 2000), *Nkx2.2* (Watada et al., 2003), and *Insm1* (Mellitzer et al., 2006; Osipovich et al., 2014) *etc*. Among these, *Arx* and *Pax4* have received most

investigation and they are found to specify α-cell and β/δ-cell lineages, respectively. *Arx* and *Pax4* are co-expressed in the early endocrine progenitors but their expression becomes mutually exclusive and inhibits the expression of one other, committing the endocrine progenitors to different lineages (Collombat et al., 2005). *Pax4* expression is gradually restricted to β and δ lineages and is switched off after birth; *Arx* expression, on the other hand, persists in mature α cells. The mutual repressive relationship between *Arx* and *Pax4* has been interrogated with many loss-of-function and ectopic expression experiments (Collombat et al., 2005; Collombat et al., 2007; Collombat et al., 2003; Collombat et al., 2009; Courtney et al., 2013; Dhawan et al., 2011). For instance, *Pax4* knockout results in the loss of β and δ lineages (Sosa-Pineda, 2004). In contrast, loss of *Arx* leads to increased β and δ cells at the expense of α cells (Collombat et al., 2003). Not surprisingly, *Pax4* ectopic expression in α cells converts them to β cells *in vivo* (Collombat et al., 2009) while misexpression of *Arx* in β cells leads to their conversion to glucagon$^+$ and PP$^+$ cells (Collombat et al., 2007). The homeodomain transcription factor Nkx2.2 is necessary for both β-cell specification and maintenance (Doyle and Sussel, 2007; Sussel et al., 1998). *NeuroD*, a basic helix-loop-helix transcription factor and an *Ngn3* downstream target (Huang et al., 2000), is required for endocrine differentiation. Loss of *NeuroD* results in the reduction of all endocrine cells and a reduced β-to-α and β-to-δ cell ratio, suggesting the significance of *NeuroD* in regulating endocrine differentiation, especially β-cell lineage specification (Naya et al., 1997). Nkx2.2-deficient mice display hyperglycemia due to the lack of insulin-secreting β cells and die at neonatal stages (Sussel et al., 1998). In β cells, Nkx2.2 can recruit a repressor complex composed of Groucho 3 (Grg3), DNA methyltransferase 3a (Dnmt3a) and histone deacetylase 1 (Hdac1) to enhance the methylation of the upstream regulatory elements of *Arx* gene and thus prohibit the acquisition of α-cell fate (Mastracci et al., 2011; Papizan et al., 2011). Like Nkx2.2, the homeodomain transcription factor family member

Nkx6.1 is also shown to directly repress the expression of *Arx* and thus responsible for β-cell specification (Schaffer et al., 2013). The zinc finger transcription factor Myt1 forms a feed-forward loop with Ngn3 as the two promote the expression of each other (Wang et al., 2008). Loss of *Myt1* in the developing pancreas results in abnormal multi-hormone positive cells, suggesting that endocrine differentiation or maturation is disrupted (Wang et al., 2007). *Insm1*, another *Ngn3* downstream gene (Mellitzer et al., 2006), is also important in regulating pancreatic endocrine differentiation through a gene network that involves cell adhesion, cell migration, extracellular matrix remodeling, cell proliferation, and mRNA alternative splicing *etc*. Loss of *Insm1* decreases the delamination of pro-endocrine progenitors, leads to the alternative splicing of Ngn3 mRNA, and decreases β-cell production (Osipovich et al., 2014). Besides transcription factors, a gene that encodes a secreted protein, connective tissue growth factor (CTGF), also influences endocrine lineage allocation. Deletion of *CTGF* results in reduced β-to-α cell ratio and altered islet morphology in the adult mice (Crawford et al., 2009). The basic-leucine-zipper MafA and MafB transcription factors, though not employed in lineage specification, are important to endocrine cell maturation. In the developing mouse pancreas, insulin$^+$ cells switch from MafB$^+$ to MafA$^+$ with the concomitant high *Pdx1* expression, marking the maturation of β cells (Artner et al., 2010; Nishimura et al., 2006). Reviews by Jensen (Jensen, 2004) and Pan (Pan and Wright, 2011) provide comprehensive summaries of transcription factors involved in pancreas development.

Ngn3 not only dictates endocrine differentiation but also affects duct branching morphogenesis. Unlike other organs such as lung, kidney and mammary gland, pancreatic ductal network does not develop from the elongation and bifurcation of a single duct tube. Instead, a plexus comprising numerous small lumens first forms and gradually remodels and coalesces into a single-lumen ductal system (Iber and Menshykau, 2013; Villasenor et al., 2010) (Fig 1.2 B'', C''). Loss of *Ngn3* leads to

reduced branching and dilated pancreatic plexus. In the opposite, $Ngn3^+$ domain expansion by inhibiting Notch signaling results in thinning plexus (Magenheim et al., 2011). It is speculated that endocrine differentiation and ductal morphogenesis are coupled events and perturbation of pancreatic branching morphogenesis could influence endocrine differentiation, which is an active research topic (Pan and Wright, 2011; Rieck et al., 2012).

In summary, pancreas development initiates from two $Pdx1^+$ pancreatic buds in the foregut endoderm. The multipotent $Pdx1^+Ptf1a^+Cpa1^+$ progenitor cells bifurcate into the acinar cell lineage and a $Pdx1^{low}Sox9^{hi}Ptf1a^-Cpa1^-Hnf1\beta^+$ bipotent progenitor population. The bipotent progenitors further diverge into the duct cell lineage and the endocrine lineage mainly based on the activation of *Ngn3*, the endocrine master gene. Differentiation of $Ngn3^+$ pro-endocrine progenitors into various endocrine lineages depends on the concerted interaction of various *Ngn3* downstream genes, whose activation/inactivation determines which endocrine lineage $Ngn3^+$ progenitors are able to commit. Nonetheless, the pancreatic endocrine differentiation process is far from well understood. When and how the $Ngn3^+$ progenitors are specified to different endocrine cell lineages is one of the many mysteries, and this is the main focus of this thesis research.

1.4 Diabetes and Significance of This Study

Diabetes is a major health concern in the US and worldwide. According to the data of American Diabetes Association, the United States has witnessed a steady increase of diabetic cases during the last 50 years. By 2012, 29.1 million children and adults, roughly 9.3% of the total US population, had diabetes (http://www.diabetes.org/diabetes-basics/diabetes-statistics/). Worldwide, the prevalence

of diabetes was about 9% among adults above eighteen years old in 2014. WHO predicts that in 2030 diabetes will become the 7th leading cause of death worldwide and 4th in high-income countries (http://www.who.int/mediacentre/factsheets/fs312/en/).

There are two main types of diabetes: type I and type II (Zimmet et al., 2001). Type I diabetes is characterized by the loss of β cells and thus insulin insufficiency. It is generally believed that genetic susceptibility and environmental triggers induce β-cell auto-immune response, characterized by the presentation of autoantigens (insulin, glutamic acid decarboxylase (GAD65), tyrosine phosphatase islet antigen 2 (IA2 or ICA512)) and subsequent emergence of autoantibodies and autoreactive T cells capable of recognizing these autoantigens and destructing β cells (Atkinson, 2012; Knip et al., 2005; Van Belle et al., 2011). Immunosuppressive drugs like cyclosporine delay type I diabetes progression but not prevent it (Van Belle et al., 2011). Type I diabetic patients eventually require insulin injection to compensate for the loss of β cells and maintain blood glucose within physiological range. Recently, intestinal K cells are engineered to express insulin. By taking advantage of the K cells' glucose-responsive machinery and their ability to escape autoimmune attack, researchers are able to protect non-obese diabetic (NOD) mice with an insulin transgene in the K cells from developing diabetes (Mojibian et al., 2014). The ethical controversy over human genome manipulation, however, precludes its clinical application in the short term. Type II diabetes is the dominant type of diabetes (~90% of diabetes cases) and is characterized by insulin resistance in insulin-responsive organs or tissues. The pathogenesis of type II diabetes is a result of many factors, including genetics, nutrition and lifestyle, *etc*. (Ashcroft and Rorsman, 2012). An overly simplified view holds that nutrient overload exposes tissues to deleterious metabolic intermediates, activates the endoplasmic reticulum (ER) stress and innate immune pathways, and eventually disrupts insulin signaling and causes insulin resistance, although the detail mechanisms are more sophisticated (Muoio and

Newgard, 2008; Samuel and Shulman, 2012). While calorie-rich diet style and lack of physical exercise are well-known risk factors of type II diabetes, recent research has also identified more than a dozen of single nucleotide polymorphisms (SNPs) associated with increased risk of type II diabetes by using genome-wide association study (GWAS) analysis (Ashcroft and Rorsman, 2012). These susceptible genes include *Tcf7l2, Slc30a8, Kcnq1, etc*., demonstrating the involvement of genetics in type II diabetes pathogenesis, though the mechanisms for these susceptible genes remain elusive. The current treatment for type II diabetes enhances insulin secretion from β cells to counteract insulin resistance, as well as targets periphery organs to inhibit gluconeogenesis. For instance, sulfonylurea targets K_{ATP} channels of β cells to increase glucose-stimulated insulin secretion; exendin-4 targets GLP1 receptor to enhance insulin secretion; metformin targets the liver to inhibit glucose production, *etc*. (Moller, 2001; Park et al., 2007; Viollet and Foretz, 2013). In late stage type II diabetes, impaired β-cell function such as glucose-stimulated insulin secretion (GSIS) and loss of β-cell mass arise as they become exhausted from excessive insulin secretion to counteract insulin resistance. This poses a new challenge and requires intervention with insulin injection (Ashcroft and Rorsman, 2012).

Unfortunately, there is no cure for diabetes currently and diabetic patients have to take medicines and/or insulin injections regularly as well as adjusting their diet and lifestyle, which all have notorious patient compliance issues (Beckman et al., 2002; Grundy et al., 2005). Islet transplantation-based therapy is a promising direction and will cure insulin-dependent diabetes once and for all. In the next section, I will summarize recent progresses in the field of β-cell regeneration, reprogramming and the limitations of these studies at the moment.

1.5 Beta-cell Regeneration and Reprogramming as Diabetes Therapy

In order to replenish β cells in diabetic patients, there are several strategies: increase the proliferation of existing β cells, induce the differentiation of endocrine progenitors in the adult pancreas, transplant cadaveric islets, produce transplantable β cells from *in vitro* differentiation of human embryonic stem cells (hESCs) or induced pluripotent stem cells (iPSCs), or convert other cell types into β cells (Bonner-Weir and Weir, 2005; Desgraz et al., 2011; Pagliuca and Melton, 2013; Ziv et al., 2013).

One remedy for β-cell loss is accelerating β-cell replication. β cells are mostly generated during embryogenesis and the perinatal stage. In adult mice, β-cell proliferation is limited and decreases as the mice age (German, 2013; Pagliuca and Melton, 2013). Various factors have been found to enhance β-cell proliferation, including GLP1/exendin-4, betatrophin, CTGF, *etc*. (Riley et al., 2015; Stoffers et al., 2000; Xu et al., 1999; Yi et al., 2013), although controversies regarding their capacity of promoting β-cell replication in humans still remain (Burcelin and Dejager, 2010; Espes et al., 2014; Jiao et al., 2014; Wang et al., 2013). Some small molecules are also found through high-throughput screening to promote β-cell replication (Wang et al., 2015; Wang et al., 2009b). These small molecules could be potentially used to treat diabetes but more rigorous studies should be carried out in *in vivo* mouse models and isolated human islets before they are introduced into clinical trials. Meanwhile, we should be wary of their potential tumorigenesis effects.

Another way of replenishing β cells is through the differentiation of $Ngn3^{+}$ progenitors. In adult mice, β-cell turnover rate is very low and its replenishment is completed by the proliferation of existing β cells but not differentiation from endocrine progenitors, as *Ngn3* expression is low in adult islets (Dor et al., 2004). However, Xu et al. reported the reactivation of *Ngn3* expression program and a two-fold increase of

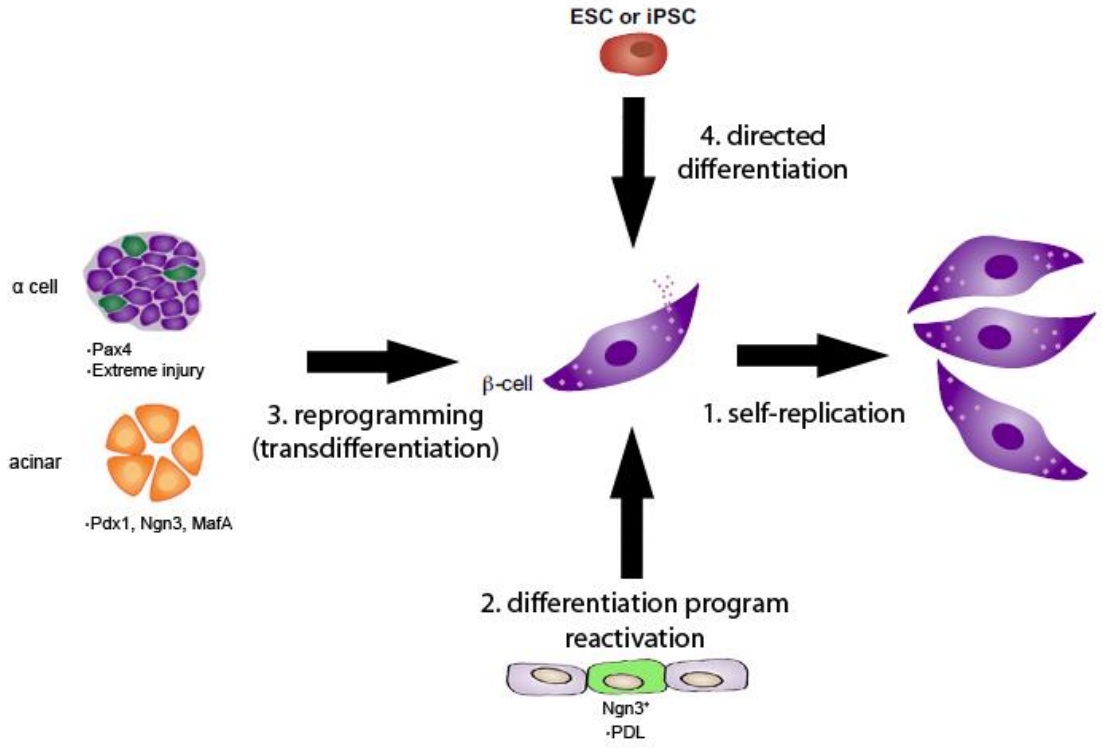


Fig 1.4 Sources of β-cell regeneration. Several strategies of β-cell regeneration are proposed: (1) increase the proliferation of existing β cells; (2) induce the differentiation of endocrine progenitors, although the existence of dormant $Ngn3^{+}$ progenitors in adults is controversial (see text); (3) convert other cell types into β cells either *in vivo* or *in vitro*; (4) *in vitro* differentiation of β cells from hESCs or iPSCs. Figure is adapted from Pagliuca and Melton, 2013.

insulin$^+$ cell mass in a partial pancreatic duct ligation (PDL) mouse model (Xu et al., 2008) . These Ngn3$^+$ cells line the duct epithelium and are considered as latent endocrine progenitors in adult mice. It is also reported that *Ngn3* is re-expressed in the duct epithelial cells upon α-to-β conversion mediated by *Pax4* ectopic expression (Al-Hasani et al., 2013). Nevertheless, whether PDL awakens endocrine progenitor differentiation program and increases actual β-cell number has been questioned. Some studies suggest that β cells are not regenerated in the PDL mouse model (Rankin et al., 2013) or that no endocrine/β cells are differentiated from these resurging Ngn3$^+$ cells (Kopp et al., 2011; Xiao et al., 2013).

The third strategy of compensating β-cell loss is islet/β-cell transplantation. The Edmonton group led by Dr. Shapiro et al. published clinical trial results in which type I diabetic patients become insulin-independent after islet transplantation in conjugation with a glucocorticoid-free immunosuppressive regime. Some patients remain insulin-independent for more than two years (Shapiro et al., 2000; Shapiro et al., 2006). Although this strategy is dampened by the difficulty of islet transplantation procedure, necessity of immunosuppression regimen to suppress host rejection, gradual loss of islet transplants, as well as scarcity of islet source (Merani and Shapiro, 2006; Rother and Harlan, 2004), this pioneer work represents the initial success that islet transplantation could be a promising cure for diabetes. Besides islet/β-cell transplantation, it is also notable that subcutaneous implantation of embryonic brown adipose tissue (BAT) can reverse hyperglycemia in type I diabetic mouse models without increasing β cells or insulin level (Gunawardana and Piston, 2012, 2015). The use of embryonic tissue is controversial and attempts to use BAT stem cells or BAT-secreted adipokines shall be made in the future. This topic will not be discussed in detail.

To overcome the islet/β-cell scarcity obstacle, scientists have developed various protocols to generate β cells *in vitro*. In 2006, D'Amour and colleagues were able to

generate pancreatic endocrine cells from hESCs by using a stepwise induction protocol but the resulted β-like insulin-secreting cells respond poorly to glucose stimulation, indicative of functional immaturity (D'Amour et al., 2006). A few years later, the same group generated glucose-responsive insulin-secreting cells after transplanting hESCs-derived pancreatic endoderm into immunocompromised mice. Most importantly, these cells protect mice from streptozotocin (STZ)-induced diabetes (Kroon et al., 2008). However, the transplanted pancreatic endoderm requires as long as three months of maturation time to become responsive to glucose and secret insulin (Kroon et al., 2008; Rezania et al., 2012). This maturation phase is ill-understood and it is doubtful whether this process could be replicated in humans. The Melton group reported a scalable method to differentiate β cells from hESCs (Pagliuca et al., 2014). These cells respond well to repeated glucose stimulation and reverse hyperglycemia in NRG-Akita mice within as fast as 18 days. Nonetheless, the authors also admitted that human insulin secreted in these mice is lower than that of mice transplanted with human cadaveric islets on a per cell basis, revealing the gaps of functionality between induced and authentic β cells. One possibility is that the presence of other endocrine cell types and proper vascularization is necessary to fine-tune the function of hESC-derived β cells (Brissova et al., 2014; Reinert et al., 2013; Rodriguez-Diaz et al., 2011). To this end, amid the current studies mostly focusing on *in vitro* generation of β cells, the differentiation of other endocrine cell type, three-dimensional tissue construction, and proper encapsulation and delivery strategies should be given equal attention in order to achieve therapeutic effectiveness. Meanwhile, the risk of teratoma formation accompanying hESC- or iPSC-based therapy should not be overlooked if the derived β cells are not fully differentiated and purified before transplantation.

β cells can also be generated from terminally differentiated cells through transdifferentiation *in vivo* or *in vitro*. The source cells are full of choices but usually

etiologically related to β cells. For example, α cells could transdifferentiate into β cells in a diphtheria toxin-induced β-cell ablation model (Thorel et al., 2010) or in a forced *Pdx1* expression model (Yang et al., 2011). Another notable example involves the antagonizing *Arx* and *Pax4*, which determine α and β/δ cell fate, respectively. *Pax4* ectopic expression in α cells converts them into β cells *in vivo* (Collombat et al., 2009). On the contrary, misexpression of *Arx* in β cells leads to their conversion to glucagon$^+$ and PP$^+$ cells (Collombat et al., 2007). Moreover, remodeling epigenetic architecture with small molecules such as adenosine periodate oxidized (Adox) is also able to reprogram α cells to β cells (Bramswig et al., 2013). In addition to endocrine cells, exocrine cells also demonstrate the plasticity of being converted into other cell types. Duct cells contribute to endocrine cells when the latter is destroyed by diphtheria toxin or when *Ngn3* is ectopically expressed in the duct cells (Criscimanna et al., 2011; Heremans et al., 2002; Lee et al., 2013). After partial pancreatic duct ligation, acinar cells can convert into duct cell and eventually endocrine/β cells (Pan et al., 2013). A cocktail of *Ngn3*, *Pdx1*, and *MafA* can reprogram adult mouse pancreatic exocrine cells to β cells *in vivo* (Zhou et al., 2008). Instead of genetic manipulation, a transient cytokine treatment can also reprogram acinar cells to β cells and restore euglycemia in diabetic mouse models (Baeyens et al., 2014). In addition to pancreatic cell types, human hepatocytes (Zalzman et al., 2005) and mouse fibroblasts (Li et al., 2014a) are also converted into β cells with *in vitro* protocols.

The significance of β-cell reprogramming research is multi-fold. First and foremost, *in vitro* generated β cells provide an affluent source for clinical transplantation, bypassing the issue of limited cadaveric islet/β-cell donors. Second, *in vitro* generated β cells cause less immune rejection if the source cells are the patient's own hESCs, iPSCs or terminally differentiated cells as compared to cadaveric islets from orthogonal donors. This improves β cell engraftment and obviates the use of immunosuppressive reagents,

although autoimmune reaction in type I diabetic patients remains an issue to be resolved. Third, induced β cells not only can be used for transplantation but also serve as drug screening platforms, obviating the use of scarce human islets or mouse islet substitutes. Whether we are to regenerate β cells from the replication of existing β cells, direct differentiation of hESCs/iPSCs, or transdifferentiation, we need to first of all have a better understanding of pancreas development and β-cell generation during embryogenesis. This thesis work focuses on the transcription factors *Ngn3* and *Myt1* and their roles in endocrine lineage allocation. I hope my research will contribute to our current knowledge of endocrine lineage specification and foster the informed design of β-cell regeneration protocols.

1.6 Introduction to Lineage tracing and the Cre/loxP Technique

For all multi-cellular organisms, how one cell proliferates and differentiates into a mature organism is always a fascinating topic. In order to understand the cell lineage relationships, researchers have developed various methods to track cell fate (Kretzschmar and Watt, 2012). Early cell lineage tracing methods include direct observation, label retaining, retroviral transduction, and tissue transplantation, *etc*. One of the most notable achievements in developmental biology is the mapping of *C. elegans* cell lineages by 1983 with time-lapse microscopy and differential interference contrast microscopy (Sulston et al., 1983). Useful as it is with *C. elegans* development research, direct observation with microscopes is not suitable for the study of higher organisms such as mouse whose number of cells increases by several orders of magnitude and embryonic development takes place *in utero*. Direct observation is also not feasible for studying hematopoietic stem cell differentiation, which is a highly mobile tissue. Label retaining methods with vital dyes and radioactive tracers are utilized in the early years of

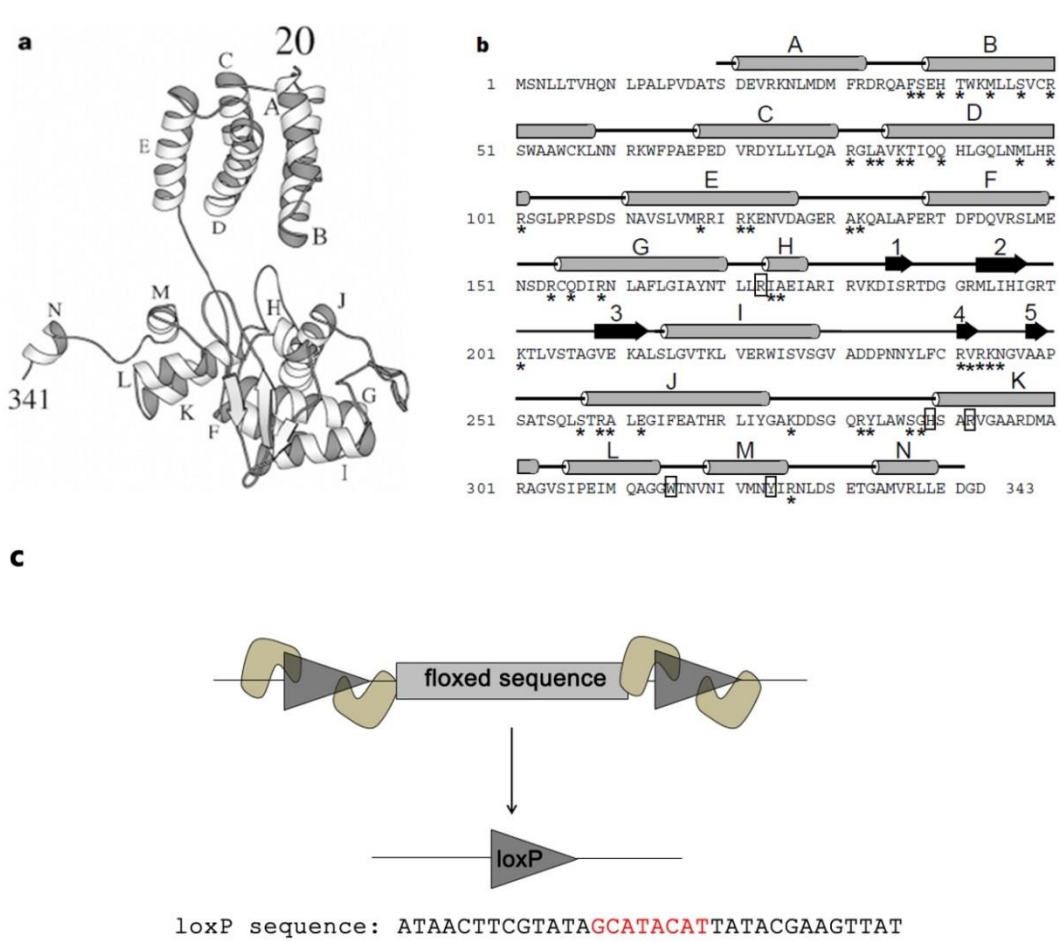


Fig 1.5 Cre recombinase structure and sequence. (a) Ribbon diagram depicting the 3D structure of Cre recombinase. Crystal structure is obtained for Cre 20-341. (b) Primary sequence of Cre recombinase and secondary structure alignment. Cylinders represent alpha helices while black arrows represent beta sheets. Active sites are boxed. Starred amino acid residues make contact with DNA. (c) Schematic representation of Cre/loxP recombination mechanism. Two Cre proteins occupy one loxP site. The floxed sequence is subsequently excised in this case where the two loxP sites are oriented in the same direction. When two loxP sites are oriented in the opposite direction, the floxed sequence is inverted. *trans* recombination (e.g., interchromosome exchange) may also occur when two loxP sites are located in two DNA molecules but the rate and efficiency is low. loxP sequence is also presented here. loxP sequence consists of two palindromic 13mers connected by a 8mer spacer (in red). Adapted from Guo et al., 1997.

developmental biology but are compromised by their drawbacks such as inaccuracy, disruption to normal development, and signal dilution, *etc*. DNA analog (e.g., BrdU, EdU) incorporation indicates DNA replication; it is not an accurate readout for cell differentiation but is still used for short term lineage tracing under certain circumstances. Transfection and viral transduction of genetic markers such as *GFP* and *lacZ* gene are superior to label-retaining methods because of marker inheritability across many cell generations. However, transfection and viral transduction lack precision and cellular resolution and thus are not versatile tools to label a specific cell type. Intra- and interspecies tissue transplantation is used to study organ origin during gastrulation but the lack of single cell resolution and the need of surgery and sometimes irradiation render this method unpopular.

Modern developmental biologists now use genetic methods to trace cell lineages. The two most commonly used genetic cell lineage tracing systems are the Cre/loxP system and the FLP/FRT system (Branda and Dymecki, 2004; Lewandoski, 2001). Both systems consist of two components: the Cre or FLP recombinase and the loxP or FRT recognition sequence. Both systems function in very similar mechanisms. The FLP/FRT system is inferior to and less commonly used than the Cre/loxP system in the mammalian models because of its lower recombination efficiency (Anastassiadis et al., 2009). Most mouse models use the Cre/loxP system and I will focus on the Cre/loxP system in the following.

The Cre recombinase is a 38 kilo-Dalton protein (343 amino acid residues) that belongs to the integrase family. It is first found in the P1 bacteriophage (Sternberg and Hamilton, 1981). Unlike other bacteriophage such as lamda phage, P1 bacteriophage does not integrate its DNA into the host genome. Rather, its DNA is circulized and maintained like a plasmid. The function of Cre recombinase is to promote P1 DNA cyclization after infection and resolve plasmid multimer to stabilize plasmid copy number

(Lobocka et al., 2004). Crystal structure has been solved for Cre recombinase (Fig 1.5 a, b) (Guo et al., 1997). It has two distinct N-terminal and C-terminal domains connected by a short peptide. Both domains contribute to DNA binding and contain active amino acid residues that coordinate nucleotides at the site of attack and concertedly nick DNA (Gibb et al., 2010; Van Duyne, 2001).

The binding sequence of Cre recombinase is the 34bp loxP site. loxP sequence is composed of two 13bp palindromic elements connected by an 8bp spacer. In a floxed sequence (sequence flanked by two loxP sites), each Cre protein occupies a 13mer half-site, thus forming a Cre_4loxP_2 complex. Because the 8bp spacer is asymmetrical and thus directional, the flanked sequence can be either excised or inverted depending on the relative orientation of the two loxP sites (Fig 1.5 c).

A cell lineage tracing reporter is usually a transgene with a fluorescent protein or *lacZ* gene following a floxed stop cassette. When used with a reporter, the Cre recombinase recognizes the loxP sites and excises the stop cassette, activating the reporter. Most but not all reporters are inserted into the *Rosa26* locus. *Rosa26* locus provides a constitutively active gene expression environment which allows the inserted transgene to be expressed in all types of tissues and at all life stages. *Rosa26* does not encode a functional protein and the insertion of exogenous DNA does not cause developmental or functional abnormality (Friedrich and Soriano, 1991; Zambrowicz et al., 1997). Once Cre recombinase excises the stop cassette from a reporter inserted into the *Rosa26* locus, the reporter is activated and expressed thereafter, regardless of the expression pattern of the Cre recombinase afterwards. In this way, cells will be permanently marked and their origins can be traced back.

Due to the success when this method was initially introduced, many modifications and improvements of the Cre/loxP system have been done over the years (Hayashi and McMahon, 2002; Lewandoski, 2001). First, to study the cell lineages from

a specific group of progenitor cells, various tissue-specific Cre-drivers are created, such as *Ngn3*^{Cre} (endocrine pancreas), *RIP*^{Cre} (β cell), *Lgr5*^{CreERT2} (intestinal stem cell), and *Albumin*^{Cre} (hepatocyte), *etc*. (Barker et al., 2007; Gu et al., 2003; Lewandoski, 2001; Postic et al., 1999). A list of common Cre drivers used in pancreas research can be found in the review by Dr. Magnuson and Dr. Osipovich (Magnuson and Osipovich, 2013).

Second, to control the timing of Cre-mediated recombination, a Cre-estrogen receptor (ER) fusion protein and its various derivatives (CreER, CreERT2) are designed and applied in mammalian models (Feil et al., 1997; Indra et al., 1999; Metzger et al., 1995). CreER is retained in the cytoplasm by heat shock protein (HSP) 70 and 90. Upon the administration of tamoxifen (TM) or its active metabolite 4-hydroxyltamoxifen (4OH-TM), 4OH-TM binds with the estrogen receptor and dissociates the CreER-HSP complex, enables CreER to translocate to the nucleus (Metzger et al., 1995). A fusion protein of Cre and progesterone receptor is also created but not widely used (Kellendonk et al., 1996). Similar to CreER, doxycycline controlled tetON/OFF systems also serve as temporal control switches in cell lineage analysis (Belteki et al., 2005; Lewandoski, 2001; Urlinger et al., 2000).

Third, researchers have developed a palette of reporters. These include lacZ reporters (Soriano, 1999) and reporters with fluorescent proteins of various colors (Madisen et al., 2010; Mao et al., 2001; Srinivas et al., 2001). Dual color reporters, which show different colors before and after recombination, not only track cell lineages but also allow the visualization of the transition state (Muzumdar et al., 2007; Novak et al., 2000). An MADM (mosaic analysis with double markers) reporter is designed to allow interchromosomal recombination and mosaic analysis (Tasic et al., 2012; Zong et al., 2005). The confetti mouse model advances reporters to a new level and is powerful in clonal analysis (Livet et al., 2007). Recently, other permissive gene loci, such as *Hprt*

and *Cd6*, are also reported to be good sites for reporter insertion (Ichise et al., 2014; Tasic et al., 2011). The combinatorial utilization of multiple reporters knocked into these loci may be useful in certain situations.

The application of Cre/loxP-based methods provides a powerful tool for developmental biologists to track cell lineages. All these toolkits enable researchers to analyze cell lineages in great spatial and temporal resolution. In addition to its power in cell lineage tracing, Cre/loxP-mediated recombination is also widely used in conditional gene activation, inactivation, and cell ablation, *etc*. (Ivanova et al., 2005; Lewandoski, 2001; Zhang and Lutz, 2002). Cre-mediated conditional gene manipulation allows versatile control of gene expression.

Despite its power as a cell lineage tracing tool, the Cre/loxP system should be used with caution. In Chapter II, I will discuss the limitations of Cre/loxP-mediated cell lineage tracing and gene manipulation. Specifically, I will report the observation of non-parallel recombination of multiple floxed alleles in the same cell and caution that non-parallel recombination should be kept in mind when interpreting the results from Cre/loxP-mediated DNA recombination experiments. Furthermore, in addition to the floxed alleles/reporters, I will review issues around Cre drivers and what we should consider in selecting or designing Cre drivers in the discussion section of Chapter II.

CHAPTER II

NON-PARALLEL RECOMBINATION LIMITS CRE/LOXP-BASED REPORTERS AS PRECISE INDICATORS OF CONDITIONAL GENETIC MANIPULATION

This chapter is adapted from a publication under the same title (Liu et al., 2013).

2.1 Abstract

Cre/loxP-mediated recombination allows for conditional gene activation or inactivation. When combined with an independent lineage-tracing reporter allele, this technique traces the lineage of presumptive genetically modified Cre-expressing cells. Several studies have suggested that floxed alleles have differential sensitivities to Cre-mediated recombination, which raises concerns regarding utilization of Cre reporters to monitor recombination of other floxed loci of interest. Here, we directly investigate the recombination correlation, at cellular resolution, between several floxed alleles induced by Cre-expressing mouse lines. The recombination correlation between different reporter alleles varied greatly in otherwise genetically identical cell types. The chromosomal location of floxed alleles, distance between loxP sites, sequences flanking the loxP sites, and the level of Cre activity per cell all likely contribute to observed variations in recombination correlation. These findings directly demonstrate that, due to non-parallel recombination events, commonly available Cre reporter mice cannot be reliably utilized, in all cases, to trace cells that have DNA recombination in independent-target floxed alleles, and that careful validation of recombination correlations are required for proper interpretation of studies designed to trace the lineage of genetically modified populations, especially in mosaic situations.

2.2 Introduction

The introduction of Cre/loxP-mediated DNA recombination technique has facilitated the investigation of cell origins and the manipulation of gene expression. With the advent of a vast array of tissue and cell-specific Cre drivers, fluorescent reporters, and high resolution microscopy, this technique has been wildly used in developmental biology, immunology, cancer research, and countless other fields. Compared to other cell lineage tracing methods, such as dye/radioactive labelling, BrdU incorporation *etc*., genetically-based Cre/loxP method permits a more accurate means of cell lineage analysis (Kretzschmar and Watt, 2012). In addition to its power in cell lineage tracing, Cre/loxP-mediated recombination is also widely used in conditional gene activation, inactivation, and cell ablation, *etc*. (Ivanova et al., 2005; Lewandoski, 2001; Zhang and Lutz, 2002).

Nonetheless, the efficiency of Cre does not guarantee 100% recombination. Incomplete recombination results in the missing of certain cells in lineage tracing and mosaic pattern in the case of ectopic gene expression or gene inactivation. Incomplete recombination does not exhibit prohibiting issues as long as proper quantification and interpretation are applied, but it is problematic when there are two floxed alleles in one cell and their recombination doesn't occur simultaneously. For instance, Cre/loxP-based lineage tracing reporters are often utilized to determine the consequence of genetic manipulation at another floxed locus within individual cells or populations of cells. However, this assumes that the activation of a reporter allele indicates the recombination of the other floxed locus (Dzierzak and Speck, 2008; Gu et al., 2003; Herrera et al., 1998; Kawaguchi et al., 2002; Kretzschmar and Watt, 2012; Lao et al., 2012; Spence et al., 2009). Nonetheless, because the chromosomal location of loxP sites (Vooijs et al., 2001), distances between loxP sites (Collins et al., 2000; Koike et al., 2002; Zong et al., 2005),

and cell-type-specific epigenetic context of floxed loci (Hameyer et al., 2007; Long and Rossi, 2009) all affect recombination efficacy, the occurrence of recombination in one allele may not predict recombination in the other within the same cell. Here, we directly demonstrate such non-parallel recombination with commonly utilized Cre reporter alleles and several Cre/CreERT mouse lines (Table 1). We report that simultaneous recombination is not always guaranteed, which complicates results and makes data interpretation difficult. We conclude that non-parallel recombination exists and should be taken into consideration when examining and interpreting experimental results involving Cre/loxP.

2.3 Results

In order to examine the factors that affect Cre-mediated recombination, we utilized several pancreatic gene-based Cre driver and several commonly used reporter mouse lines to investigate reporter activation. The Cre lines include *Pdx1*^{Cre}, *Pdx1*^{CreERT}, *Ngn3*$^{B\text{-}Cre}$, and *Sox9*^{CreERT2}, which induce recombination in pancreatic or endocrine progenitors. The reporters include *R26R*^{Ai9}, *R26R*^{eYFP}, *R26R*^{eGFP}, *and R26R*^{mTmG}, and a transgenic *Z/EG* reporter. Besides the transgenic Z/EG reporter, the other Cre reporters are all knocked into the Rosa26 locus but they differ in their floxed sequences and ancillary CAG promoter usage. A *Pdx1*^{FLOX} allele is also used in line with the reporters.

Floxed alleles in the same cell are not always recombined simultaneously

First, we derived *Ngn3*$^{B\text{-}Cre}$*; R26R*$^{Ai9/eYFP}$ neonatal mice, wherein the reporter alleles at the *Rosa26* locus are *R26R*^{Ai9} and *R26R*^{eYFP}, and examined the extent of parallel reporter activation in endocrine and exocrine pancreatic lineages. Progenitor cells that express high levels of Ngn3 become endocrine islet cells, whereas cells that

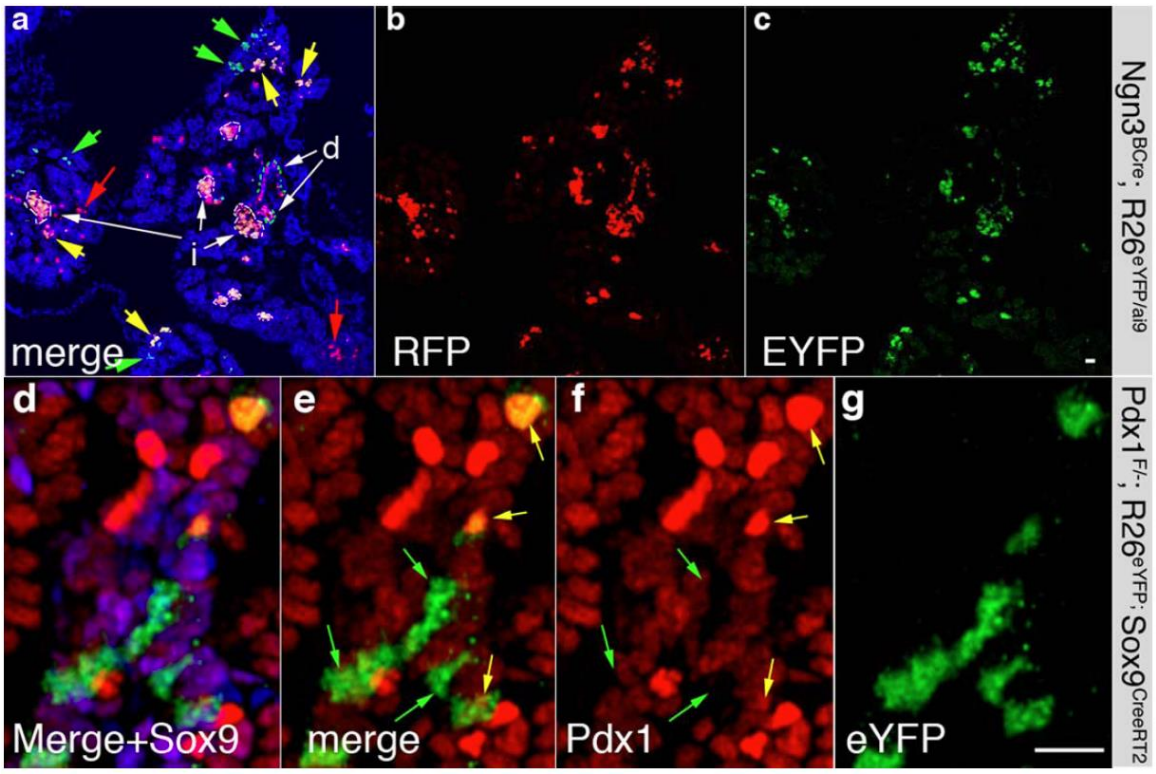


Fig 2.1 Recombination in multiple floxed alleles are independent events. (a–c) Reporter activation in neonatal *Ngn3*$^{B\text{-}Cre}$*; R26R*$^{eYFP/Ai9}$ pancreas. RFP (tdTomato), eYFP, and a merged channel are presented. Islet cells, broken-lined circle (panel a). Green arrows indicate eYFP only cells; red arrows indicate RFP only cells; and yellow arrows indicate cells expressing both fluorescent proteins. (d–g) E15.5 pancreatic sections from *Pdx1*$^{FLOX/-}$*; Sox9*^{CreERT2}*; R26R*^{eYFP} embryos (0.3 mg TM injected at E12.5) with immunodetection for Pdx1, Sox9, and eYFP. Two merged images [d (Pdx1, Sox9, and eYFP) and e (Pdx1 and eYFP)] and two individual channels [f (Pdx1) and g (eYFP)] are shown. Yellow arrows, Pdx1^{+}eYFP^{+} cells. Green arrows, Pdx1^{-}eYFP^{+} cells. Bars=20 μm.

express low levels of Ngn3 become exocrine cells (Schonhoff et al., 2004; Wang et al., 2010). These properties allowed us to assess the influence of differential *Ngn3*$^{B\text{-}Cre}$-driven promoter activity on the recombination correlation between *R26R*^{Ai9} and *R26R*^{eYFP} (express RFP (tdTomato) or eYFP, respectively). Most, if not all, endocrine islet cells (recognizable as tightly packed cell clusters) in *Ngn3*$^{B\text{-}Cre}$*; R26R*$^{Ai9/eYFP}$ neonatal pancreas produced both reporters. In contrast, many acinar and duct cells only produced a single reporter, indicating non-parallel recombination (Fig 2.1 a–c). These above findings suggest that "high Ngn3-expressing cells" produced high Cre levels sufficient to recombine both alleles within individual cells, whereas low-Ngn3- expressing cells recombine one allele but not the other.

To assess the incidence of non-parallel recombination under mosaic experimental conditions, we used a *Sox9*^{CreERT2} transgene to drive CreERT2 (a tamoxifen (TM)- inducible Cre), to recombine a Cre reporter (*R26R*^{eYFP}) and a floxed *Pdx1* allele (*Pdx1*^{FLOX}) whose recombination results in a null mutation. Most, if not all, pancreatic progenitor cells express Sox9 and a low level of Pdx1 (Pdx1^{Lo}). When pancreatic progenitor cells differentiate into β cells, *Pdx1* expression is upregulated (Pdx1^{Hi}), while *Sox9* becomes inactivated (Fujitani et al., 2006; Kopp et al., 2011). Therefore, any Sox9^{+} pancreatic progenitor cell that has inactivated *Pdx1* will be incapable of becoming a Pdx1^{Hi}Sox9^{-} cell. We administered 0.3 mg/mouse TM to plugged females at E12.5 to activate CreERT2 in *Pdx1*$^{FLOX/+}$*; Sox9*^{CreERT2}*; R26R*^{eYFP} mouse pancreas in a mosaic fashion, and scored eYFP^{+} cells for *Sox9* and *Pdx1* expression statuses. Three days after TM administration, about half of the eYFP^{+} cells retained Pdx1 production, with a portion of these cells displaying a high Pdx1 signal (Fig 2.1 d–g), demonstrating that the *Pdx1*^{FLOX} allele is not inactivated even though recombination in the *R26R*^{eYFP} allele had occurred in some cells. Together, the above

Table 1. Reporter alleles, Cre drivers, and conditional alleles used for non-parallel recombination study

Reporter name	Official strain name	Floxed sequences/transgene	References
$R26R^{Ai9}$	(ROSA)26Sor$^{tm9(CAG-tdTomato)Hze/J}$	3XpolyA	Madisen et al., 2010
$R26R^{eYFP}$	Gt(ROSA)26Sor$^{tm1(EYFP)Cos/J}$	Pgk-Neo-3XpolyA	Srinivas et al., 2001
$R26R^{EGFP}$	Gt(Rosa)26Sor$^{<tm1(rtTA,EGFP)Nagy>/J}$	Pgk-Neo-3XpolyA	Belteki et al., 2005
$R26R^{mTmG}$	Gt(ROSA)26Sor$^{tm4(ACTB-tdTomato,-EGFP)Luo/J}$	Membrane-tagged td-Tomato-PolyA	Muzumdar et al., 2007
Z/EG	Tg(CAG-Bgeo/GFP)21Lbe/J	LacZ-polyA	Novak et al., 2000
$Ngn3^{B-Cre}$	Tg(Neurog3Cre)C1Able/J	Cre	Schonhoff et al., 2004
$Pdx1^{Cre}$	Tg(Pdx1-Cre)89.1	Cre	Gu et al., 2002
$Pdx1^{CreERT}$	Tg(Pdx1-cre/Esr1*)35.10Dam	CreERT	Gu et al., 2002
$Pdx1^{FLOX}$	Pdx1-tm4Cvw	Conditional LOF allele	Gannon et al., 2008
$Sox9^{CreERT2}$	Tg(Sox9-cre/ERT2)1Msan	CreERT2	Kopp et al., 2011

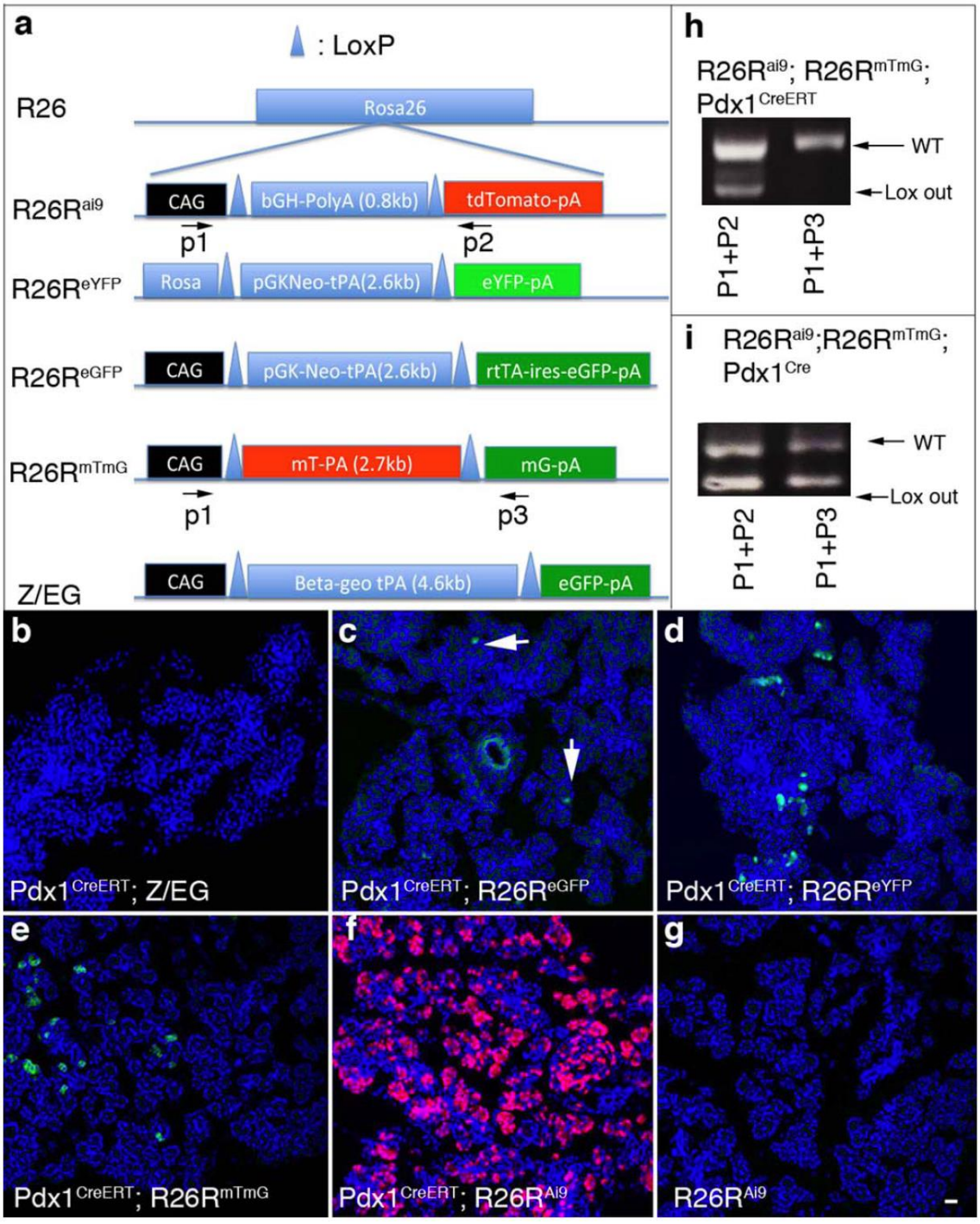

a
: LoxP
R26
Rosa26
R26R^ai9
CAG
bGH-PolyA (0.8kb)
tdTomato-pA
p1
p2
R26R^eYFP
Rosa
pGKNeo-tPA(2.6kb)
eYFP-pA
R26R^eGFP
CAG
pGK-Neo-tPA(2.6kb)
rtTA-ires-eGFP-pA
R26R^mTmG
CAG
mT-PA (2.7kb)
mG-pA
p1
p3
Z/EG
CAG
Beta-geo tPA (4.6kb)
eGFP-pA
h
R26R^ai9; R26R^mTmG;
Pdx1^CreERT
WT
Lox out
P1+P2
P1+P3
i
R26R^ai9;R26R^mTmG;
Pdx1^Cre
WT
Lox out
P1+P2
P1+P3
b
Pdx1^CreERT; Z/EG
c
Pdx1^CreERT; R26R^eGFP
d
Pdx1^CreERT; R26R^eYFP
e
Pdx1^CreERT; R26R^mTmG
f
Pdx1^CreERT; R26R^Ai9
g
R26R^Ai9

Fig 2.2 Cre-reporter alleles have different sensitivity to Cre-induced recombination. (a) Diagrammatic representations of the reporter alleles utilized in this study. (b–g) Representative neonatal pancreatic sections of *Pdx1*^{CreERT} transgene in combination with reporter lines: *Z/EG, R26R*^{eYFP}*, R26R*^{eGFP}*, R26R*^{mTmG}*, and R26R*^{Ai9} without TM administration. Natural fluorescence of the expressed FPs is shown. White arrows in panel c point to two eGFP^{+} cells. An image of *R26R*^{Ai9} pancreatic section (without *Pdx1*^{CreERT}) is also included, to show that its reporter activation depends on the presence of CreERT (g). DAPI marks all nuclei, including non-pancreatic mesenchymal cells that do not express Cre. (h and i) PCR-based detection of DNA products after expected recombination in *Pdx1*^{CreERT}*; R26R*^{Ai9}*; R26R*^{mTmG} (h) and *Pdx1*^{Cre}*; R26R*^{Ai9}*; R26R*^{mTmG} (i) pancreas, respectively. The locations of utilized oligos, P1, P2, and P3 are noted in panel (a). A wild control band (WT, a fragment in the Myt1 locus) was utilized as PCR control (450bp). The Lox out bands were only detectable after Cre-mediated recombination (250bp). Bar=20 μm.

findings demonstrate that different levels of Cre influence the efficiency with which one can recombine two independent floxed alleles in an individual cell.

Cre reporters have different sensitivity to Cre recombinase activity

Several available reporters including *R26R*^{eGFP}, *R26R*^{Ai9}, *R26R*^{mTmG}, and *R26R*^{eYFP} are derived by *Rosa26*-based targeting, and contain different stop signals and reporter genes (Table 1). Conversely, *Z/EG* reporter is an insertion-based transgene (Lobe et al., 1999; Novak et al., 2000). Recombination events in *R26*^{Ai9}, *R26R*^{mTmG}, *R26R*^{eYFP}, and *Z/EG* lines activate a downstream fluorescence reporter only, whereas recombination in *R26R*^{eGFP} results in an iRES-based bi-cistronic mRNA that produces both rtTA and eGFP (Fig 2.2 a). Thus, *R26R*^{eGFP} produces lower levels of eGFP compared with other reporters after recombination. Yet, the eGFP expression pattern in *R26R*^{eGFP} faithfully identifies cells that have undergone recombination (Belteki et al., 2005). To evaluate, within a linear range, the level of Cre required to activate each reporter gene, we took advantage of a *Pdx1*^{CreERT} line that maintains a low level of Cre activity in pancreatic progenitor cells (Gu et al., 2002) in the absence of TM (see below). No TM-independent recombination was observed in *Pdx1*^{CreERT}*; Z/EG* pancreatic cells (Fig 2.2 b). Similarly, less than 0.1% of all pancreatic cells of *Pdx1*^{CreERT}*; R26R*^{eGFP} underwent recombination (n=6; Fig 2.2 c). Both *Pdx1*^{CreERT}*; R26R*^{eYFP} and *Pdx1*^{CreERT}*; R26R*^{mTmG} mice displayed between 0.4 and 2.7% pancreatic cells with recombination (n=6–8; Fig 2.2 d-e). Surprisingly, over one-third of all pancreatic cells in *Pdx1*^{CreERT}*; R26R*^{Ai9} mice recombined to express RFP (Fig 2.2 f; n=5). None of the reporter mice express detectable FPs in the absence of the Cre-expressing transgene (Fig 2.2 g and data not shown). To confirm that the lack of reporter gene expression was not a result of gene silencing after recombination, we examined DNA recombination in *Pdx1*^{CreERT}*; R26R*$^{Ai9/mTmG}$ (with two reporter alleles at the *Rosa26* locus) pancreas by PCR analysis.

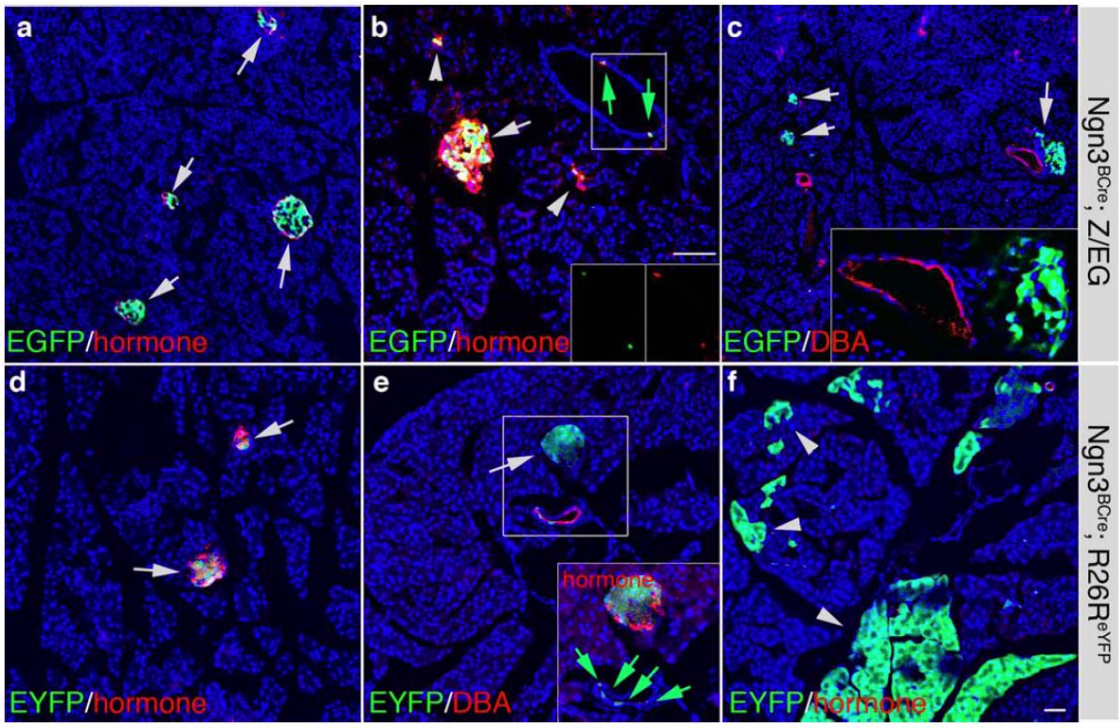


Fig 2.3 Cre-reporter allele sensitivity can lead to different lineage tracing outcomes. *Ngn3*$^{B\text{-}Cre}$-mediated recombination of the *Z/EG* (a–c) and *R26R*^{eYFP} (d–f) reporter alleles. Sections from 2-month-old pancreata were utilized [counter-stained with DAPI (blue)]. Islets are marked with arrows. In sections of *Ngn3*$^{B\text{-}Cre}$*; Z/EG* pancreata, GFP^{+} cells did not react with DBA lectin, which specifically marks pancreatic ducts (c). Note that eGFP^{+} cells observed in duct (green arrows) expressed endocrine hormones (insets in b). Inset in c, a higher magnification image to show lack of green cells in a duct section. In *Ngn3*$^{B\text{-}Cre}$*; R26R*^{eYFP} mice, significant numbers of duct (e, green arrows) and acinar cells (f, white arrowhead) expressed eYFP reporter. In e, the inset shows hormone staining within the boxed area. Also note that the acinar labeling in pancreatic tissues is not randomly distributed, so that some microscopic fields do not have eYFP^{+} exocrine cells (d and e). Bar=50 μm.

Recombinant DNA product was detected from the *R26R*^{Ai9} allele but not from *R26R*^{mTmG} (Fig 2.2 h). As a positive control for PCR detection, recombinant products were detected at *R26R*^{Ai9} and *R26R*^{mTmG} loci in *Pdx1*^{Cre}*; R26R*$^{Ai9/\ mTmG}$ pancreas (Fig 2.2 i). Taken together, these data demonstrate differential recombination efficiencies between select reporter alleles in a model for low-level Cre activity.

The above findings strongly suggest that experimental lineage tracing results may vary in a reporter line–dependent fashion. We tested this possibility by following the lineage of Ngn3-expressing cells using different reporter alleles, to reconcile observations suggesting in one case that Ngn3^{+} cells only give rise to endocrine islet cells (Gu et al., 2002), and in another case that Ngn3^{+} cells also give rise to exocrine pancreatic cells (Schonhoff et al., 2004; Wang et al., 2010). We derived *Ngn3*$^{B\text{-}Cre}$*; Z/EG* and *Ngn3*$^{B\text{-}Cre}$*; R26R*^{eYFP} adult pancreata to examine eGFP or eYFP expression. Most of the identifiable eGFP^{+} cells in *Ngn3*$^{B\text{-}Cre}$*; Z/EG* pancreata were localized in cell clusters producing endocrine hormones (Fig 2.3 a–c). Several single eGFP-producing cells were found and within the duct of *Ngn3*$^{B\text{-}Cre}$*; Z/EG* pancreas; yet, these lineage-traced cells also expressed endocrine hormones (inset in Fig 2.3 b). In contrast, a large number of eYFP^{+} cells were found in pancreatic duct and acinar tissue of *Ngn3*$^{B\text{-}Cre}$*; R26R*^{eYFP} mice (Fig 2.3 d–f). These findings suggest that reporter selection in Cre-reporter-based lineage tracing influence experimental outcomes.

High level of Cre activity cannot fully normalize non-parallel recombination

Finally, we investigated whether high levels of Cre activity can normalize the observed differential sensitivities of *R26R*^{eGFP} and *R26R*^{eYFP} reporters. Both *Pdx1*^{Cre} and *Pdx1*^{CreERT} under high-TM dose conditions were used to produce high Cre activities. One milligram of TM was administered to E14.5 pregnant female mice to activate CreERT in *Pdx1*^{CreERT}*; R26R*^{eYFP} or *Pdx1*^{CreERT}*; R26R*^{eGFP} embryos, respectively. Reporter

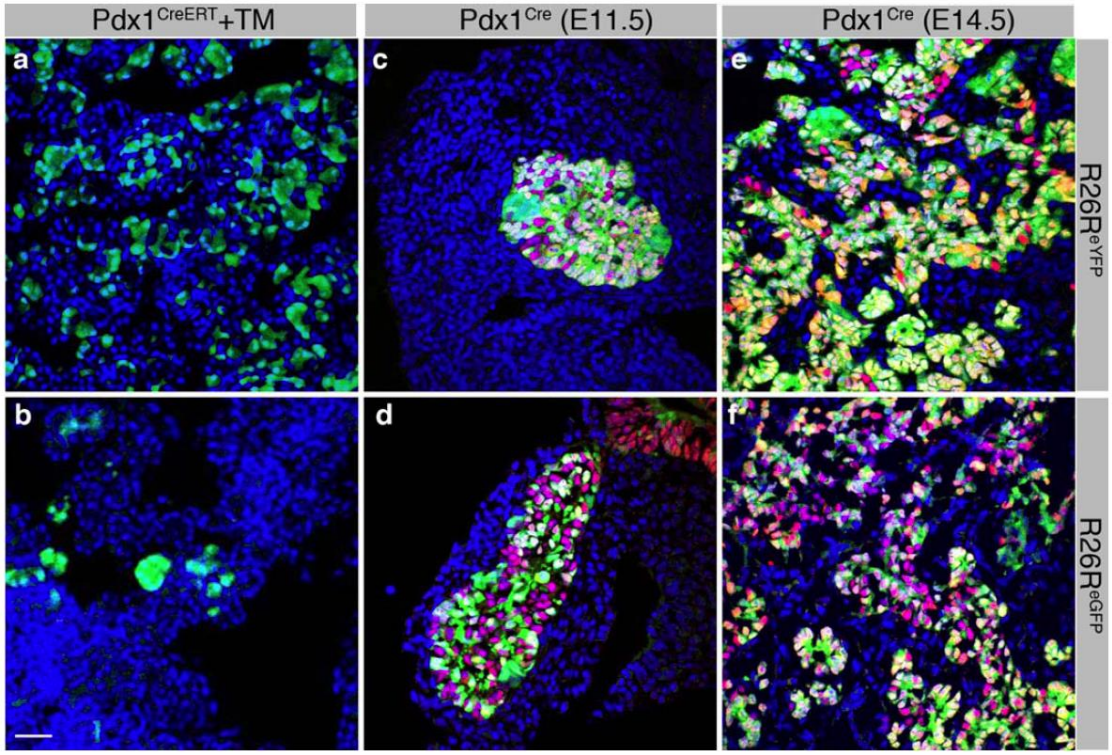


Fig 2.4 Increased Cre activity cannot normalize the variable sensitivity of different floxed alleles. Neonatal pancreatic sections of different reporter mice when activated by increased levels of Cre with *Pdx1*^{CreERT} plus TM administration or *Pdx1*^{Cre}. In the presence of TM, *Pdx1*^{CreERT} induced over a third of pancreatic cells to activate reporter expression in *R26R*^{eYFP} (a). Less than 8% pancreatic cells activated eGFP in *R26R*^{eGFP} mice with TM+*Pdx1*^{CreERT} (b). Similarly, *R26R*^{eYFP} also showed higher recombination than *R26R*^{eGFP} when a *Pdx1*^{Cre} line was analyzed at E11.5 and E14.5 (c–f). Note the red only cells (Pdx1 staining to visualize the pancreatic cells) in panels c–f, which indicate the cells that failed to undergo recombination. DAPI, blue. Bar=50 μm.

expression was characterized in newly born neonates (n=3). Over one-third of pancreatic cells in *Pdx1*^{CreERT}*; R26R*^{eYFP} mice expressed eYFP (Fig 2.4 a), whereas less than 8% of pancreatic cells in *Pdx1*^{CreERT}*; R26R*^{eGFP} mice expressed eGFP (Fig 2.4 b). Similarly, *Pdx1*^{Cre} induced eYFP expression in more than 85% of *Pdx1*^{Cre}*; R26R*^{eYFP} pancreatic cells at E11.5 (Fig 2.4 c), but only 35% of pancreatic cells activated eGFP expression in *Pdx1*^{Cre}*; R26R*^{eGFP} pancreata at the same stage (Fig 2.4 d). By E14.5, over 93% of pancreatic cells in *Pdx1*^{Cre}*; R26R*^{eYFP} mice activated eYFP expression (Fig 2.4 e), whereas only 72% of pancreatic cells activated eGFP expression in *Pdx1*^{Cre}*; R26R*^{eGFP} mice (Fig 2.4 f; n=3). These data suggest that increased Cre activity is not sufficient to normalize the differential sensitivity detected for these two floxed reporter alleles.

2.4 Discussion

Our studies demonstrate that *R26R*^{Ai9} is the most sensitive reporter of Cre-mediated recombination, whereas *Z/EG* is the least sensitive reporter. Our studies also directly demonstrate that Cre-mediated recombination in one floxed allele does not necessarily report recombination at another allele within the same cell. While this phenomenon has been implicated in other experimental settings, it has not been demonstrated directly (Hameyer et al., 2007; Long and Rossi, 2009; Schmidt-Supprian and Rajewsky, 2007; Vooijs et al., 2001). Importantly, observed non-parallel recombination is particularly severe in cells with low levels of Cre activity, which can occur in many experimental settings, such as weak promoter-driven Cre or TM-inducible CreER activation for mosaic analysis. Our data further implies that reporter sensitivity inversely correlates with the distance between loxP sites in the R26-reporter transgenes (Table 1). This implication also seemed to be applicable to the *Z/EG* line, which has the greatest inter-loxP distance and lowest sensitivity to recombination, albeit with the

additional variable that the *Z/EG* cassette is located on a different chromosome from the R26-based reporters. Furthermore, one unexpected finding from our analysis is that *R26R*^{eYFP} and *R26R*^{eGFP} have vastly different recombination sensitivities despite identical floxed stop signals. This observation suggests that sequences outside the loxP sites may influence Cre-based recombination efficiencies. At least two mechanisms could contribute to this difference. It is likely that the different sequences downstream the loxP sites of *R26R*^{eYFP} and *R26R*^{eGFP} result in different methylation status, which has been suggested to affect recombination efficiency (Long and Rossi, 2009). Alternatively, the different sequences surrounding loxP sites could form different nucleasomal structures that affect the accessibility of loxP sites to Cre enzyme. Our current data do not allow us to differentiate these two possibilities. Future efforts to unravel these possibilities will likely facilitate engineering floxed alleles with specific recombination efficiencies.

In the above, we primarily focused on the Cre reporters or floxed conditional alleles. Nonparallel recombination within a cell results in discrepancies in cell lineage tracing results and compromises the use of fluorescent reporters as an accurate surrogate marker for gene activation/inactivation. We tapped on the recombinase activity level from different Cre drivers and found that high Cre activity level do not fully normalize the sensitivity variation from different Cre reporters. In the following discussion, I will discuss more about the issues in the design of a new Cre driver or the selection of an existing Cre driver in research.

First of all, Cre protein, especially when excessively expressed as is often the case with strong promoter-driven Cre transgenes, exerts toxicity to mammalian cells (Schmidt-Supprian and Rajewsky, 2007). The mammalian genome contains pseudo-loxP sites where Cre recombinase might bind and disrupt the genome (Thyagarajan et al., 2000). Although the binding affinity and recombination efficiency is remarkably lower on the pseudo-loxP site than on the wild type loxP site, Cre-mediated recombination

occurs or leads to DNA damage that is not faithfully repaired. For instance, Cre expression under the mouse *Protamine 1* (*Prm1*) promoter causes male sterility due to unfaithful Cre activity (Schmidt et al., 2000). It is, therefore, good practice to include proper controls, including wild-type, Cre-driver-only, and reporter-only groups, in experimental designs.

Second, the expression pattern of Cre driver one chooses should be carefully examined when tissue-specific gene manipulation is desired. Cre driver should faithfully recapitulate the expression pattern of the gene of interest. Take RIP^{Cre} for example, RIP^{Cre} is widely used in β-cell and diabetes research as a β-cell-"specific" Cre driver. However, RIP^{Cre} mice show glucose intolerance and impaired glucose-stimulated insulin secretion in both males and females, although severity of the phenotype depends on mouse genetic background (Lee et al., 2006). In fact, the *RIP* promoter is active not only in pancreatic β cells but also in the brain, especially the hypothalamus (Lee et al., 2006; Wicksteed et al., 2010). Undesirable gene editing in the brain might confound β-cell-related phenotype and cause ambiguity in interpreting results, providing that the hypothalamus is pivotal in regulating food intake and energy metabolism (Koch and Horvath, 2014; Mighiu et al., 2013; Sousa-Ferreira et al., 2014). Therefore, extensive characterization should be performed the first time a new Cre driver mouse strain is generated. In general, knock-in Cre alleles are preferred than transgenes. Transgenes are randomly inserted into the genome and they may not fully recapitulate the expression pattern of the gene of interest. In addition, random insertion may lead to undesirable activation/inactivation of endogenous genes. The Jackson Laboratory has comprehensively characterized commonly used Cre drivers and the information is publically accessible (Heffner et al., 2012). Cre drivers commonly used in pancreas research are also reviewed (Magnuson and Osipovich, 2013).

Third, germ line recombination might occur. It has been reported in several cases that recombination of floxed alleles occurs in all mouse embryos even when not all of them inherit the Cre driver (Hafner et al., 2004; Ramirez et al., 2004; Schmidt-Supprian and Rajewsky, 2007). This is probably because Cre recombinase remains active in the oocytes until fertilization. To avoid germ line recombination, it is better to maintain Cre driver and floxed allele in different parents.

Fourth, recombination efficiency should be examined. Cre recombinase may not be expressed at a substantial amount, especially when it is expressed from a tissue/cell type-specific promoter (Long and Rossi, 2009; Postic and Magnuson, 2000). Suboptimal Cre expression causes incomplete recombination, leading to mosaic gene activation/inactivation and loss of track of progenitors *etc*., which confounds the quantification and interpretation of results. The efficiency of a new Cre driver could be tested *in vitro* and benchmarked against well characterized Cre drivers. In some cases, an indicator protein (EGFP) is expressed along with a Cre driver in a bicistronic fashion or as fusion proteins, allowing researchers to visualize the Cre^{+} domain and estimate the recombination efficiency (Arnes et al., 2012; Choi et al., 2012; Hudson et al., 2011; Magnuson and Osipovich, 2013; Passegue et al., 2004; Woodhead et al., 2006).

Implementation and development of vast site-specific DNA recombination toolkits, including Cre/loxP and FLP/FRT, *etc*. permits us to interrogate the origins of cell lineages in an unprecedented level of versatility and resolution, as well as manipulate gene activation or inactivation in a spatiotemporally controlled manner at our will. Despite its invaluable power as an indispensable research tool, the interpretation of Cre/loxP-mediated DNA recombination and cell lineage tracing should always be treated with care.

2.5 Materials and methods

Mouse Strains and Care

Mouse lines *R26R*^{eYFP}, *R26R*^{Ai9}, *Z/EG*, and *R26R*^{eGFP} (see Table 1 for details) were from Jackson Laboratory (Bar Harbor, ME). *Ngn3*$^{B\text{-}Cre}$ was a gift from A. Leiter (Schonhoff et al., 2004). *Sox9*^{CreERT2} was a gift from M. Sander (Kopp et al., 2011). *Pdx1*^{CreERT}, *Pdx1*^{Cre}, and *Pdx1*^{FLOX} were previously reported (Gannon et al., 2008; Gu et al., 2002; Wang et al., 2007). Tissue collection and section preparation followed published methods (Wang et al., 2008).

Conventional PCR (31 cycles) was utilized to detect the genomic sequences derived from recombination of *R26R*^{Ai9} and *R26R*^{mTmG} alleles. Control oligos to amplify genomic DNA: 5'-CCATGCATATGCCTGGTGCTTGT-3' and 5'-GGGTTAGGATTAAGAGTTTTAGT-3'. Oligos for detecting the recombination product P1: 5'-GGTTCGGCTTCTGGCGTGTGAC-3', P2: 5'-GCACCTTGAAGCGCATGAACTC-3' and P3: 5'-ACGCTGAACTTGTGGCCGTTTAC-3' (see Fig 2.2 a).

All mice were housed and cared in the Vanderbilt Division of Animal Care and in compliance to IACUC regulations.

Immunostaining and Confocal Imaging

Antibodies utilized were as follows: chicken anti-GFP (Aves Labs, Inc., Tigard, OR), 1:500; rabbit anti-Sox9 (Millipore, Billerica, MA); guinea pig anti-insulin, guinea pig anti-glucagon, rabbit anti-SS, goat anti-PP, FITC-conjugated donkey anti-rabbit, FITC-conjugated donkey anti-guinea pig, Cy3- conjugated donkey anti-rabbit, Cy3-conjugated donkey anti-guinea pig, Cy3-conjugated donkey anti-goat, Cy5- conjugated donkey anti-rabbit, Cy5-conjugated donkey anti-goat, and FITC-conjugated streptavadin were all from Jackson Immunoresearch. Biotin-DBA was from Sigma Aldrich. Antibody detection

followed standard protocols (Wang et al., 2007). All antibodies were utilized at a 1:1000 dilution ratio, or as noted.

Both confocal imaging and epifluorescence microscopy were used. For semi-quantification, a quarter (embryonic pancreata) to one-tenth (postnatal pancreata) of the pancreas, in 10–20μm sections, was analyzed. Images (usually low magnification to image a large area, but high enough to discern single cells properly) were captured at >20 microscopic views/slide. Cell numbers were quantified with ImageJ.

CHAPTER III

ACTIVATION OF MYT1 IN A SUBSET OF NGN3$^+$ PROGENITORS FACILITATES BETA-CELL FATE CHOICE

3.1 Abstract

In the developing pancreas, the basic helix-loop-helix (bHLH) transcription factor Neurogenin3 (Ngn3) specifies endocrine cells, including insulin-expressing β cells, glucagon-expressing α cells, δ, PP, and ε cells. How Ngn3$^+$ progenitors differentiate into various endocrine cell types is not clearly understood. Interestingly, we found that Ngn3$^+$ progenitors are heterogeneous in terms of their co-expression pattern with other transcription factors, such as Myt1. This heterogeneous co-expression pattern divides Ngn3$^+$ progenitors into subsets (i.e., Ngn3$^+$Myt1$^+$ and Ngn3$^+$Myt1$^-$ subsets). By using a novel bipartite Cre lineage tracing method, we were able to perform lineage tracing of the Ngn3$^+$Myt1$^+$ progenitor subpopulation and found that Ngn3$^+$Myt1$^+$ progenitors favored β cells over α cells. Preliminary epigenetic studies showed that DNA methylation of a regulatory element, Myt1 region 2, decreased from E10.5 to E15.5, which may facilitate enhanced Myt1 expression and β cell differentiation. Perturbing methylation by chemical inhibition and *Dnmt3b* (DNA methyltransferase 3b) overexpression both altered β-to-α ratio, although detailed mechanism awaits further investigation. The bipartite Cre cell lineage tracing technique will enable us to delineate the differentiation program of pancreatic endocrine progenitors and holds great potential in mapping stem cell fate in other systems.

3.2 Introduction

The pancreas is a pivotal organ of the human body. It is composed of both exocrine and endocrine cells. The exocrine cells, including duct cell and acinar cells, are responsible for secreting and delivering digestive enzymes to the duodenum. The endocrine pancreas is scattered among the acinar cells as islets and comprises α, β, δ, PP, and ε cells. Endocrine pancreas development is regulated by a cascade of transcription factors. During embryogenesis, a domain within the foregut epithelium expressing Pdx1 is reserved for pancreas (Jonsson et al., 1994; Offield et al., 1996). *Ngn3* is then activated in the $Pdx1^+$ primordia, which marks the endocrine progenitors (Gradwohl et al., 2000; Gu et al., 2002). $Ngn3^+$ pro-endocrine progenitors then differentiate into all types of endocrine cells, as a result of the interplay of various transcription factors, including *Pax4*, *Arx*, *Myt1*, *Nkx2.2*, *Insm1*, and *Nkx6.1*, *etc.* (Collombat et al., 2006; Jensen, 2004; Pan and Wright, 2011).

Although $Ngn3^+$ progenitors as a population produce all types of endocrine cells, it is unclear how individual $Ngn3^+$ progenitor makes its decision to become one endocrine cell type. Johansson et al. proposed a competence window model that the specification of endocrine cells occurs in sequential time windows but not in random (Johansson et al., 2007). α cells are produced first, starting from as early as E9.5. Next, during E12.5–E16.5, there is a major wave of β-cell generation, with many α cells also being differentiated during this time period. δ and PP cells do not appear until late gestational stages. The phenomenon of successive (but partially overlapping) differentiation of endocrine cells suggests that $Ngn3^+$ endocrine progenitors pass through consecutive temporal windows during which their differentiation competence changes. Johansson et al. also found that the competence of $Ngn3^+$ progenitors is intrinsic to the epithelium but not influenced by signaling from the surrounding

mesenchyme. When they recombined embryonic pancreatic epithelium and mesenchyme in a stage-matched or heterochronic manner, they discovered that the distribution of endocrine cells is similar between groups if the pancreatic epithelia are of the same stage, regardless of the stage of the mesenchyme used (Johansson et al., 2007). It appears endocrine differentiation in the pancreatic epithelium is autonomously regulated. I therefore focused on the many transcription factors pertaining to the endocrine progenitors but not external signaling and investigated the roles of transcription factors in determining endocrine cell fate.

A cascade of transcription factors regulates the endocrine differentiation process (Jensen, 2004; Wilson et al., 2003) and many of them are *Ngn3* downstream targets. These include *Arx* (Collombat et al., 2003), *Pax4* (Smith et al., 2003), *NeuroD* (Huang et al., 2000), *Nkx2.2* (Watada et al., 2003), and *Insm1* (Mellitzer et al., 2006; Osipovich et al., 2014) *etc*. In the early pancreatic endocrine progenitors, *Arx* and *Pax4* are co-expressed (Collombat et al., 2005). Then, *Pax4* is gradually restricted to β and δ lineages and is switched off after birth; *Arx* expression, on the other hand, is restricted to α cells and its expression persists in adult α cells. Experimental evidences suggest that *Arx* and *Pax4* specify α cell and β/δ cell lineages, respectively. Deletion of *Pax4* results in the loss of β and δ endocrine lineages (Sosa-Pineda, 2004). In contrast, loss of *Arx* leads to increased β and δ cell number at the expense of α cells (Collombat et al., 2003). Expectedly, *Pax4* ectopic expression in α cells converts them into β cells *in vivo* (Collombat et al., 2009) while misexpression of *Arx* in β cells leads to their conversion to glucagon^{+} and PP^{+} cells (Collombat et al., 2007). *NeuroD*, a basic helix-loop-helix transcription factor and an *Ngn3* downstream target (Huang et al., 2000), is required for endocrine differentiation. Loss of *NeuroD* results in the reduction of all endocrine cells and a reduced β-to-α and β-to-δ cell ratio, suggesting the significance of *NeuroD* in regulating endocrine differentiation, especially β-cell lineage specification (Naya et al.,

1997). The homeodomain transcription factor *Nkx2.2* functions in both β-cell specification during embryogenesis and β-cell function maintenance in the adulthood (Doyle and Sussel, 2007; Sussel et al., 1998). *Nkx2.2*-deficient mice develop hyperglycemia due to the lack of β cells and die shortly after birth (Sussel et al., 1998). In β cells, Nkx2.2 can recruit a repressor complex composed of Groucho 3 (Grg3), DNA methyltransferase 3a (Dnmt3a) and histone deacetylase 1 (Hdac1) to enhance the methylation of *Arx* upstream regulatory elements and thus prohibit the acquisition of α cell fate (Mastracci et al., 2011; Papizan et al., 2011). The zinc finger transcription factor *Myt1* is expressed in endocrine progenitor cells as well as most mature endocrine cells (Wang et al., 2007). Loss of *Myt1* in the developing pancreas results in abnormal multi-hormone positive cells, suggesting that endocrine differentiation or maturation is disrupted (Wang et al., 2007). *Insm1*, another *Ngn3* downstream gene (Mellitzer et al., 2006), regulates pancreatic endocrine differentiation through a gene network that involves cell adhesion, cell migration, extracellular matrix remodeling, cell proliferation, and mRNA alternative splicing *etc.* Loss of *Insm1* decreases the delamination of pro-endocrine progenitors, leads to alternative splicing of Ngn3 mRNA, and decreases β-cell production (Osipovich et al., 2014). In addition to transcription factors, connective tissue growth factor (CTGF), an extracellular protein, also influences endocrine lineage allocation. Deletion of *CTGF* results in a reduced β-to-α cell ratio and altered islet morphology in the adult mice, although it is not clear how *CTGF* affects endocrine lineage allocation (Crawford et al., 2009). Since CTGF contains multiple domains that modulate TGF-β, BMP, and Wnt signaling, it could transduce extracellular signal into the nucleus and regulate the expression of endocrine lineage determination effectors.

Although much effort has been devoted to investigating the role of individual transcription factors in endocrine lineage determination, little information is known concerning how transcription factors concertedly specify endocrine lineages. To this end,

simultaneous loss of transcription factors is achieved in mouse models and endocrine lineage allocation is altered as a result. Such double knockout experiments unveil the interactions between transcription factors. For instance, simultaneously loss of *Arx* and *Pax4* promotes δ cell fate at the expense of α and β cell fate (Collombat et al., 2005). Compound loss of *Arx* and *Nkx2.2* results in somatostain^{+}ghrelin^{+} cells at the expense of α and β cells (Kordowich et al., 2011; Mastracci et al., 2011). Co-loss of *Nkx2.2* and *Pax4* results in the loss of β cells (Wang et al., 2004). In addition to global gene knockout, conditional deleting or overexpressing key endocrine transcription factors specifically in the Ngn3^{+} domain also causes the reallocation of endocrine lineages. For instance, loss of *Nkx6.1* in Ngn3^{+} progenitors results in the loss of β cells and increases non-β endocrine cells. The remaining insulin^{+} cells lack *MafA* and *Pdx1* expression and are probably dysfunctional, suggesting that *Nkx6.1* is necessary for both β-cell differentiation and maturation (Schaffer et al., 2013). On the contrary, *Ngn3^{Cre}*-mediated *Nkx6.1* overexpression favors β-cell allocation (Schaffer et al., 2013). In another report, ectopic *Pdx1* overexpression in the Ngn3^{+} domain promotes β-cell differentiation at the expense of α cells (Yang et al., 2011). A “combinatorial transcription factor code” is proposed that co-expression of transcription factors designates certain endocrine cell lineage(s) (Jorgensen et al., 2007). Indeed, we observed that the Ngn3^{+} progenitors are a heterogeneous population with a mosaic expression pattern with *Myt1*, *Nkx2.2*, *Pax4*, among many others (Fig 3.1 and references: Collombat et al., 2003; Soyer et al., 2010). Thus, we hypothesize that different Ngn3^{+} subpopulations (e.g., Ngn3^{+}Myt1^{+} and Ngn3^{+}Myt1^{-}) possess different differentiation potential.

However, the above data are obtained through loss-of-function or ectopic expression experiments. No definitive cell lineage tracing experiments have been done to directly test the possibility of transcription factor combination-mediated endocrine lineage determination. The currently available recombinase-based cell lineage tracing

techniques, including Cre/loxP and FLP/FRT systems, can hardly answer the above question because they can only capture progenitor cells marked by one molecular marker.

In order to analyze the differentiation potential of $Ngn3^+$ progenitor subsets, e.g., $Ngn3^+Myt1^+$ progenitor subpopulation, we utilized a novel bipartite Cre cell lineage tracing system. This bipartite Cre system allowed us to track cell lineages of double marked progenitors. With this tool, we found that $Ngn3^+Myt1^+$ progenitors favored β cells over α cells. It should be noted that pancreatic endocrine cells are differentiated successively according to a competence window model (Johansson et al., 2007). α cells are produced starting from E9.5, while β cells are produced mainly from E12.5 to E16.5. Our preliminary epigenetic studies showed that methylation of a *Myt1* regulatory region in the $Ngn3^+$ progenitors decreased from E10.5 to E15.5, which was correlated with elevated *Myt1* expression and β cell production. It is possible that enhanced *Myt1* expression in the $Ngn3^+$ progenitors initiated the β-cell differentiation program, although the involvement of other transcription factors or signaling pathways could not be excluded.

3.3 Results

$Ngn3^+$ progenitors are heterogeneous

In order to dissect the heterogeneity of endocrine progenitors, I first examined the co-expression pattern between *Ngn3* and several other endocrine transcription factors, including *Myt1*, *Nkx2.2*, *Pdx1*, and *Pax4* (Fig 3.1). In the E14.5 pancreas, 34.0±2.1% (n=6) of the $Ngn3^+$ progenitors were $Ngn3^+Myt1^+$ and 67.0±1.7% (n=8) of the $Ngn3^+$ progenitors were $Ngn3^+Nkx2.2^+$. This heterogeneity was not observed for every transcription factor though. For instance, *Pdx1* expression was detected in all $Ngn3^+$

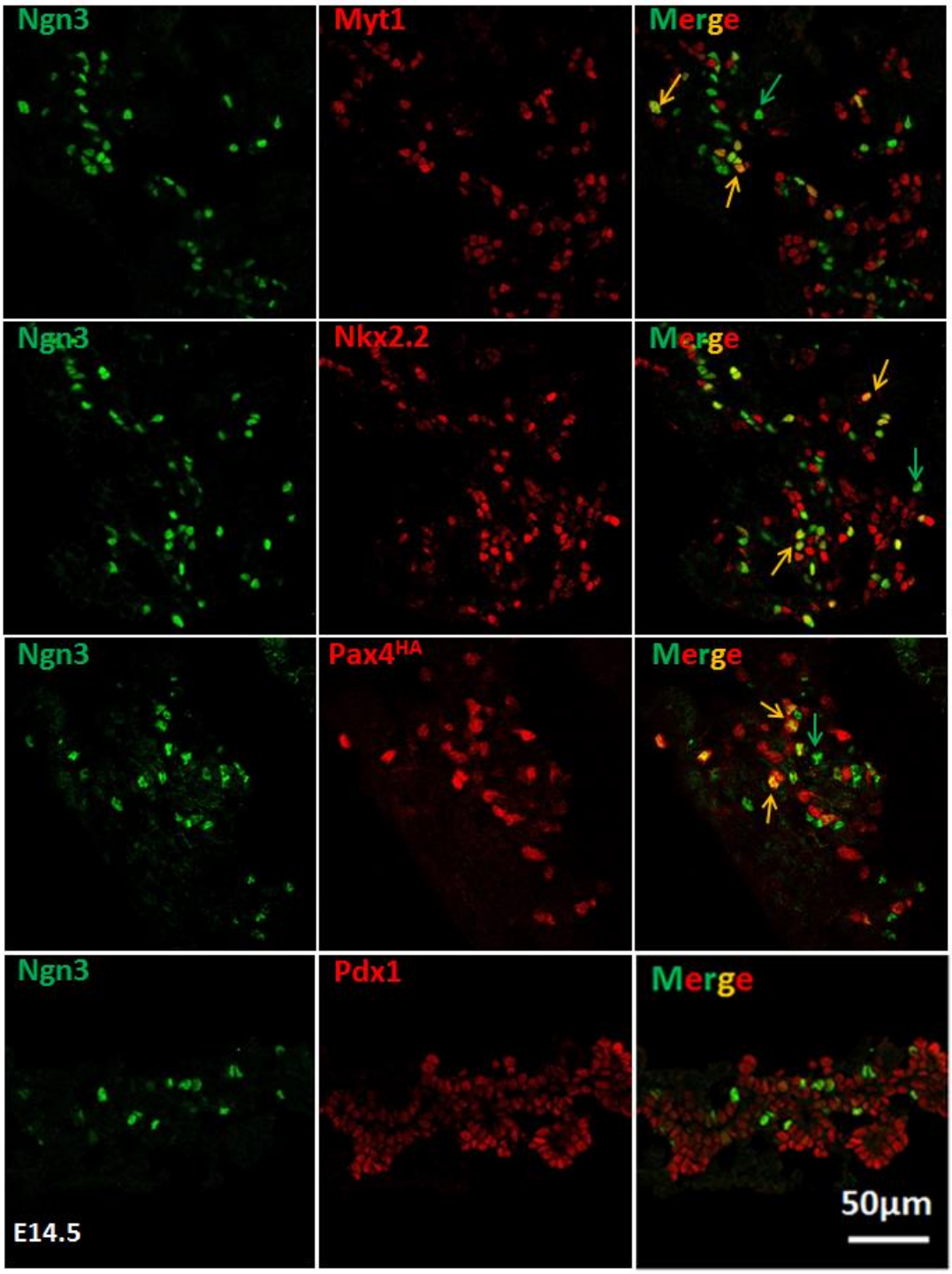


Fig 3.1 Ngn3^{+} progenitors are heterogeneous. Co-staining of Ngn3 and other transcription factors (TFs) Myt1, Nkx2.2 or Pax4^{HA} showed a mosaic co-expression pattern. Both Ngn3 single (green arrow) and Ngn3/TF double positive cells could be found. Pax4 expression was visualized by staining against HA tag in the *Pax4*$^{HA/+}$ mouse pancreas. *Pdx1* was expressed in all Ngn3^{+} progenitors. Stage: E14.5. Scale bar: 50µm.

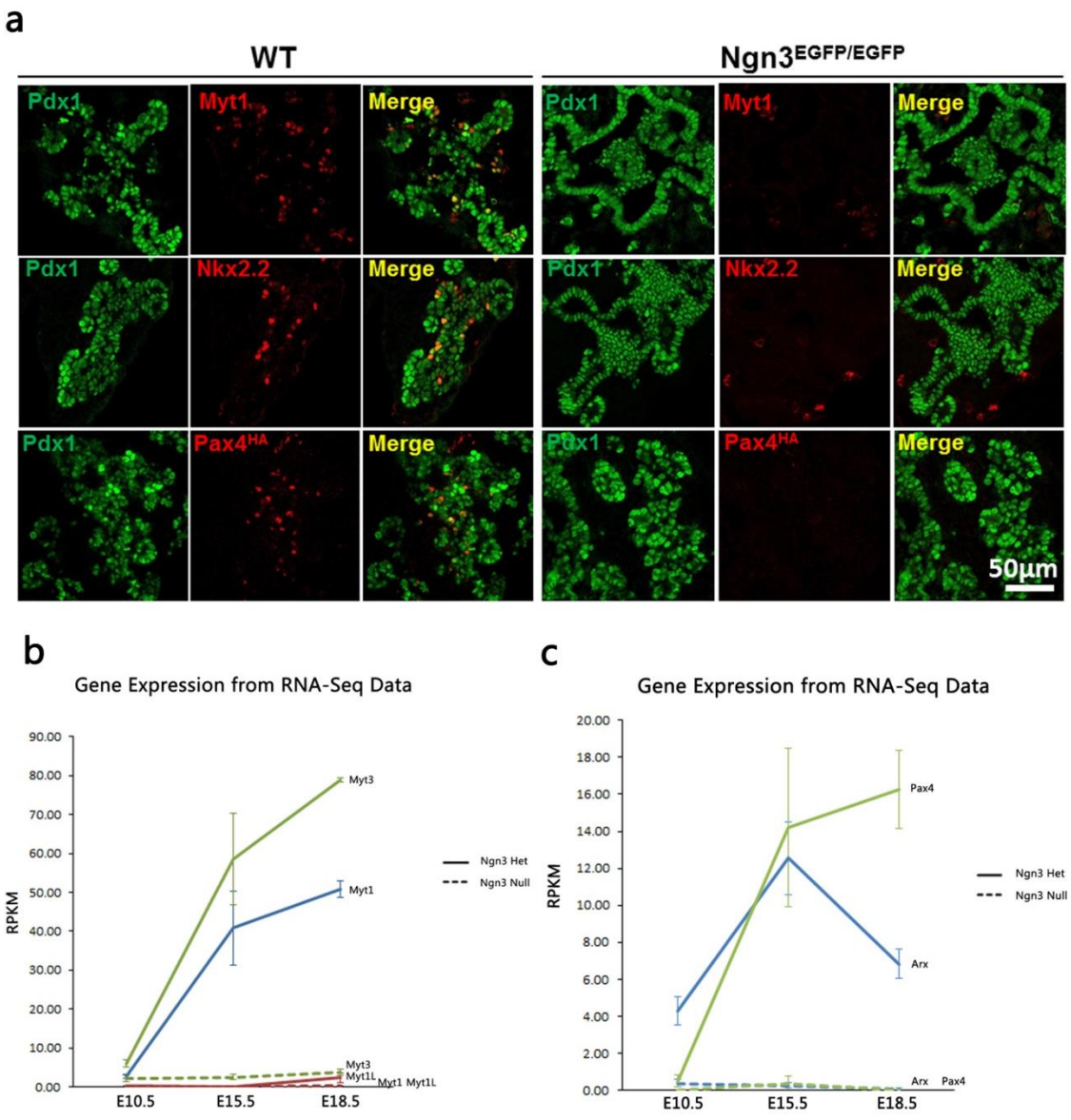


Fig 3.2 *Myt1* etc. expression is *Ngn3*-dependent. (a) Embryonic pancreas sections were stained with Myt1, Nkx2.2, HA (for Pax4), in together with Pdx1. Left panel: wild-type pancreas. Right panel: $Ngn3^{EGFP/EGFP}$ (Ngn3-null) pancreas. The expression of *Myt1*, *Nkx2.2* and *Pax4* was almost absent in the $Ngn3^{EGFP/EGFP}$ pancreas whereas *Pdx1* expression was largely unaffected. Stage: E14.5. Scale bar: 50μm. (b, c) Gene transcription data from RNA-Seq. y axis: normalized transcript abundance RPKM (readings per kilobase per million reads); x axis: three embryonic stages. Solid lines represent data from the $Ngn3^{EGFP/+}$ sample while dashed lines represent data from the $Ngn3^{EGFP/EGFP}$ sample. Lines of the same color in each graph represent the same gene.

cells, although *Pdx1* is a key transcription factor in establishing and maintaining β-cell fate (Gao et al., 2014; Yang et al., 2011). I also examined the co-expression pattern at E16.5 and the percentages were not significantly different: 32.7±1.3% (n=8) of $Ngn3^+$ progenitors were $Ngn3^+Myt1^+$ and 63.6±2.8% were $Ngn3^+Nkx2.2^+$ (n=8), implying that the heterogeneous pattern did not shift along development. Though without quantification, I also found that *Pax4* was expressed only in a subset of $Ngn3^+$ progenitors (Fig 3.1).

Interested in whether this heterogeneous co-expression pattern was *Ngn3*-dependent, I performed co-staining in both wild-type and $Ngn3^{EGFP/EGFP}$ pancreas. $Ngn3^{EGFP}$ allele is an *Ngn3*-knockout allele in which the EGFP cassette is knocked into the endogenous *Ngn3* locus to replace the entire coding region of *Ngn3* (Lee et al., 2002). I did not detect *Myt1* or *Nkx2.2* expression in the pancreatic epithelium of the $Ngn3^{EGFP/EGFP}$ pancreas, indicating that the expression of *Myt1*, *Nkx2.2* and *Pax4* depended on *Ngn3* (Fig 3.2 a).

Ngn3-dependent expression pattern was corroborated by RNA-Seq data. Previously, the Gu lab obtained RNA-Seq data for $Ngn3^+$ progenitors (using EGFP as a surrogate marker for *Ngn3* expression; data unpublished) at E10.5, E15.5 and E18.5. At all stages, *Myt1*, *Nkx2.2*, *Pax4*, and *Arx* showed negligible expression in the $Ngn3^{EGFP/EGFP}$ pancreas in contrast to the $Ngn3^{EGFP/+}$ pancreas (Fig 3.2 b, c), suggesting that these transcription factors were *Ngn3*-dependent. The RNA-Seq data were consistent with immunostaining results (Fig 3.2 a) and previous reports that these transcription factors were *Ngn3* downstream targets (Collombat et al., 2003; Smith et al., 2003; Wang et al., 2008; Watada et al., 2003).

Design of a novel bipartite Cre cell lineage tracing system

In order to understand the endocrine lineage allocation process in pancreas development, I wanted to follow the cell fate of $Ngn3^+$ progenitor subsets, for example, $Ngn3^+Myt1^+$ cells. However, the conventional recombinase-based cell lineage tracing system (Cre/loxP or FLP/FRT) is not sufficient to track the cell fate of double-marked cells. Dual-reporters, which take advantage of both the Cre/loxP and FLP/FRT recombination system, are created (Imayoshi et al., 2012; Yamamoto et al., 2009). Such dual-reporters usually have two tandem cassettes each flanked by loxP or FRT sites. The activation pattern of the dual-reporters varies depending on how the dual-reporters are designed and should be examined case-by-case. The principle underlying the dual-reporters is the same; that is, reporter activation pattern reflects whether one or two genes of interest are active. However, these dual-reporters cannot address our research question either because the recombination events mediated by Cre and FLP are independent. Dual-reporters cannot distinguish between successive recombination and simultaneous recombination events. The expression time frames of the two genes used to drive the expression of Cre and FLP do not necessarily have to overlap for reporter activation. In the pancreas, *Ngn3* and *Myt1* are both expressed in progenitor cells. The expression window of *Ngn3* in an individual progenitor cell is less than 48 hours (Gu et al., 2002; Schwitzgebel et al., 2000). Then *Ngn3* expression diminishes in the mature endocrine cells whereas *Myt1* expression persists in most endocrine cells (Wang et al., 2007). Therefore, these dual-reporters would not allow us to capture and only capture the real $Ngn3^+Myt1^+$ progenitors as long as two recombination events are involved.

To circumvent this obstacle, the Gu lab designed a bipartite Cre cell lineage tracing system (Fig 3.3). In this system, the Cre protein is split into two halves, the N-terminal half (nCre) and the C-terminal half (cCre). Neither of the two halves is functional unless they reassemble into a full Cre. We also added nuclear localization signal (NLS)

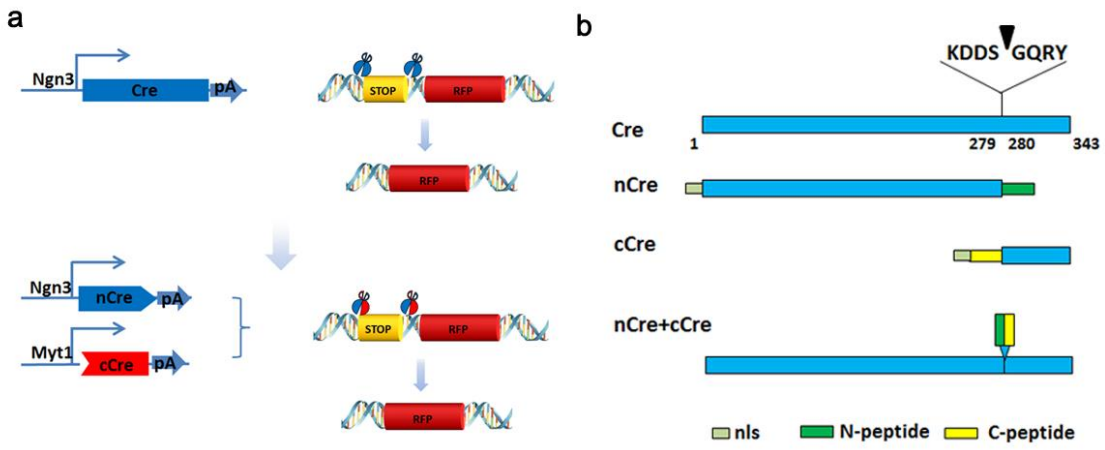


Fig 3.3 Design of the bipartite Cre. (a) Conception of the bipartite Cre. In conventional Cre-mediated DNA recombination, one promoter (e.g., Ngn3) drives the expression of full length Cre, which subsequently excises a stop cassette and activates reporter expression (e.g., red fluorescent protein (RFP) reporter). However, this strategy only allows tracking progenitor cells with one marker (e.g., Ngn3). In the bipartite Cre design, full length Cre recombinase is split into two halves and driven by distinct promoters (e.g., Ngn3 and Myt1). Neither Cre half is functional unless they are expressed in the same cell and reconstitute into a full Cre, which in turn activates the reporter. (b) Schematic design of bipartite Cre. Wild type Cre is split between S279 and G280. Leucine zipper (N- and C- peptide) and nuclear localization signal (NLS) sequences are fused to nCre and cCre to facilitate reconstitution and nuclear localization, respectively.

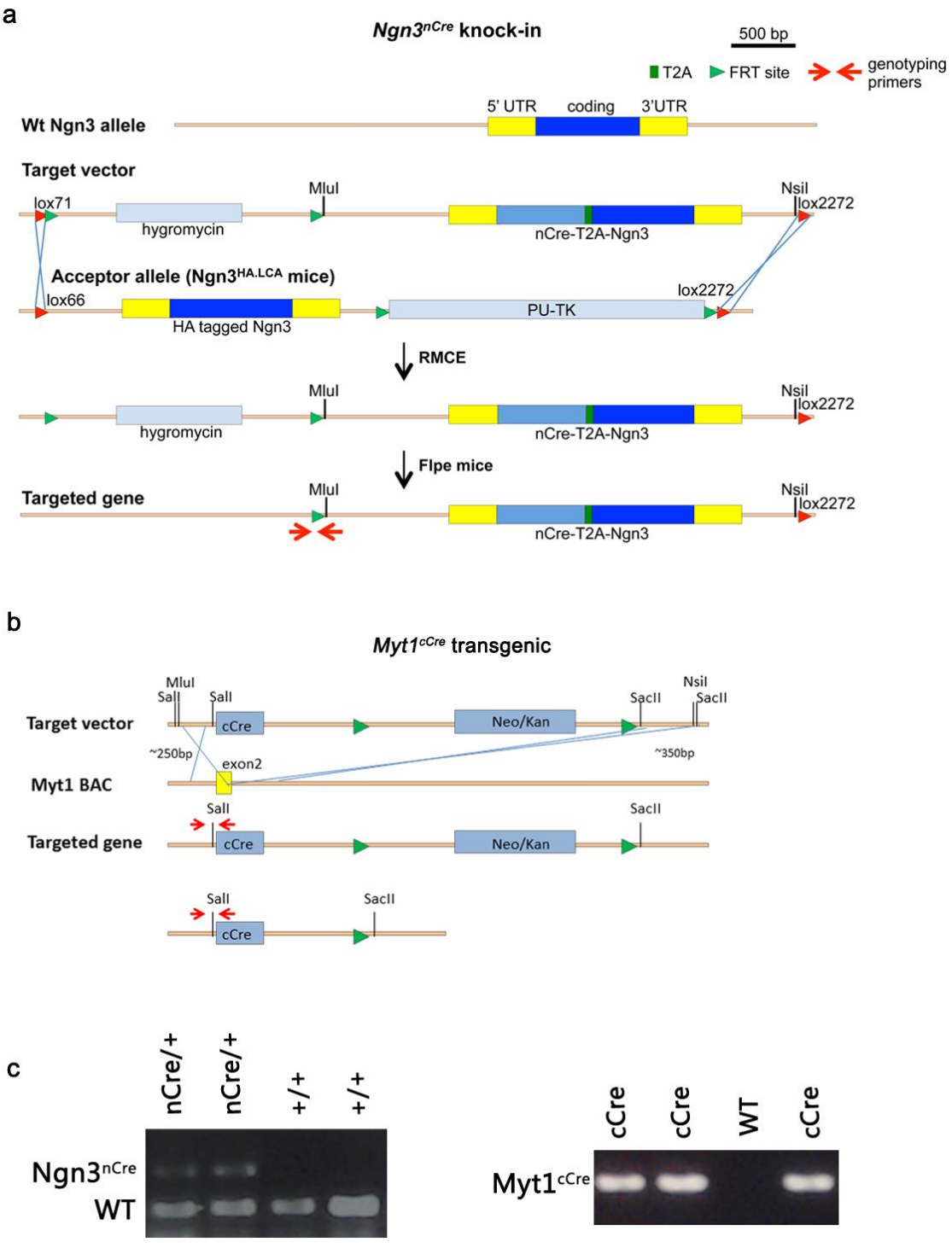

a
Ngn3nCre knock-in
500 bp
T2A
FRT site
genotyping primers
Wt Ngn3 allele
5' UTR
coding
3'UTR
Target vector
lox71
MluI
NsiI
lox2272
hygromycin
nCre-T2A-Ngn3
Acceptor allele (Ngn3HA.LCA mice)
lox66
lox2272
HA tagged Ngn3
PU-TK
RMCE
Flpe mice
Targeted gene
b
Myt1cCre transgenic
MluI
SalI
NsiI
SacII
Target vector
cCre
Neo/Kan
~250bp
~350bp
exon2
Myt1 BAC
Targeted gene
c
nCre/+
+/+
Ngn3nCre
WT
cCre
WT
Myt1cCre

Fig 3.4 Generation of mouse lines and gene targeting scheme. (a) Gene targeting scheme of *Ngn3*^{nCre} allele. A targeting vector containing nCre-T2A-Ngn3 is inserted into the endogenous locus of *Ngn3* through RMCE. An existing *Ngn3*$^{HA.LCA}$ allele serves as the loxed cassette acceptor. The nCre-T2A-Ngn3 coding sequence follows the ATG starting codon of endogenous *Ngn3*. The FRT-flanking hygromycin selection marker is removed via crossing with a transgenic FLPe mouse. (b) Generation of the transgenic *Myt1*^{cCre} mouse. *Myt1*^{cCre} mouse is generated through BAC recombineering. A targeting vector containing cCre is first inserted into the Myt1 BAC to harness Myt1 regulatory sequence. Recombined BAC is then introduced into the mouse genome by pronuclear injection. (c) Genotyping bands.

and leucine zipper sequences to both the nCre and cCre to facilitate their nuclear localization and reconstitution. The initial work was published in *Nucleic Acids Research* in 2007 (Xu et al., 2007). Since then, the Gu lab has optimized this bipartite Cre system, including optimizing codon usage for the expression in mammals, and selecting a new split site to improve the reassembly and the reconstituted recombinase activity. In the current version of bipartite Cre, the split site is between Ser299 and Gly280. This position resides in a flexible region between α-helix J and K, minimizing the disruption of Cre three-dimensional structure (Guo et al., 1997). A similar split Cre system has been used to map the fate of neural stem cells (Beckervordersandforth et al., 2010; Hirrlinger et al., 2009b), but ours is the first one to be used in pancreas development research. Besides, the split site choice and the peptide used for facilitating reconstitution are different in the two designs.

I wanted to use *Ngn3* and *Myt1* promoters to drive the expression of nCre and cCre respectively so that I could perform lineage tracing for the $Ngn3^{+}Myt1^{+}$ progenitors. Because *Ngn3* expressing level affects cell lineage allocation during pancreas development (Wang et al., 2010), I decided to make an Ngn3-T2A-nCre knock-in mouse by utilizing the T2A peptide to avoid potential haploinsufficiency effect. T2A peptide is an 18-amino acid (AA) peptide that allows bicistronic expression from a single mRNA through a ribosome skip mechanism (Donnelly et al., 2001). After its cleavage, the T2A peptide leaves a 17-AA tail in the leading protein and a proline in the N-terminus of the following protein. Concerned about the effect of the 17-AA tail and the leading proline on the stability and function of Ngn3 and nCre (Varshavsky, 1996), I first compared Ngn3-T2A-nCre and nCre-T2A-Ngn3 in a cell transfection-based reporter assay. I found that nCre-T2A-Ngn3 gave higher recombinase activity than Ngn3-T2A-nCre when other conditions were the same. 62.7±1.8% (n=3) of transfected cells underwent recombination by the nCre-T2A-Ngn3 (in together with a cCre) while only 49.4±2.7%

(n=3) of transfected cells were recombined by the Ngn3-T2A-nCre (p value = 0.01). Therefore, I decided to create an *Ngn3*$^{nCre\text{-}T2A\text{-}Ngn3}$ knock-in allele (*Ngn3*^{nCre} thereafter). In collaboration with the Transgenic Mouse/ES Share Resource at Vanderbilt, we used the recombination-mediated cassette exchange (RMCE) strategy to generate the *Ngn3*^{nCre} knock-in mouse. The nCre-T2A-Ngn3 target vector was electroporated into embryonic stem (ES) cells derived from the *Ngn3*$^{HA.LCA}$ mouse (mouse generated by the Gu lab; data unpublished yet) and knocked into the endogenous *Ngn3* locus (Fig 3.4 a).

Instead of making a knock-in *Myt1*^{cCre} mouse, we created a transgenic *Myt1*^{cCre} allele because a previously made knock-in *Myt1*^{cCre} failed to achieve robust recombination activity. In making the *Myt1*^{cCre} mouse, we utilized a bacteria artificial chromosome (BAC) recombineering strategy such that as many regulatory sequences as possible were preserved in the transgenic mice (Fig 3.4 b).

Because reconstituted bipartite Cre had lower activity than full length Cre (Xu et al., 2007), we decided to use the *Rosa26*^{tdTomato} (*Ai9* for short) reporter for lineage tracing analysis (Madisen et al., 2010). *Ai9* is a very sensitive reporter possibly due to its relatively shorter stop cassette than most other reporters (Liu et al., 2013). tdTomato displays bright signal under the microscope and its signal is preserved even after paraformaldehyde fixation, making it an ideal fluorescent protein reporter for cell lineage tracing.

Characterization of mouse models

I first characterized the mouse models before proceeding to perform lineage tracing experiments.

First of all, the transgenes did not affect the overall mouse physiology. Mouse litters were born in a Mendelian ratio. Their body weight and blood glucose level were measured and no significant differences were observed between the

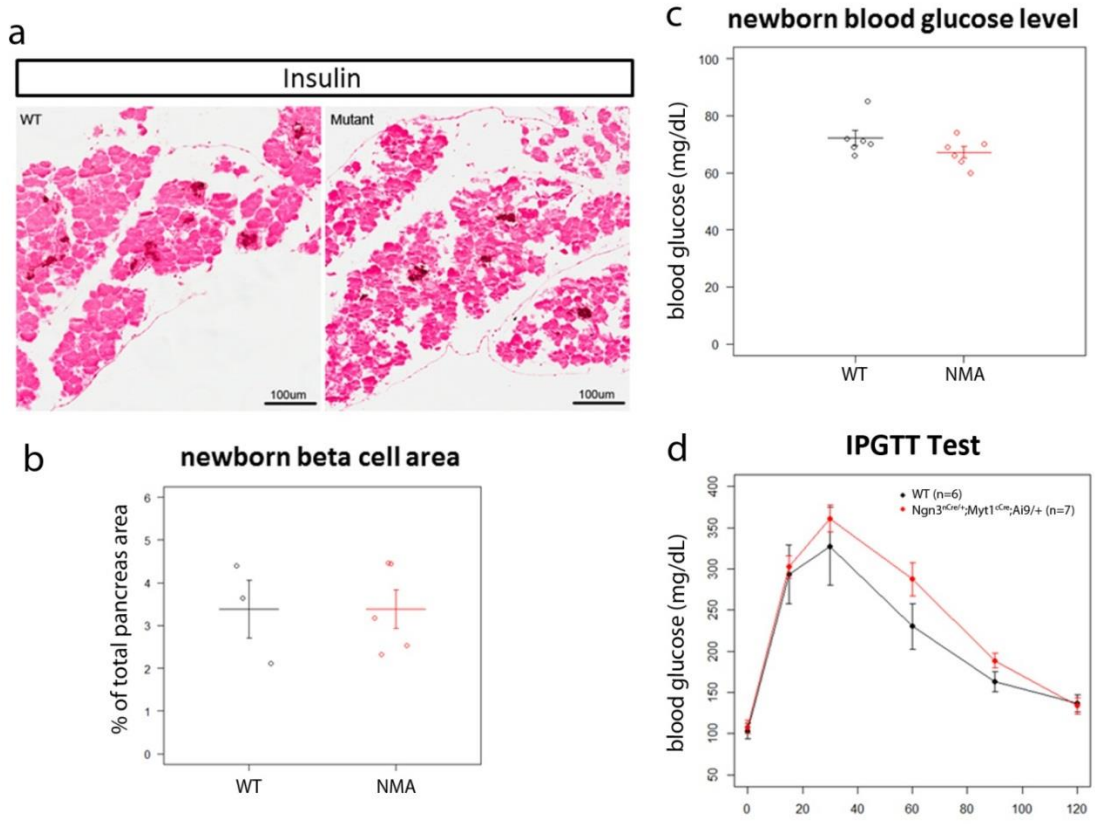


Fig 3.5 Characterization of the bipartite Cre cell lineage tracing mouse model. (a, b) Quantification of β-cell amount in P0 pancreas. (a): Insulin staining. P0 pancreas sample is stained with insulin antibody and signal is amplified with DBA staining. Sample is counter-stained with eosin. (b): quantification of β-cell amount. β-cell amount is measured as the percentage of total pancreas area. There is no significant difference between the wild-type pancreas and the *Ngn3*$^{nCre/+}$*;Myt1*^{cCre}*;Ai9/+* (NMA) pancreas. (c) Measurement of newborn blood glucose level. No significant difference in blood glucose concentration is detected between the wild-type and the *Ngn3*$^{nCre/+}$*;Myt1*^{cCre}*;Ai9/+* neonates . (d) Intraperitoneal glucose tolerance test (IPGTT) of the wild-type and the *Ngn3*$^{nCre/+}$*;Myt1*^{cCre}*;Ai9/+* adult mice (6 weeks). Glucose clearance capacity is not impaired in the *Ngn3*$^{nCre/+}$*;Myt1*^{cCre}*;Ai9/+* mice.

Ngn3$^{nCre/+}$*;Myt1*^{cCre}*;Ai9/+* trigenic mice and their wild-type littermates. Resting blood glucose of the trigenic neonates was 67.2±2.0mg/dL (n=6), which was indistinguishable from the wild-type newborns (72.2±2.5mg/dL, n=6; p value=0.171) (Fig 3.5 c).

Second, β-cell mass (measured by the percentage of insulin^{+} cells out of total pancreas area) was not significantly different between the trigenic mice and the wild-type mice. At P0, 3.39±1.02% (n=5) of the total pancreas area of the trigenic mice was insulin^{+} and the percentage for the wild-type mice was 3.38±1.16% (Fig 3.5 a-b) (n=3, p value=0.996).

Mouse physiology was also normal in the adulthood (Fig 3.5 d). At 6 weeks, the blood glucose level of the *Ngn3*$^{nCre/+}$*;Myt1*^{cCre}*;Ai9/+* mice after overnight fasting was 107.9±8.0mg/dL (n=7) and was not statistically different from that of the wild-type mice (102.8±9.6mg/dL, n=6; p value=0.695). I also performed intraperitoneal glucose tolerance test (IPGTT) assay on these mice and did not observe any significant differences either. Although the *Ngn3*$^{nCre/+}$*;Myt1*^{cCre}*;Ai9/+* mice showed higher blood glucose level (361.0±16.3mg/dL) than their wild-type littermates (327.3±47.6mg/dL) 15 minutes after IP glucose challenge, the difference was not statistically significant (p value=0.528) and the trigenic mice could quickly clear excessive blood glucose. By 2 hours, the blood glucose level returned to normal (133.6±10.0mg/dL for trigenic versus 136.8±10.5mg/dL for wild type). There appeared no significant differences between female and male trigenic mice.

Last but not least, this bipartite Cre system was stringently regulated. As expected, tdTomato was only expressed in the islets when the *Ai9* reporter was activated (Fig 3.6). In addition, of all the genotypes I examined, only the *Ngn3*$^{nCre/+}$*;Myt1*^{cCre}*;Ai9/+* pancreas expressed tdTomato. This suggested that the bipartite Cre system is a very robust system and only labels double-marked cells.

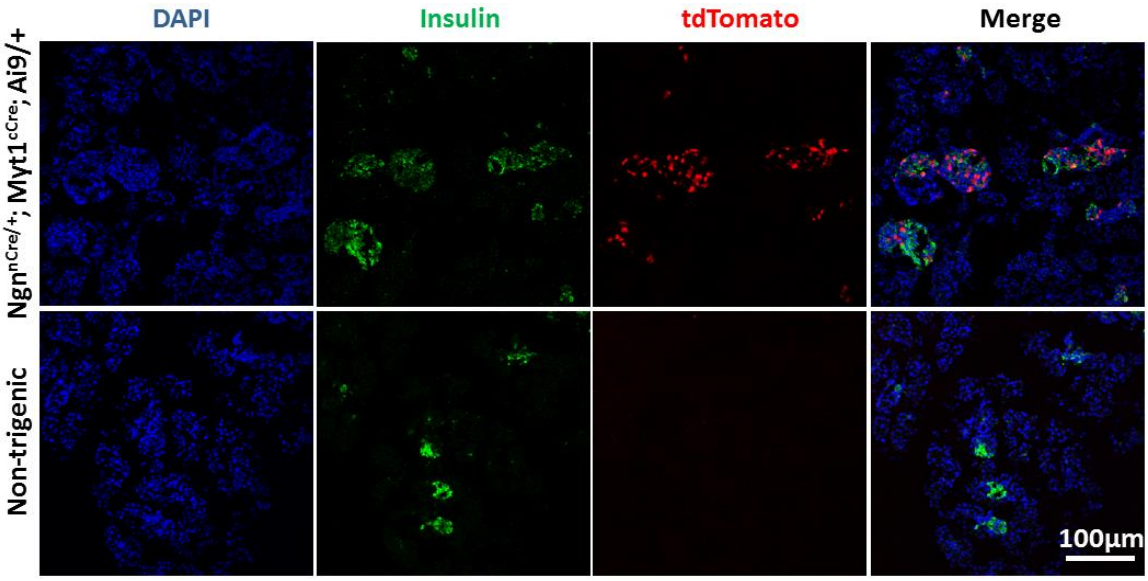


Fig 3.6 Overall labelling of pancreas of the bipartite Cre mouse model. low magnification confocal images showing the overall expression pattern of tdTomato. Upper panel: *Ngn3*$^{nCre/+}$;*Myt1*^{cCre};*Ai9*/+mouse pancreas section. Lower panel: representative images from all other non-trigenic mice. tdTomato (*Ai9* reporter) is only activated in the *Ngn3*$^{nCre/+}$;*Myt1*^{cCre};*Ai9*/+ mouse islets. Stage: P0. Scale bar: 100µm.

Lineage tracing shows β-cell biased cell fate of Ngn3^{+}Myt1^{+} progenitors

I and Dr. Gu then examined the cell lineages derived from Ngn3^{+}Myt1^{+} progenitors with the *Ngn3$^{nCre/+}$;Myt1^{cCre};Ai9/+* mouse model. P0 pancreas was dissected and the labeling of each endocrine cell type by the *Ai9* reporter was examined by immunostaining, imaging and quantification.

Cell lineage tracing quantification showed β-cell biased choice made by the Ngn3^{+}Myt1^{+} progenitors (Fig 3.7). At P0, we found that 52.6±2.9% of the insulin^{+} cells were labelled by the *Ai9* reporter while this number was only 12.7±1.9% for the glucagon^{+} cells, 35.5±2.4% for the SS^{+} cells, and 55.9±4.0% for the PP^{+} cells (n=8-13). Since dramatically more insulin^{+} cells than glucagon^{+} cells were labelled by the *Ai9* reporter ($p<0.0001$), this suggested a β-cell biased choice made by the Ngn3^{+}Myt1^{+} progenitors. This cell fate bias appeared as early as E16.5 when the Ngn3^{+}Myt1^{+} progenitors contributed to 57.1±4.9% of the β cells but only 12.9±1.5% of the α cells ($p<0.0001$, n=7-8). Moreover, we obtained consistent results when using the *Rosa26^{EYFP}* reporter (Srinivas et al., 2001) that the Ngn3^{+}Myt1^{+} progenitors favored β cells over α cells. These results suggested that earlier Ngn3^{+}Myt1^{+} progenitors preferentially gave rise to β cells, but not α cells. At late gestational stages, the Ngn3^{+}Myt1^{+} progenitors also preferentially gave rise to δ and PP cells.

DNA methylation of a Myt1 regulatory element in the Ngn3^{+} progenitors decreases over development

The Gu lab previously obtained RNA-Seq data for Ngn3^{+} progenitors at E10.5, E15.5 and E18.5 and the data allowed us to assess the temporal expression pattern of major endocrine transcription factors. For instance, *Myt1* and its homologous gene *Myt3* all showed low expression at E10.5 but dramatic increase at E15.5 and continued high expression at E18.5. Another *Myt* gene member, *Myt1L*, was minimally expressed

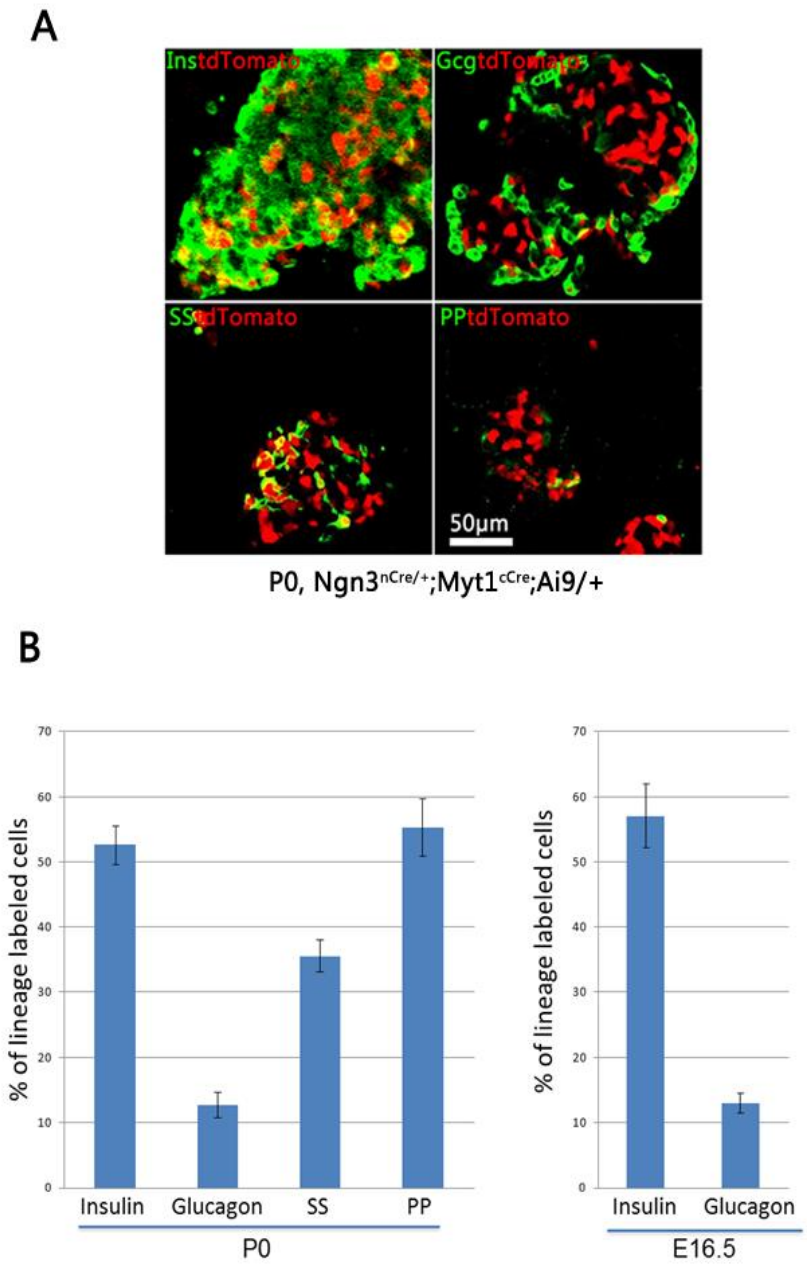


Fig 3.7 Bipartite Cre cell lineage tracing result and quantification. (A) Pancreas sections were stained with anti-endocrine hormone antibodies (green). Red: tdTomato. Stage: P0. Genotype: *Ngn3*$^{nCre/+}$*;Myt1*^{cCre}*;Ai9/+*. Scale bar: 50μm. (B) Quantification of cell lineage tracing results at two time points, P0 and E16.5.

across all stages, probably due to genetic redundancy and expression repression (Wang et al., 2007) (Fig 3.2 b). *Pax4* expression was negligible at E10.5 while *Arx* displayed robust expression. At 15.5, although both *Pax4* and *Arx* showed relatively high mRNA amount, *Arx* was outcompeted by *Pax4* (Fig 3.2 c). The balance between *Pax4* and *Arx* is important in specifying β- versus α-cell lineages because of their antagonizing relationship (Collombat et al., 2003). Our RNA-Seq data of *Pax4* and *Arx* fitted well in the competence window model (Johansson et al., 2007): *Arx* was initially expressed and α cells were mainly produced; later on, *Pax4* dominated *Arx* and β cells were generated during the secondary transition.

To examine what caused the change in gene transcription, Dr. Gu and I investigated the epigenetic status of $Ngn3^{+}$ progenitors. We focused on DNA methylation of key transcription factors using a bisulfite sequencing assay. The bisulfite sequencing data will provide us with information of gene regulation and endocrine lineage specification.

In choosing candidate loci for DNA methylation analysis, I consulted both the literature and bioinformatics prediction. To this end, I focused on a regulatory element we coined as Myt1 region 2. Myt1 region 2 lies in the first intron of *Myt1* gene and is a predicted CpG island by the MethPrimer algorithm (Li and Dahiya, 2002). In addition, this Myt1 region 2 overlaps with a reported conserved promoter region across species, which is supposedly bound by Ngn3 (Wang et al., 2008).

We made several interesting discoveries about the DNA methylation status of Myt1 region 2 (Fig 3.8). First, it seemed that DNA methylation of Myt1 region 2 was *Ngn3*-independent because the average DNA methylation rate of the region was not significantly different between the $Ngn3^{EGFP/+}$ and the $Ngn3^{EGFP/EGFP}$ pancreas and this was true for both E10.5 and E15.5 stages. This implied that *Ngn3* was not involved in the deposition, maintenance, or removal of DNA methylation marker in Myt1 region 2.

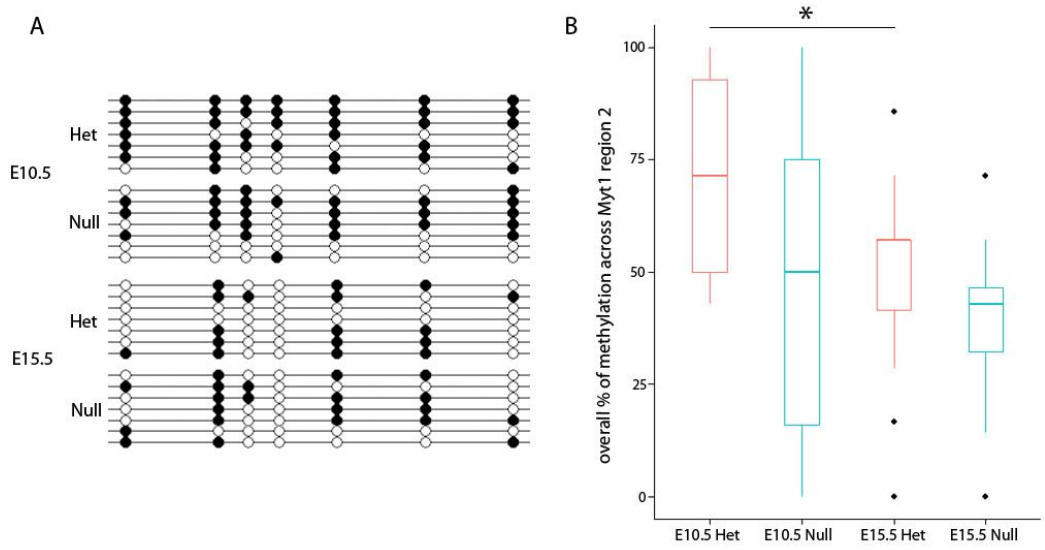


Fig 3.8 Temporal change of DNA methylation status of Myt1 region 2. (A) Bisulfite sequencing results of Myt1 region 2 (+3684bp to +3865bp). Two types of samples were used: Het (*Ngn3*$^{EGFP/+}$) and Null (*Ngn3*$^{EGFP/EGFP}$). EGFP^{+} cells were sorted by FACS and subject to bisulfite conversion, amplification and sequencing. Black dots represent methylated CpG dinucleotides. Open circles represent unmethylated CpG dinucleotides. (B) Box-whisker graph of the quantification of CpG methylation. EGFP^{+} progenitors of E10.5 had higher DNA methylation at Myt1 region 2 than that of E15.5.

Secondly, DNA methylation of Myt1 region 2 was statistically higher in the EGFP^{+} progenitors isolated from E10.5 *Ngn3*$^{EGFP/+}$ pancreas (71.4±9.4%, n=7) than the E15.5 counterpart (47.5±5.2%, n=23). DNA methylation data was consistent with our RNA-Seq data. It is possible that DNA methylation-mediated *Myt1* transcription repression restricted *Myt1* expression at a low level at E10.5, when α cells were primarily generated. As development progresses, hypomethylation of Myt1 region 2 led to elevated *Myt1* expression, which facilitated β-cell generation. Whether this was passive loss of DNA methylation or active demethylation was not clear though.

Interfering with DNA methylation alters β-to-α cell ratio

Because the methylation status of Myt1 region 2 decreased from E10.5 to E15.5, Dr. Gu and I wanted to know whether the disruption of DNA methylation would change endocrine cell lineage allocation. To this end, we attempted two methods to manipulate DNA methylation.

In the first method, we cultured embryonic pancreas *in vitro* in the presence of a small chemical inhibitor, adenosine periodate oxidized (Adox) to inhibit overall methylation. In this *in vitro* culture assay, E11.5 pancreas was dissected out and cultured on Matrigel or porous membrane filter for three days before fixation and examination by either whole mount staining or section staining. Though preliminary, our data showed that 10μM Adox treatment increased the insulin^{+} to glucagon^{+} cell ratio (Fig 3.9). In order to validate this result and examine the changes of gene expression that caused this elevated β-to-α cell ratio, I performed quantitative real-time PCR (qRT-PCR) to detect the expression level of hormone and endocrine transcription factor genes. However, due to the fact that the enveloping mesenchyme cannot be fully dissected away from the pancreatic epithelium, qRT-PCR data fluctuated depending on how much the epithelium was contaminated by the mesenchyme (data not shown). An epithelium-

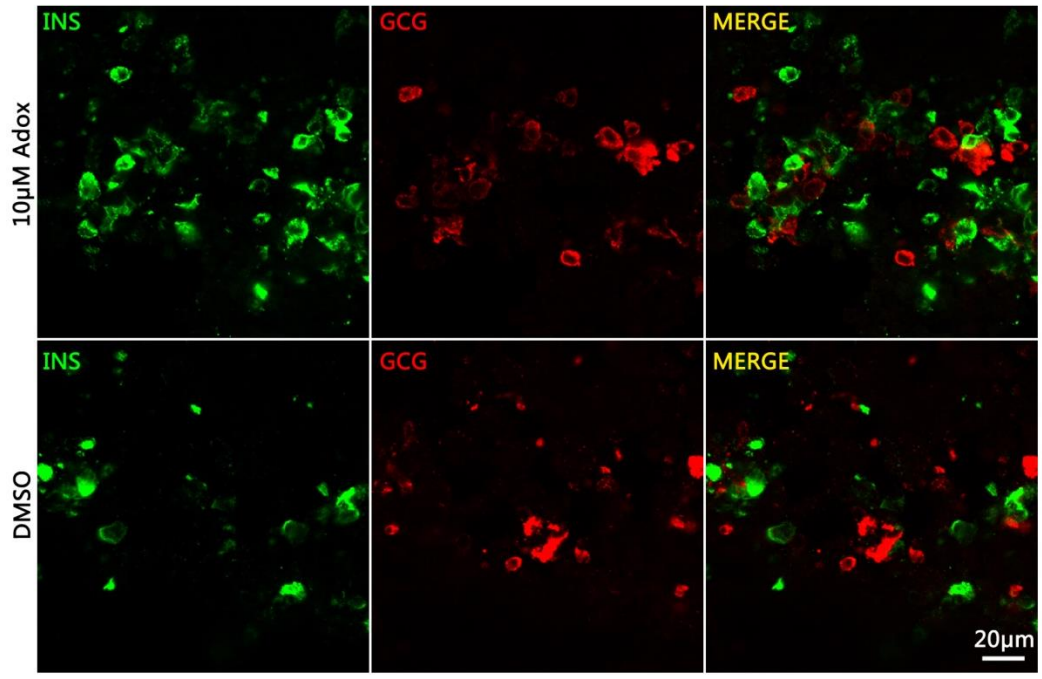


Fig 3.9 Adox treatment increases β-to-α cell ratio in *in vitro* cultured pancreatic buds. Either 10μM Adox or DMSO was added into the culture medium for *in vitro* pancreatic buds culturing. Pancreatic buds were dissected out from E11.5 wild-type embryos and cultured for 3 days. Sections were prepared on the 3rd day and stained with anti-insulin and anti-glucagon antibodies.

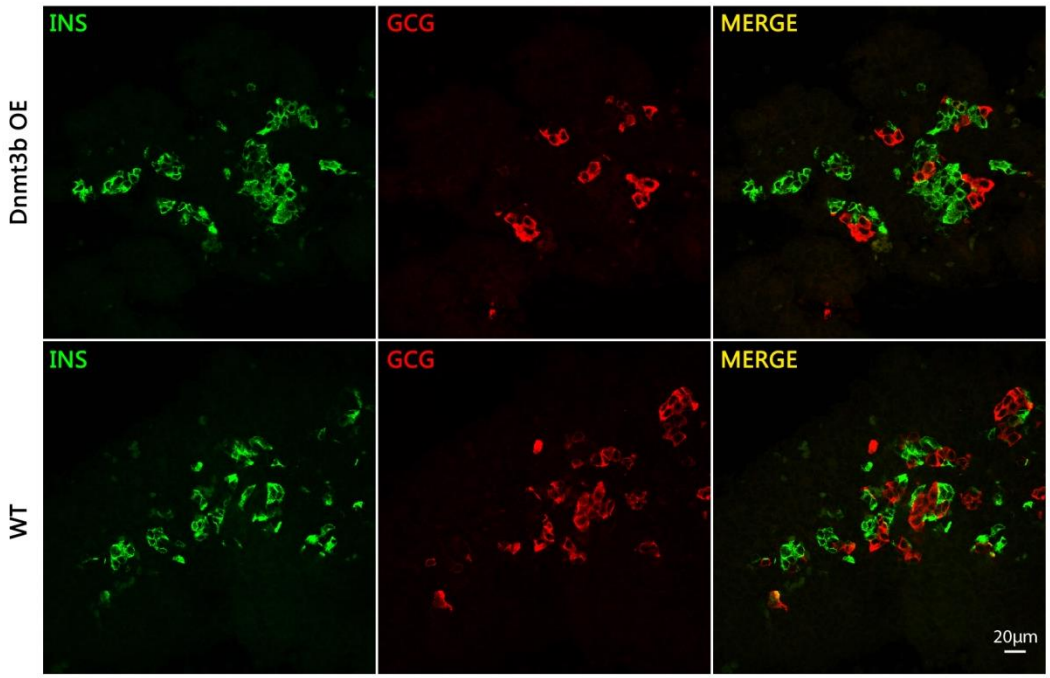


Fig 3.10 *Dnmt3b* overexpression results in an increased β-to-α cell ratio. Timed pregnant female mice were administered with doxycycline intraperitoneally from E8.5 to E15.5 to induce *Dnmt3b* overexpression in the *Rosa26*$^{M2\text{-}rtTA/+}$;*ColA1*$^{tetOP\text{-}Dnmt3b/+}$ fetus. Pancreas was dissected at E15.5 and analyzed with standard IHC by staining sections with anti-insulin and anti-glucagon antibodies. Scale bar: 20μm.

specific normalization control, in contrast to the commonly used house-keeping genes *GAPDH* or *PPIA*, will possibly solve this problem. Nonetheless, whether the expression of such an epithelium-specific gene is influenced by Adox treatment remains unclear. It will be helpful to include several epithelium-specific genes to determine whether anyone of them can serve as a proper normalization control.

In a second strategy, we attempted to increase DNA methylation with an inducible *Dnmt3b* (DNA methyltransferase 3b) overexpression model (Linhart et al., 2007). In this *Rosa26*$^{M2\text{-}rtTA/+}$;*Col1A1*$^{tetOP\text{-}Dnmt3b/+}$ model, *Dnmt3b* overexpression was induced by doxycycline administration. We treated timed pregnant female mice with daily intraperitoneal injection of 150μl 5mg/ml doxycycline from E8.5 to E15.5. At the end of E15.5, we examined the endocrine lineage allocation of the *Rosa26*$^{M2\text{-}rtTA/+}$;*Col1A1*$^{tetOP\text{-}Dnmt3b/+}$ embryos with immunostaining. We observed that *Dnmt3b* overexpression increased β-to-α cell ratio (Fig 3.10 and preliminary quantification data not shown).

3.4 Discussion

How seemingly equivalent progenitor cells give rise to different cell types remains an outstanding question. We utilize endocrine islet cell differentiation in the mouse pancreas to dissect this process. In this case, pancreatic progenitor cells that transiently produce a bHLH transcription factor Ngn3 differentiate into all endocrine cell types (α, β, δ, PP and ε). Because each *Ngn3*-expressing cell only gives rise to one endocrine cell, we postulate that the *Ngn3*-expressing cells can be divided into subpopulations, each of which has a preference for specific endocrine lineage(s). Indeed, we found that Ngn3^{+} progenitors can be classified into two cell populations based on their co-expression with *Myt1* (i.e., Ngn3^{+}Myt1^{+} and Ngn3^{+}Myt1^{-} cells). We developed an effective bipartite Cre-based technique to reconstitute Cre activity in cells that co-expressed the two proteins.

This method allowed us to examine whether the subset of $Ngn3^+$ cells that co-expressed *Myt1* represented specialized progenitor cells. Genetic lineage tracing showed that the Ngn3 and Myt1 proteins preferably marked progenitors for β cells, but not that for α cells. Interestingly, we further showed that the activation of *Myt1* in subsets of $Ngn3^+$ cells depended on the activity of *Ngn3*. Because *Myt1* is a downstream target of *Ngn3*, the phenomenon of *Myt1* expression in selective $Ngn3^+$ progenitors implies that activation of *Myt1* in the $Ngn3^+$ progenitors probably arises from unidentified stochastic events. Preliminary DNA methylation analysis demonstrated the temporal change of a regulatory element of *Myt1* gene, Myt1 region 2. This element was hypermethylated in E10.5 $Ngn3^+$ progenitors, which was correlated with low *Myt1* expression as revealed by RNA-Seq and more α-cell production at this stage. At E15.5, loss of DNA methylation derepressed *Myt1* expression, leading to substantial *Ngn3* and *Myt1* co-expression in the progenitors, which were funneled into a β-cell differentiation program.

Allocation of endocrine lineages from pro-endocrine progenitors is a cryptic process and deciphering this process requires investigation at multiple levels that regulate gene expression and ultimately affect cell fate choice, including cell signaling, epigenetic regulation, transcriptional and translational regulation, as well as post-transcriptional and post-translational regulation, *etc*. Here, I focused on transcriptional analysis with RNA-Seq and DNA methylation with bisulfite sequencing. It is desirable to extract the most information from our multi-stage RNA-Seq data by means of bioinformatics tools such as gene clustering and gene ontology annotation, which will provide invaluable information of the changes that $Ngn3^+$ pro-endocrine progenitors undergo during endocrine differentiation. In addition, a systematic profiling of the DNA methylome and histone modification will reveal important clues for decoding the enigma of endocrine differentiation. Another intriguing and important question regards the differences between $Ngn3^+Myt1^+$ cells and $Ngn3^+Myt1^-$ cells. It is an appealing idea to

harvest these two cell populations and scrutinize their similarities and differences with regard to their transcription profile and epigenetic landscape. Here, I will mainly discuss the roles DNA methylation and histone modification play during endocrine differentiation, as well as immediate experiments that can be executed based on the methodologies we have developed so far. More discussions about epigenetic regulation and potential approaches to tackle its engagement and function in pancreatic endocrine differentiation, along with sorting $Ngn3^{+}Myt1^{+}$ cells and $Ngn3^{+}Myt1^{-}$ cells, will be continued in the future directions section in Chapter VI.

DNA methylation of transcription factor genes other than Myt1

There are three DNA methyltransferase genes in mammals, *Dnmt1*, *Dnmt3a*, and *Dnmt3b* (Moore et al., 2013). Dnmt1 maintains DNA methylation propagation during cell division by methylating the newly synthesized DNA strand based on the DNA methylation pattern of the mother strand. In contrast, Dnmt3a and Dnmt3b are *de novo* DNA methyltransferases that deposit new DNA methylation markers but not simply copying existing DNA methylation markers.

DNA methylation-mediated gene expression control is not an uncommon mechanism in regulating endocrine transcription factor genes. For example, DNA methylation of the upstream regulatory (UR) elements UR1, UR2 and UR3 of *Arx* mediates *Arx* repression and is required to maintain β-cell identity (Dhawan et al., 2011). *Dnmt1* deletion in β cells leads to reduced DNA methylation of these *Arx* upstream regulatory elements and the conversion of β cells to α cells (Dhawan et al., 2011).

In Fig 3.10, *Dnmt3b* overexpression resulted in increased β-to-α cell ratio, but how that was achieved remained unclear. Specifically, what gene regions besides Myt1 region 2 were methylated to cause β-cell lineage expansion at the expense of α-cell lineage? In the $Rosa26^{M2\text{-}rtTA/+};ColA1^{tetOP\text{-}Dnmt3b/+}$ mice, induced *Dnmt3b* overexpression

does not cause overwhelming DNA methylation. Although the researchers lacked whole genome sequencing evidence, they found that the amount of 5'methylcytosine is not increased in the *Dnmt3b* overexpression mice. Rather, DNA methylation is increased at selective loci (Linhart et al., 2007). Thus, instead of being limited to Myt1 region 2, we need to assess more genomic loci or even global methylome to unveil the underlying mechanism. Next, I will discuss some candidate loci that are worth investigation.

DNA elements of some key endocrine transcription factors are reported to be binding sites of regulatory proteins and are predicted to be CpG islands (CpG islands are predicted by MethPrimer), but no explicit DNA methylation analysis has been done for these DNA elements. For instance, the AEG element of the *Pax4* promoter (Brink et al., 2001; Brink and Gruss, 2003), which is a binding site for transcription activators Ngn3, Pdx1, and NeuroD, is located close to a predicted CpG island. Whether DNA methylation of AEG regulates *Pax4* expression could be tested with bisulfite sequencing. An Hnf3β/Ngn3 binding site of the *Nkx2.2* promoter (Watada et al., 2003), the B1 and the PBE site of the *Nkx6.1* gene (Iype et al., 2004; Watada et al., 2000) are all such examples and good candidate loci.

I have encountered many problems with bisulfite sequencing, primarily because of the difficulties of obtaining high quality PCR fragments. After bisulfite conversion, the complexity of the genomic DNA is reduced and thus non-specific primer binding occurs more frequently. Besides, DNA secondary structures frequently form after bisulfite conversion, especially when there are long AT-rich stretches. DNA instability after bisulfite conversion is also a potential reason prohibiting successful PCR. I have thus far optimized the bisulfite PCR protocol to balance the trade-off between efficiency and specificity. I have designed nested primers (Table 3) and used a touchdown and nested PCR strategy to amplify regions of interest. In the future, we should obtain more colonies

for Myt1 region 2 in order to reach a more robust conclusion. Meanwhile, I am also testing many other regions, including those regulating *Pax4*, *Arx*, *Nkx2.2*, and *Nkx6.1 etc*.

Histone modification and pancreatic endocrine lineage allocation

Histone modification has been shown to be important in specifying endocrine lineages during pancreas development in many studies. For instance, human α cells exhibited distinct histone modification patterns from exocrine cells and β cells, with many more genes at a poised status marked by both the activating H3K4me3 marker and the repressing H3K27me3 marker (Bramswig et al., 2013). β-cell-specific deletion of *Ezh2*, the histone methyltransferase subunit of the Polycomb Repressive Complex 2 (PRC2), derepresses the *Ink4a/Arf* locus and reduces β-cell proliferation (Chen et al., 2009). Class IIa histone deacetylase (Hdac) Hdac4, 5, and 9 inhibit the β/δ cell lineages. Loss of *Hdac4* or *Hdac5* results in more δ cells, while loss of *Hdac5* or *Hdac9* increases the β-cell pool (Lenoir et al., 2011).

Our *in vitro* culture assay showed that Adox increased β-to-α cell ratio. A similar result is reported by Bramswig et al., in which they described murine α-to-β cell conversion upon Adox treatment (Bramswig et al., 2013). However, it should be noted that Adox targets the S-adenosyl-homocysteine (SAH) hydrolase and decreases the production of the methyl-group donor S-adenosyl-methionine (SAM) (Vaes et al., 2010). Thus, Adox inhibits global methylation events, including DNA methylation and histone methylation. Adox-induced β-to-α cell ratio increase suggests the involvement of epigenetic regulation in pancreatic endocrine lineage allocation, but it is not clear which epigenetic modification is accounted for the altered endocrine lineage allocation. Thus, other more specific inhibitors, such as 5-Aza-2'-deoxycytidine (AZA, an DNA methyltransferase inhibitor) (Christman, 2002), BIX 01294 or UNC 0638 (inhibitors of G9a, a histone methyltransferase for depositing H3K9me2 marker), 3-deazaneplanocin

(DZNep) or EPZ-6438 (inhibitors of the H3K27me3 methyltransferase Ezh2) among many others (Helin and Dhanak, 2013; Kubicek et al., 2007; Vedadi et al., 2011) could be tested in the future. The *in vitro* pancreas culture experiment was hindered mostly by technical issues. Due to their small size, cultured pancreatic buds were frequently lost during section preparation. We obtained a very small sample size and the robustness of these results will be characterized with further experimentation. Complementary to immunostaining on sections, we could perform qRT-PCR with these samples to examine gene expression alteration and endocrine cell allocation change. I believe that the $Ngn3^{+}$ progenitor subsets differ not only in their transcription profile and DNA methylation as discussed above, but also in their histone modification landscape, which could be investigated in the future by chromosome immunoprecipitation (ChIP) assay, small molecule interference, and genetic mouse models.

Diabetes is a major healthcare issue threatening the world. Type I diabetes is caused by the loss of β cells, presumably because of the immune attack on β cells due to defects in the immune system (Knip et al., 2005). Type II diabetes, on the other hand, is characterized by insulin resistance in the periphery organs or tissues, although β-cell loss is also observed as the disease deteriorates (Ashcroft and Rorsman, 2012). Islet or β-cell transplantation is a promising therapy for diabetes and several successful cases of islet transplantation in type I diabetic patients have been reported (Shapiro et al., 2000; Shapiro et al., 2006). The remaining problems are the lack of sufficient islet donor sources and the immune rejections that impair successful engraftment (Rother and Harlan, 2004). Generating β cells from hESCs or iPSCs or even terminally differentiated cells is a promising alternative to cadaveric islets. Many research teams have claimed the generation of β cells but these β cells have intrinsic drawbacks to be used as a therapy. For example, they do not show robust insulin secretion in response to glucose

stimulation and lack β-cell maturation markers (Basford et al., 2012; Bruin et al., 2013; D'Amour et al., 2006; Efrat, 2013; Kroon et al., 2008; Pagliuca et al., 2014). Therefore, it is necessary to have a better understanding of the pancreas development process, especially β-cell generation, in order to come up with a refined *in vitro* β-cell differentiation protocol. Our bipartite Cre cell lineage tracing experiment as well as the ongoing epigenetic analysis will contribute to the understanding of transcription factor interactions and the endocrine lineage commitment mechanism, which will provide educated suggestions to the design of β-cell regeneration protocol.

3.5 Materials and Methods

Mouse strains and care

The *Ngn3*^{nCre} knock-in mouse was generated by RMCE (recombination-mediated cassette exchange). A targeting vector was constructed through conventional molecular cloning and it contained a lox71 site, FRT flanked hygromycin selection cassette, 3.5kb Ngn3 5' region, nCre-T2A-Ngn3, polyA signal and a lox2272 site. The loxed cassette acceptor (LCA) allele is *Ngn3*$^{HA.LCA}$ (unpublished), which contained a lox66 site inserted 3.5kb upstream of the transcription initiation site of Ngn3 and a lox2272 site 1kb downstream of the Ngn3 polyA signal. Targeting vector was prepared through E.coli amplification followed by QIAGEN Maxiprep (#12162) and was electroporated into ES cells possessing the *Ngn3*$^{HA.LCA}$ allele. ES cell clones were screened through PCR and Southern blot for correct recombination. Selected ES cell clones were injected into mouse blastocoel and implanted into the uterus of pseudo-pregnant female mice. Pups were genotyped and germ line transmissible strains were kept. The FRT-site flanked hygromycin expression cassette was used for positive selection of targeted ES cells and was later removed by crossing with a transgenic FLPe mice.

The *Myt1*^{cCre} allele was generated through BAC recombineering. A targeting vector containing ~250bp 5' arm, cCre coding sequence, FRT-flanked Neo/Kan selection cassette, and ~350bp 3' arm was constructed. The 5' and 3' homologous arm were designed around exon2 of Myt1 such that cCre was inserted there. The BAC that comprises Myt1 gene was electroporated into EL250 cells. Subsequently, the targeting vector was electroporated into the EL250 cells and allowed for recombineering. Colonies were screened with PCR for correct recombineering. Neo/Kan selection cassette was removed with L-arabinose induction. BAC was proliferated, purified and used for pronuclear injection. Pups were genotyped by tail biopsy and correct ones were retained.

Rosa26$^{tdTomato/+}$ (*Ai9*) (Madisen et al., 2010) and *Ngn3*$^{EGFP/+}$ (Lee et al., 2002) mice were purchased from the Jackson Laboratory. Genotyping primers used for all mouse lines in this study were listed in Table 2.

Mice were generated in collaboration with the ES/Transgenic Shared Resource at Vanderbilt University. All mice were housed and cared in the Vanderbilt Division of Animal Care and in compliance to IACUC regulations.

Immunostaining

Immunostaining methods followed protocols described before (Wang et al., 2008; Wang et al., 2010; Xu et al., 2007). Either paraffin or cryo-preserved sections were used. In general, dissected pancreata were fixed in 4% paraformaldehyde overnight. They were then embedded in OCT cryo-preservant (Sakura 4583) and stored in -80°C freezer. Pancreas tissue was cut at 10μm. Sections were washed with 1X PBS three times, 5min each. They were then permeablized with permeablization solution (0.2% Trition-X 100 in 1XPBS) for 10min. After washing again, sections were blocked with blocking solution (0.1% Tween-20 in 1X PBS, 5% donkey serum, 1% BSA). Then primary antibodies were

added and incubated at 4°C overnight. The second day, slides were washed. Secondary antibodies were added and incubated at room temperature for 30-60min. After washing, slides were ready for examination and imaging. For long time preservation, slides can be sealed in mount medium. Staining protocol was similar for paraffin sections, except that tissues had to go through dehydration and rehydration processes. Depending on the antibody used, it was sometimes necessary to perform antigen retrieval between the permeablization and blocking steps. Slides were boiled in 10mM Tris-HCl, 0.5mM EGTA pH 9.0 (cytoplasmic antigen) or 10mM Sodium Citrate pH 6.0 (nuclear antigen) antigen retrieval buffer in water bath for 10min.

Primary antibodies used: guinea pig anti-Ngn3, goat anti-Ngn3, rabbit anti-Myt1, guinea pig anti-Pdx1, mouse anti-Nkx2.2 (74.5A5-c, DSHB), guinea pig anti-insulin (DAKO), rabbit anti-glucagon (DAKO), rabbit anti-somatostain (Invitrogen), guinea pig anti-pancreatic polypeptide (Invitrogen), *etc*. Secondary antibodies used: FITC-conjugated donkey anti-guinea pig, Cy3-conjugated donkey anti-rabbit, *etc*. (Jackson Immunoresearch).

Confocal microscopy

All confocal images were taken with Leica TCS-SP5 scanning confocal microscopy or Olympus FV-1000 confocal microscopy.

Quantification of β cell mass

P0 pancreas was prepared into paraffin sections and stained with guinea pig-anti-insulin, followed by DAB (3, 3'-diaminobenzidine)-Peroxidase signal amplification (Vector, SK-4100). Sections were counter stained with eosin and then scanned by Leica Aperio ScanScope. Image processing and quantification was done with ImageScope Viewer from Leica. Total insulin$^+$ area from one animal was normalized to its eosin$^+$

pancreas area. Blood vessels and non-pancreatic tissue were excluded manually. At least three animals for each genotype were quantified.

Blood glucose measurement and IPGTT test

Random blood glucose level was measured with blood glucose monitor (Nova Max). Blood samples were obtained with tail bleeding. For IPGTT test, 6 week old mice were fasted overnight. In the next morning, glucose was given through IP injection at 2mg glucose (concentration: 10% (w/v)) per gram body weight. Blood glucose level was monitored before giving IP glucose administration (0min), and 15min, 30min, 60min, 90min, 120min after IP glucose injection with blood glucose monitor (Nova Max). Blood samples were obtained with tail bleeding.

Bisulfite sequencing

Ngn3$^{EGFP/+}$ male and female mice were crossed to obtain E10.5 and E14.5/E15.5 embryos. Both *Ngn3*$^{EGFP/+}$ and *Ngn3*$^{EGFP/EGFP}$ embryos were collected and pancreata dissected. After dissociation, pancreatic cells were subject to fluorescence-activated cell sorting (FACS) to collect EGFP^{+} cells. After FACS, cells were subject to bisulfite conversion with EZ DNA Methylation-Direct Kit (Zymo Research, D5020). Bisulfite conversion was done according to the manufacturer's protocol. Then, DNA region of interest was amplified with nested PCR (primer sequence: Table 3). PCR product from the first round of PCR with outer primer pair was used as template for the second round PCR with inner primer pair. A touchdown PCR method was used to ensure specificity. PCR program: 1st PCR: 94°C$^{5:00}$ [94°C$^{0:30}$;60-54/0.5°C$^{0:30}$;72°C$^{1:00}$]$_{12}$ [94°C$^{0:30}$;54°C$^{0:30}$;72°C$^{1:00}$]$_{25}$ 72°C$^{5:00}$12°C$^{5:00}$; 2nd PCR: 94°C$^{5:00}$ [94°C$^{0:30}$;54°C$^{0:30}$;72°C$^{1:00}$]$_{35}$ 72°C$^{5:00}$12°C$^{5:00}$. The PCR fragment was ligated into the pBluescript KS II SmaI site using Takara ligation kit. Ligation product was transformed

into DH5α cells followed by blue/white screening. White colonies were picked and checked with PCR (with T3 and T7 primers). Colonies with inserts were propagated and plasmids were prepared for sequencing. Sequence alignment and visualization was done with CLC Sequence Viewer 7 and R programming, respectively.

In vitro pancreas bud culture assay

E11.5 pancreas buds were dissected and placed on the Matrigel or porous membrane filters. Pancreas buds were cultured in RPMI 1640 medium (Gibco, 11875093) supplemented with 10% FBS (Gibco, 26140079) and 100 unit/ml penicillin-streptomycin (Gibco, 15140-122). Adox was added into the culture medium at a final concentration of 10μM. Culture medium was changed daily. After 3 days of culturing, pancreas buds were retrieved and prepared into sections for staining. Alternatively, whole pancreas buds were stained in test tubes without sectioning.

Quantitative real time-PCR (qRT-PCR)

qRT-PCR experiments was done to quantify the mRNA level of cells of interest. First, RNA was extracted from dispersed cultured pancreatic buds with TRIzol (Life Technologies, Cat#: 15596026). Then, cDNA was prepared with the High Capacity cDNA Reverse Transcription Kit (Cat #: 4368814) or Bio-Rad iScript cDNA Synthesis Kit (Cat #: 170-8890). Next, real-time PCR was done with the Bio-Rad SsoAdvanced Universal SYBR Green Supermix (Bio-Rad, Cat #: 172-5271) following manufacturer's instructions. PCR reaction was performed with Bio-Rad CFX96 Real Time System. qRT-PCR primer sequences are listed in Table 4. qRT-PCR experiments was also done for the FACS-sorted cells (see Fig 6.3).

Statistical analysis

Data in the figures were all represented as mean ± s.e.m. Student's t-test was used for statistical comparison. P-values less than 0.05 were considered significant.

Table 2. Genotyping primers

Allele		Primer name	Primer sequence (5'-3')
Ai9		oIMR9020	AAGGGAGCTGCAGTGGAGTA
		oIMR9021	CCGAAAATCTGTGGGAAGTC
		Ai9-1	GCGGCCACTACCTGGTGGAGT
		Ai9-2	CCACGCCACGTTGCCTGACAA
$Ngn3^{nCre}$		ngn3ncreO1	GACTTGAGCAGGGACCGTCTCT
		ngn3ncreO2	CTCAGAGAGGGAAACGGCTTGT
	or	ngn3ncre5a	CCAAAGGGTGGATGAGGGGCG
		ngn3ncre5s	ATGTGGCGTCCACGGGGAGT
$Myt1^{cCre}$		myt1cCreOs	GGCAAACTTCTGACCCAGAGGT
		myt1cCreOas	GTTGGTCCATCCGCCAGCCTGCA
	or	myt1cCreOs	GGCAAACTTCTGACCCAGAGGT
		myt1cCre_geno	ACGTTGGTCCATCCGCCAGC
$Nkx2.2^{cCre}$		nkxcCreout2	CTGGAAGGGCGTGCTCCAGGCT
		nkxcz-300	GCTCGCTCCAACCTGGGCCATT
	or	NkxcCreOS	GTCCAGGCCCAGCAGTGGACTT
		myt1cCre_geno	ACGTTGGTCCATCCGCCAGC
$Ngn3^{EGFP}$		Ngn3_A_KK	ATACTCTGGTCCCCCGTG
		Ngn3_B_KK	TGTTTGCTGAGTGCCAACTC
		Ngn3_C_KK	GAACTTGTGGCCGTTTACGT

Table 3. PCR primers for bisulfite sequencing

Locus	Primer sequence (5’ to 3’)	
Arx UR2	Outer primers:	ATAGAGAGTTTGAAAGTTTTTTGGG; TCTCATCTAACTTTAACCCAATATT
	Inner primers:	TGGATTTATTTTTATTTTATTTTTTGC; AACTCTTAAATTTCCCTCAAACTTC
Pax4 -2.1K	Outer primers:	AGGAGAATGATTTTGGATTTGTGG; AACCATTCCAACTCCTTCCTCACCT
	Inner primers:	AAAGGGTAGATGAGGATTAGGATTT; ATTAAAATTCCTCCCTACTTCCTTC
Pax4 AEG	Outer primers:	GGAGTTTTTTAAAGGTAGGAGTTAA; CTTCACTATCTAACTCTCCTAACTAC
	Inner primers:	TTATATTTAGGTGAGGAAGGAGTTG; CTATACCCAACCCCCAAACTACTA
Myt1 region 2	Outer primers:	GATTTATTTTTTTGATGTTATATTGTAAATTTATTT; AACAACTACCTCCCTCACCACAATCACATATA
	Inner primers:	AGAGAGTAAATAGATATATTGGAGTTTTAAGGG; TAAATTTCAATTAAATATCTTCTCCCCTC

Table 4. qRT-PCR Primers used for gene transcription quantification

Gene target	Primer name	Primer sequence (5'-3')
Ngn3	qmNgn3F	GAGGCTCAGCTATCCACTGC
	qmNgn3R	TGTGTCTCTGGGGACACTTG
Myt1	qmMyt1F	TGTGCTGGAGAATGATGAGG
	qmMyt1R	TCTCCATGGACGAGATCTGA
Nkx2.2	qmNkx2.2-F1	CTTTCTACGACAGCAGCGAC
	qmNkx2.2-R1	CCCTGGGTCTCCTTGTCATT
Pdx1	mPdx1-F	GAAATCCACCAAAGCTCACG
	mPdx1-R	CGGGTTCCGCTGTGTAAG
Pax4	mPax4-F	TCCCAGGCCTATCTCCAAC
	mPax4-R	TATGAGGAGGAAGCCACAGG
Arx	mArx-F	TTCCAGAAGACGCACTACCC
	mArx-R	TCTGTCAGGTCCAGCCTCAT
Nkx6.1	mNkx6.1-F	ACTTGGCAGGACCAGAGAG
	mNkx6.1-R	GCGTGCTTCTTTCTCCACTT
Sox9	Sox9 RT2 3646F	TTGTGACACGGGACAACACA
	Sox9 RT2 3762R	CCAGCCACAGCAGTGAGTAA
MafA	qmMafAF	ACCACGTGCGCTTGGAGGAG
	qmMafAR	ATGACCTCCTCCTTGCTGAA
MafB	qmMafB-F1	TGGGATTATCTCTTCGCCCC
	qmMafB-R1	TCGTGGGTGTGTGTATGTCA
Pax6	qmPax6-F1	TCACCATGGCAAACAACCTG
	qmPax6-R1	CCATGGGCTGACTGTTCATG
Ins1	ins1a	GGGACCACAAAGATGCTGTT
	ins1s	CAGCAAGCAGGTCATTGTTT
Ins2	qmIns2-F1	TGAAGTGGAGGACCCACAAG
	qmIns2-R1	GTAGTGGTGGGTCTAGTTGC
Gcg	glc2a	TGGTGGCAAGATTATCCAGA
	glc2s	GCGAGACTTCCCAGAAGAAG
PPY	pp1a	AGAGAGGCTGCAAGTCCATT
	pp1s	GGAGGAGAACACAGGTGGAC
SST	ss1a	ACTTGGCCAGTTCCTGTTTC
	ss1s	CCCAGACTCCGTCAGTTTCT
Ptf1a/p48	p482a	TAATTCTTCAGGCACCATGC
	p482s	TGCTCCTGCTACTACTGCCA
GAPDH	mGAPDH-F	AACTTTGGCATTGTGGAAGG
	mGAPDH-R	GGATGCAGGGATGATGTTCT
PPIA	qmPPIA-F1	AAGCATACAGGTCCTGGCATC
	qmPPIA-R1	ATGCCTTCTTTCACCTTCCCAAA
Glut2	glut21a	AAGAACACGTAAGGCCCAAG
	glut21s	AGCAACTGGGTCTGCAATTT

CHAPTER IV

DESIGN AND CHARACTERIZATION OF AN INDUCIBLE BIPARTITE CRE

4.1 Abstract

Our bipartite Cre-based lineage tracing allows us to examine the lineages of double-marked progenitor cells ($Ngn3^{+}Myt1^{+}$). Yet it does not allow us to examine the stage-specific fate of progenitor cells. It should be noted that the pancreatic endocrine cells are produced in a sequential yet partially overlapping manner. Glucagon-secreting α cells emerge at ~E9.5 and are the first endocrine cell type generated during pancreas development. Insulin-secreting β cells appear later, reach the peak of production in the secondary transition and decline gradually after perinatal period. δ and PP cells are produced in late gestational stages. In Chapter III, we found that $Ngn3^{+}Myt1^{+}$ progenitors favor β-cell fate over α-cell fate by utilizing a novel bipartite Cre lineage tracing method. However, this cell lineage tracing result represents the overall differentiation potential of $Ngn3^{+}Myt1^{+}$ progenitors across embryonic stages; it does not reflect the differentiation potential of $Ngn3^{+}Myt1^{+}$ progenitors at specific embryonic stages. It is possible that $Ngn3^{+}Myt1^{+}$ progenitors have different endocrine cell fate preferences as they pass through competence windows. As an initial step to address this hypothesis, I created a tamoxifen-inducible bipartite Cre system and characterized its recombination activity here. I found one inducible bipartite Cre construct that showed negligible background recombinase activity but substantial recombinase activity after 4-hydroxyltamoxifen (4OH-TM) induction in a dosage-dependent manner. This construct could be used to generate transgenic mouse and perform stage-specific cell lineage tracing experiments in the future.

4.2 Introduction

The pancreas is composed of endocrine islets of Langerhans and exocrine cells (duct and acinar cells) (Edlund, 2002; Gittes, 2009; Pan and Wright, 2011). The islets of Langerhans are mainly composed of insulin-secreting β cells and glucagon-secreting α cells, with a small percentage of δ, PP and ε cells, all of which are critical hormones to maintain blood glucose homeostasis (Edlund, 2002; Gittes, 2009; Pan and Wright, 2011). The bHLH transcription factor *Neurogenin3* (*Ngn3*) specifies pancreatic endocrine cells (Gradwohl et al., 2000; Gu et al., 2002). Notably, endocrine cells are generated in ordered temporal windows, during which $Ngn3^+$ progenitors are competent to generate each endocrine subtype (Johansson et al., 2007). α cells are detected as early as E9.5; β cells are generated in concurrence with the secondary transition while δ and PP cells do not emerge until E15.5.

The temporally controlled differentiation of $Ngn3^+$ progenitors is a manifestation of regulated and hierarchical expression of various transcription factors. For example, the expression of *Pax4*, *Pdx1*, *Nkx2.2* and *Nkx6.1* in $Ngn3^+$ progenitors facilitates β-cell production, while the expression of *Arx*, *Brn4* and *MafB* marks α cells (Collombat et al., 2006; Murtaugh, 2007; Pan and Wright, 2011). As discussed in Chapter III, it is possible that combinations of different transcription factors specify endocrine subtypes. However, most of the published data are obtained through loss-of-function and ectopic expression experiments. No definitive cell lineage tracing experiments have been done to test this possibility. In Chapter III, I described an innovative bipartite Cre cell lineage tracing system, which allowed me to investigate the differentiation potential of the $Ngn3^+Myt1^+$ progenitors, and found that the $Ngn3^+Myt1^+$ progenitors favored β cells over α cells.

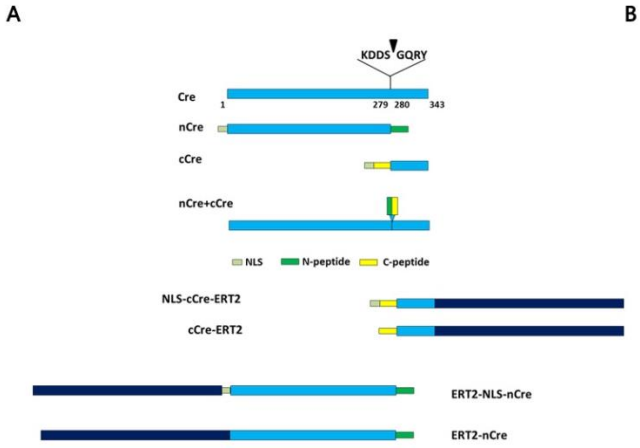
Because pancreatic endocrine cells are differentiated in sequential competence windows, I wonder whether the differentiation potential of $Ngn3^+Myt1^+$ progenitors

changes over developmental stages. The cell lineage tracing result from Chapter III only represents the overall differentiation potential of $Ngn3^{+}Myt1^{+}$ progenitors, but does not reveal the dynamic properties of progenitors over time. To unveil the cell lineage preference of $Ngn3^{+}Myt1^{+}$ progenitors at different embryonic stages, it is preferable to perform a "pulse-chase" experiment and only label the $Ngn3^{+}Myt1^{+}$ progenitors for a short time period and then track their descendent cells. To this end, I designed bipartite Cre and estrogen receptor (ERT2) fusion proteins and tested their recombinase activity *in vitro*. I compared different fusion proteins and found one nCreERT2/cCreERT2 pair that possessed optimal induced recombinase activity and minimal leaky activity. This inducible bipartite CreERT2 could be used to generate transgenic mouse models in the future and investigate the dynamics of $Ngn3^{+}Myt1^{+}$ progenitors' differentiation potential.

4.3 Results

Bipartite CreERT2 construction and comparison

I decided to create fusion proteins between bipartite Cre halves and estrogen receptor (ER) to achieve temporal regulation. CreER has been widely used and mutagenesis studies have created ER variants with improved characteristics, one of which is ERT2 (Feil et al., 1996; Feil et al., 1997; Indra et al., 1999). ERT2 has a low affinity for natural ER ligands and is highly sensitive to 4-hydroxytamoxifen (4OH-TM, the metabolized and active form of tamoxifen). This minimizes the confounding effects from endogenous estrogen and the toxicity of high tamoxifen dosage, making it a perfect choice for temporal gene regulation. Although the doxycycline-regulated tetON/OFF system can also regulate gene expression temporally (Lewandoski, 2001), it is less convenient than the ER system when it comes to mouse crossing. Therefore, I decided to create a bipartite CreERT2 cell lineage tracing system.



B

		EcoRI	XhoI	NotI
YW737	mCherry-iRES	nls-nCre	iRES	nls-cCre
YW819	mCherry-iRES	nls-nCre	iRES	cCre
YW926	mCherry-iRES	nls-nCre	iRES	nls-cCre-ERT2
YW941	mCherry-iRES	ERT2-nls-nCre	iRES	nls-cCre
YW942	mCherry-iRES	ERT2-nls-nCre	iRES	nls-cCre-ERT2
YW951	mCherry-iRES	HA-ERT2-nls-nCre	iRES	nls-cCre-ERT2-Myc
YW927	mCherry-iRES	nls-nCre	iRES	cCre-ERT2
YW936	mCherry-iRES	ERT2- -nCre	iRES	cCre
YW937	mCherry-iRES	ERT2- -nCre	iRES	cCre-ERT2
YW943	mCherry-iRES	CreERT2		

Fig 4.1 Design of the inducible bipartite CreERT2. (A) Graph representation of the inducible bipartite CreERT2 design. ERT2 module is fused to the N-terminus of nCre and C-terminus of cCre. (B) Different inducible bipartite CreERT2 constructs. YW737 and YW819 are non-inducible bipartite Cre constructs, based on which the inducible bipartite CreERT2 constructs were made. YW926, YW941, YW942 and YW951 are constructs with NLS while YW927, YW936, YW937 are ones without NLS sequences. YW951 contains an HA and an Myc tag and is derived from YW942. YW943 has a full length CreERT2. mCherry is used a transfection indicator. iRES sequences are used for multi-cistronic expression.

I strived to find a construct that had: (1) limited background activity prior to 4OH-TM induction; (2) high activity upon induction. All constructs were based on the previously used non-inducible bipartite Cre system in Chapter III. Basically, ERT2 was fused with nCre or cCre and then replaced the non-ERT2 counterparts in the original pmCherry-C1 expression vector through conventional restriction digestion and ligation (Fig 4.1). These constructs were then tested in a cell transfection assay followed by flow cytometry analysis.

There were several considerations when I constructed bipartite CreERT2 fusion proteins: (1) Do not mask the leucine zipper sequences. The leucine zipper sequences facilitate bipartite Cre reconstitution and shall not be blocked from accessing each other. Thus, I fused ERT2 with Cre halves in the opposite ends of the leucine zipper sequences. In other words, I fused ERT2 to the N terminus of nCre and the C terminus of cCre (Fig 4.1 a). (2) Inclusion of nuclear localization signal (NLS) peptide or not. The presence of NLS facilitates the nuclear localization of Cre halves but it also poses the potential problem of high background Cre activity prior to induction. Thus, I made and tested constructs with or without NLS sequences (Fig 4.1 a). (3) Copy number of ERT2. It is not clear whether and how the additional ERT2 will interfere with the expression and reconstitution of Cre halves, so I tested different combinations in which only nCre or only cCre or both were fused with ERT2 (Fig 4.1 b). (4) Since almost all inducible CreERT2 constructs have the ERT2 domain fused to the C terminus of Cre protein (Magnuson and Osipovich, 2013), I was concerned whether fusing the ERT2 domain to the N terminus of nCre protein might interrupt the folding and activity of ERT2 and nCre. Thus, I added a short linker sequence (Glycine-Glycine-Serine) to join ERT2 and nCre.

The construct YW942 stood out in the flow cytometry result. This construct had ERT2 fused to both nCre and cCre. It also had NLS sequences in both Cre halves. This construct showed negligible background Cre activity (7.0±1.3%, n=3) but substantial

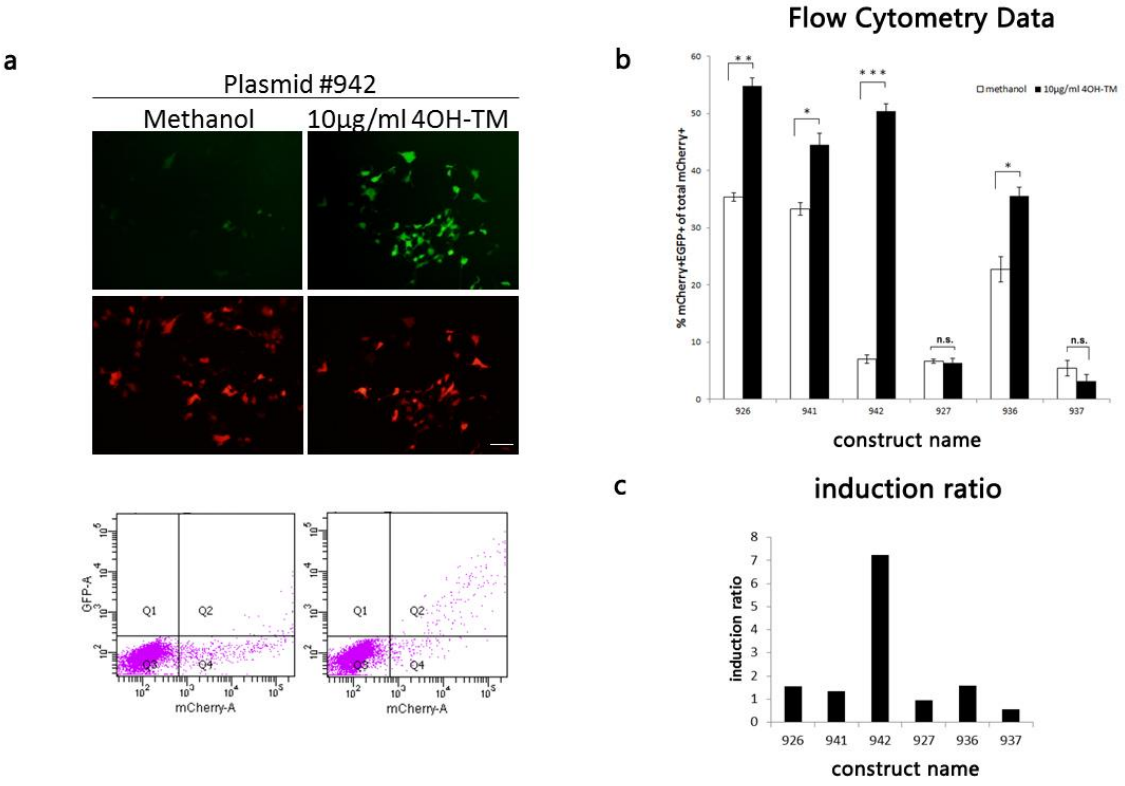


Fig 4.2 Comparison of different inducible bipartite CreERT2 constructs. (a) Representative images of transfected cells (upper panel) and flow cytometry analysis plots (lower panel). Shown here is construct YW942 treated with methanol or 10µg/ml 4OH-TM. Image: red channel: mCherry; green channel: GFP. Scale bar: 20µm. Flow cytometry plot: y-axis: GFP; x-axis: mCherry. (b) Quantification of the reconstituted recombinase activity, with or without induction. Data are represented as mean±s.e.m. p-value: ***<0.001<**<0.01<*<0.05<n.s. (c) Fold increase of recombinase activity upon 4OH-TM induction. YW942 has a remarkably higher induction ratio, suggesting its robust induced recombinase activity as well as low background activity.

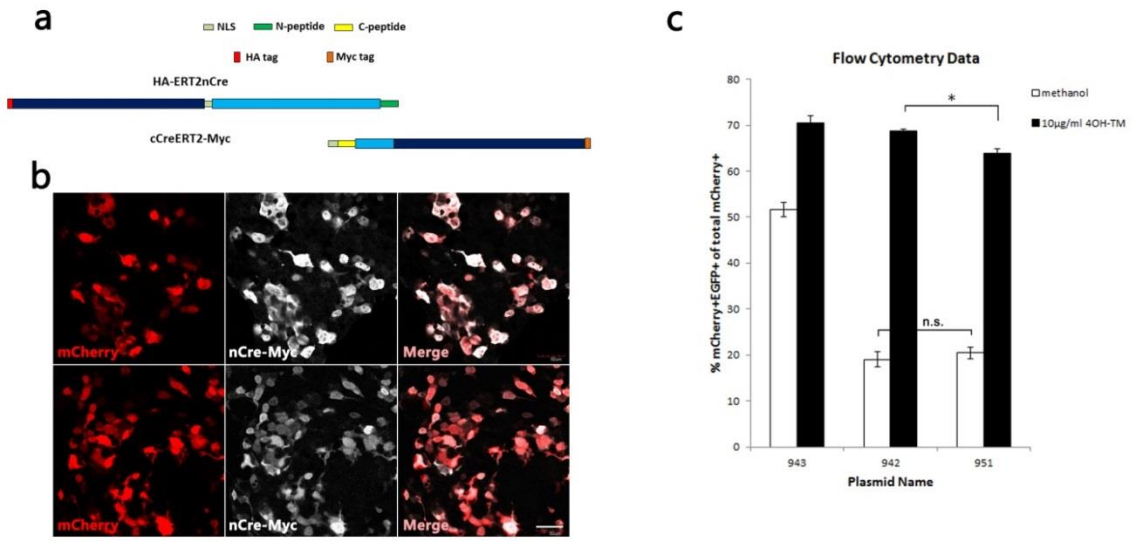


Fig 4.3 Immunotagged inducible bipartite CreERT2. (a) Graph representation of the immunotagged bipartite CreERT2 to show the position of HA/Myc tags. (b) HA and Myc tags are readily detectable by standard IHC. Scale bar: 50μm. (c) Flow cytometry analysis demonstrates that adding immunotags does not impair the reconstituted recombinase activity. Data are represented as mean±s.e.m. p-value: ***<0.001<**<0.01<*<0.05<n.s.

recombination activity upon 10μg/ml 4OH-TM treatment (50.4±2.1%, n=3) (Fig 4.2 a, b). An induction ratio of 7.2 was achieved, the highest of all constructs (Fig 4.2 c). Compared to constructs with ERT2 fused with nCre or cCre only, the background of YW942 was much lower. A possible reason could be that the nCre and cCre protein bind to each other through the interaction of leucine zipper peptides in the cytoplasm and one copy of ERT2 is insufficient to retain the reconstituted Cre in the cytoplasm. Omitting NLS impaired Cre activity and this is possibly because the nCre or cCre protein cannot be efficiently translocated into the nucleus. However, I observed one NLS-free construct that displayed appreciable recombinase activity even in the absence of 4OH-TM induction. This could probably happen because saturated cytoplasmic protein leaks into the nucleus due to the overexpression nature of cell transfection assay.

Addition of immunotags simplifies bipartite CreERT2 detection

For the convenience of future analysis, I added immunotags to both bipartite CreERT2 halves and derived a construct YW951 from the above YW942. Similar to the considerations when adding ERT2 domains, the immunotags were fused to the opposite ends of leucine zipper sequences to avoid disrupting Cre reconstitution. Specifically, an HA tag was fused to the N-terminus of ERT2-nCre, making HA-ERT2-nCre; a Myc tag was fused to the C-terminus of cCre-ERT2, making cCre-ERT2-Myc (Fig 4.3 a). Cell transfection assay and flow cytometry analysis showed that the tagged version and the untagged version had no significant differences in their background activity. Though the tagged version displayed lower response to 4OH-TM induction than the untagged version, it was sufficient to activate the reporter in most transfected cells (Fig 4.3 c). The HA and Myc tags were readily detected with standard IHC (Fig 4.3 b).

With the same transfection protocol, full length CreERT2 fusion protein showed high background activity almost as comparable when it was induced with 4OH-TM (Fig

4.3 c). In contrast, the bipartite CreERT2 protein showed lower background activity. High background activity of the full length CreERT2 can be explained by the overexpression nature of transfection assay and the strong CMV promoter I used to drive the expression of Cre recombinase. Because full length CreERT2 drivers have been successfully used in vast amount of *in vivo* studies, one could be certain that the *in vivo* expression level of CreERT2 is much lower than that of the *in vitro* assay. Thus, it is foreseeable that our bipartite CreERT2 would have trivial background activity in transgenic mouse models.

That being said, I quantified the bipartite CreERT2 recombinase activity with a series of DNA amounts for transfection and dissected the overexpression effect of cell transfection assay (Fig 4.4 a). At high DNA amount, recombination efficiency was actually lower, probably due to toxicity caused by Cre overexpression-induced cellular stress. When YW951 plasmid amount was reduced to 10ng per well of a 24-well plate, the recombination activity went higher, to 46.4%. Reducing DNA amount to 1ng decreased recombination activity by ~5%. I expected to see lower recombination activity when transfected DNA amount was further reduced. However, flow cytometry data became unreliable when a small quantity of DNA was used and the number of transfected cells was low. A strong linear correlation between the transfected DNA amount and recombinase activity was not observed in this case.

Bipartite CreERT2 shows 4OH-TM dosage-dependent response

I further characterized the immunotagged inducible bipartite CreERT2 construct YW951 by testing its responsiveness to a gradient of 4OH-TM. This inducible bipartite CreERT2 construct showed increasing activity up to 1μg/ml 4OH-TM. Beyond 1μg/ml, high 4OH-TM might have caused cell toxicity, resulting in reduced Cre recombination activity (Fig 4.4 b). This was supported by the observation of a large number of cells detaching from the Petri dish and undergoing cell death. It was also very obvious during

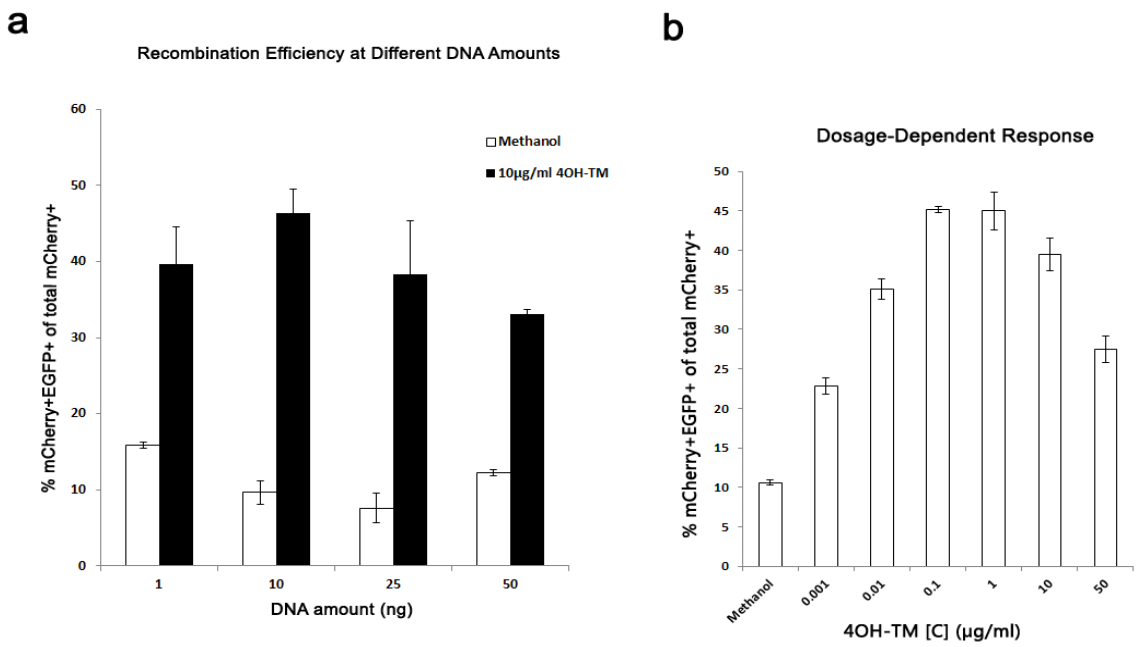


Fig 4.4 Detailed characterization of the inducible bipartite CreERT2. (a) Recombination efficiency of YW951 is tested at different DNA amount. (b) Recombination efficiency of YW951 is tested at 4OH-TM gradient concentrations. Data are represented as mean±s.e.m. p-value: ***<0.001<**<0.01<*<0.05<n.s.

flow cytometry analysis in which 4',6-diamidino-2-phenylindole (DAPI) signal formed a strong positive peak and living cell yield was reduced (data not shown).

Although the construct showed 4OH-TM dosage-dependent responsiveness and 1μg/ml 4OH-TM was found to be the optimal concentration in this *in vitro* assay, the TM dosage administered to mice should be titrated in the future if transgenic mice are generated. TM dosage varies from protocol to protocol and is determined based on the experimental design and purpose. Typically, 1-2mg per adult mouse is a good start and TM dosage is recommended to be normalized to mouse body weight.

4.4 Discussion

Pancreatic endocrine cells are generated in a temporally controlled manner, with α cells emerging first followed by β cells, δ and PP cells (Johansson et al., 2007). Though I found that $Ngn3^{+}Myt1^{+}$ progenitors give birth to more β cells than α cells, it is not clear whether the differentiation potential of $Ngn3^{+}Myt1^{+}$ progenitors is biased towards β cells at all embryonic stages. To address this question, I attempt to upgrade the bipartite Cre system into an inducible system by creating estrogen receptor fusion proteins.

I found a pair of nCre- and cCre-ERT2 fusion proteins that possessed optimal characteristics for induced DNA recombination. The nCre was fused with ERT2 at its N-terminal end and the cCre was fused with ERT2 at its C-terminal end. Both ERT2-nCre and cCre-ERT2 had NLS sequences and leucine zipper sequences for nuclear localization and facilitated reconstitution, respectively. Besides, HA and Myc immunotags were fused to the ERT2-nCre and cCre-ERT2 to assist the detection of nCre and cCre expression in the future. The recombination activity of this inducible bipartite CreERT2 was minimal without 4OH-TM induction but showed a dramatic

increase after 4OH-TM induction. Moreover, the responsiveness to 4OH-TM induction was dosage-dependent.

Since the inducible bipartite Cre construct was well characterized *in vitro*, it can be used to generate transgenic mouse models, for example, *Ngn3*$^{ERT2\text{-}nCre}$ and *Myt1*$^{cCre\text{-}ERT2}$ transgenic mice. These mice can be crossed and pregnant females be administered with tamoxifen at specific embryonic stages. In this way, cell lineage allocation of $Ngn3^{+}Myt^{+}$ progenitors at different embryonic stages can be examined and quantified with reporter labelling. Such data will give a clearer map of pancreatic endocrine specification.

The competence window model is not limited to pancreatic endocrine differentiation. Many other cell types are generated following a competence window model. In vertebrate retina, the retinal progenitors divide asymmetrically. At each stage, they can only produce one or a few cell types. For instance, ganglion cells and cone cells are the first cell types to emerge, followed by rod cells and then bipolar cells and Muller glia, albeit much overlap exists (Cepko et al., 1996; Livesey and Cepko, 2001). A similar inducible split Cre is also reported (Hirrlinger et al., 2009a). These inducible bipartite Cre toolkits can be used in many organs or tissues for cell lineage tracing, conditional gene activation and inactivation.

4.5 Materials and Methods

Plasmid construction

All Cre plasmids were constructed based on the pmCherry-C1 vector (Clontech) though the multiple cloning site (MCS) was partly modified to accommodate our inserts (Xu et al., 2007). The nCre fragment series were inserted after mCherry-iRES cassette between the EcoRI and XhoI sites, which were followed by the iRES-cCre fragment

series between the XhoI and NotI sites. Thus, mCherry, nCre and cCre portions were all connected by iRES sequences (Fig 4.1). A full length CreERT2 fragment was inserted after mCherry-iRES with restriction enzymes EcoRI and NotI. Overlap extension PCR was used to create fusion proteins of ERT2 and Cre halves (Heckman and Pease, 2007). A single nucleotide change was introduced in some primers to kill an extra XhoI restriction site without altering the amino acid encoded. As indicated in the main text, additional three amino acids (GGS) were introduced to fuse ERT2 and nCre as a linker; its coding sequence (GGTGGAAGC) was accordingly introduced into PCR primers. HA or Myc tag were added to the fusion proteins by adding coding sequences to the overlap extension PCR primers (HA tag: TACCCATACGATGTTCCAGATTACGCT; Myc tag: GAACAAAAACTCATCTCAGAAGAGGATCTG). Similar to the creation of ERT2-nCre fusion protein, the GGS linker was used to fuse HA and Myc tag to ERT2-nCre and cCre-ERT2, respectively. A Cre reporter plasmid YW421was previously constructed by inserting loxP-STOP-loxP-EGFP after a CMV promoter (Xu et al., 2007).

Three starting plasmids were used as template to construct ERT2 fusion proteins. YW737 contains nCre and cCre, both of which contain NLS sequences; YW819 contains nCre and cCre, but the cCre lacks NLS; and a pCreERT2 vector from which ERT2 was cloned. YW737 and YW819 were used in Chapter III to compare the relative recombination efficiency. The cloning strategy for YW737 and YW819 followed published methods (Xu et al., 2007) with only the codons optimized for mammalian expression.

Primers used: for cCre-ERT2: JL003 (5'gtgaatatctcgagatccgc3') + JL004 (5'agctctcatgtctccagcagatggctccagATCTCCGTCCTCCAGCAGGCGCACCATTGC3'); JL005(5'GCAATGGTGCGCCTGCTGGAGGACGGAGATctggagccatctgctggagacatgagagct3') + JL006 (5'GCGGCCGCtcagatcttcatcaagctgt3'); template: YW737 and a pCreERT2 vector. For cCre-ERT2 (without NLS), same primers were used but with YW819 as

template. For ERT2-nCre (without NLS): JL007 (5'aagaattctagatctccaccatgCTGGAGCCATCTGCTGGAGA3') + JL008 (5'CAGATTCTGGTGGACGGTGAGCAGGTTACTagctgtggcagggaaaccctctgcctcccc3'); JL009(5'ggggaggcagagggtttccctgccacagctAGTAACCTGCTCACCGTCCACCAGAATCTG3') + JL010 (5'atctcgagatattcactgtgcca3'); template: YW737 and pCreERT2. For ERT2-nCre (with NLS), same templates were used, but two primers were replaced: JL007 + JL012 (5'caggttactcaccttccgctttttctttgggcttccaccagctgtggcagggaaaccctctgcctcccc3'); JL011(5'ggggaggcagagggtttccctgccacagctggtggaagcccaaagaaaaagcggaaggtgagtaacctg3') + JL010. To make immunotagged construct YW951 using YW942 as template: for HA-ERT2-nCre portion (with NLS): JL015 (5' tagaattccaccATGTACCCATACGATGTTCCAGATTACGCTggtggaagcctggagccatctgctggagacatgagagct3') + JL010; for cCre-ERT2-Myc portion (with NLS): JL003 + JL016 (5'ttgcggccgccatcaCAGATCCTCTTCTGAGATGAGTTTTTGTTCgcttccaccagctgtggcaggaaaccct3')

All constructs were sequenced for confirmation and the end sequences are available for reference.

Cell culture and transfection

293T cells were cultured in DMEM medium (Gibco, 11995-065) supplemented with 10% FBS (Gibco, 26140079) and 100 unit/ml penicillin-streptomycin (Gibco, 15140-122). 50ng (25ng) Cre driver plasmid and 200ng (100ng) Cre reporter plasmid were used per well of a 12-well (24-well) cell culture for transfection into 293T cells. Cells were grown for 24 hours before 4OH-TM (Sigma) was added to a final concentration of 10μg/ml and cultured for another 24 hours. 4OH-TM was dissolved in methanol as a high concentration stock (10mg/ml) and methanol was always used as a vehicle control.

DNA amount and 4OH-TM concentration varied in titration experiments and were indicated in the text. After that, cells were collected for flow cytometry analysis.

Flow cytometry analysis

After transfection and induction with 4OH-TM, cells were trypsinized into single cells. DAPI was added as a viability control. These cells were analyzed by 5-laser BD LSR II or BD LSR Fortessa. Flow cytometry analysis was performed with the help from Vanderbilt Flow Cytometry Shared Resources.

Immunostaining and imaging

For immunostaining purpose, cells were cultured on cover slips. Cover slips were dipped in 90% ethanol and then sterilized over flame for a few seconds. Sterilized cover slips were put into 6-well culture plates to allow cells growing on. Transfected 293T cells were grown for 24 hours and treated with 10μg/ml 4OH-TM for another 24 hours. Cells were fixed with 4% PFA for 15min. Cells were permeablized with 0.2% Trixon X-100 and stained thereafter. Primary antibody: 1:100 rabbit anti-HA (Santa Cruz, HA-probe Antibody (Y-11): sc-805); 1:3000 rabbit anti-Myc (Millipore, gift from Dr. Wright's Lab, Vanderbilt). Secondary antibody: 1:500 Cy5-conjugated donkey anti-rabbit IgG (Jackson Immunoresearch). Confocal images were taken with Olympus FV-1000.

Statistical analysis

Data in the figures were all represented as mean ± s.e.m. Student's t-test was used for statistical comparison. P-values less than 0.05 were considered significant.

CHAPTER V

miRNA-MEDIATED INHIBITION OF NOTCH SIGNALING ENHANCES NGN3 EXPRESSION

5.1 Abstract

The pro-endocrine gene *Ngn3* marks the progenitors for all pancreatic endocrine cells. Loss of *Ngn3* or reduced *Ngn3* expression level leads to endocrine cell agenesis or reduction. It is, therefore, important to keep the regulation of *Ngn3* expression pattern and level under tight control. *Ngn3* expression regulation is subject to many factors and signaling pathways, including *Sox9*, *Hnf6*, *Pdx1*, and most importantly Notch signaling. The inhibition exerted on *Ngn3* expression by Notch signaling prevents the pancreatic epithelium from excessive endocrine cell differentiation, which is important for keeping the endocrine/exocrine balance. However, I observed side-by-side or even small clusters of $Ngn3^{+}$ progenitors, which violates the canonical lateral inhibition model of Notch signaling. Here I report preliminary data that certain miRNAs can inhibit Notching signaling pathway, potentially by targeting on its components Hes1 and Psen1, and thus derepress *Ngn3* expression. In addition, these miRNAs could possibly translocate among neighboring cells and enhance *Ngn3* expression in a non-cell-autonomous manner, unleashing lateral inhibition due to Notch signaling. The preliminary data support this miRNA-mediated inhibition of Notch signaling hypothesis and the possibility of miRNAs traveling through gap junctions. With further examination, I will be able to fill the gaps of the hypothesis and obtain a better understanding of *Ngn3* expression regulation.

5.2 Introduction

During embryogenesis, the pancreatic primordia in the foregut epithelium undergo primary transition and secondary transition to form a mature pancreas. The development of the pancreas and the allocation between endocrine and exocrine cells is a complicated process that involves the coordination of various transcription factors and signaling pathways. The transcription factor Ngn3 is best known as a pro-endocrine factor. *Ngn3*-null mice produce almost no endocrine cells (Gradwohl et al., 2000). Ectopic expression of *Ngn3* leads to precocious differentiation of pancreatic precursors into endocrine cells at the expense of exocrine lineages (Apelqvist et al., 1999; Schwitzgebel et al., 2000). Lineage tracing demonstrates that *Ngn3*-expressing cells give rise to endocrine islet cells (Gu et al., 2002).

However, the separation between endocrine and exocrine lineages does not merely depend on the presence or absence of *Ngn3* but also on Ngn3 protein level. Although Gu et al. has reported that all *Ngn3*-expressing endocrine progenitors give rise to endocrine islet cells (Gu et al., 2002), Schonhoff and colleagues find that *Ngn3*-expressing cells also give rise to a small portion of duct and acinar cells in the pancreas (Schonhoff et al., 2004). It appears that this above discrepancy may be due to the sensitivity differences of Cre reporters used in the two studies and reflects the distinct fate of progenitors that express different levels of *Ngn3* (Liu et al., 2013; Wang et al., 2010). In a previous paper published by the Gu lab, Wang et al. used a BAC-based transgenic $Ngn3^{Cre}$ driver and the $Rosa26^{EYFP/+}$ reporter to track the fate of *Ngn3*-expressing cells at different *Ngn3* expression levels. In the wild-type pancreas, more than 85% of the $Ngn3^{+}$ cells become endocrine cells. However, when the *Ngn3* expression level is decreased in the haploinsufficient $Ngn3^{+/-}$ mouse pancreas, about 45% *Ngn3*-expressing cells are shunted to the exocrine lineages, suggesting that *Ngn3*

expression should reach a threshold level to initiate the endocrine differentiation program (Wang et al., 2010). hESCs with *Ngn3* being knocked-out with CRISPR/Cas9 cannot differentiate into endocrine cells but hESCs with *Ngn3* being knocked-down with shRNA can form endocrine cells, though fewer than unmanipulated hESCs , again suggesting the importance of *Ngn3* level in endocrine specification (McGrath et al., 2015). In fact, *Ngn3*-expression level is semi-quantified with immunostaining and two populations, $Ngn3^{low}$ and $Ngn3^{high}$ cells, are reported in various articles (Seymour et al., 2008; Shih et al., 2012). These results suggest that *Ngn3* dosage can affect the differentiation potential of $Ngn3^{+}$ progenitors, underscoring the importance of understanding *Ngn3* expression regulation.

Many factors have so far been reported to regulate *Ngn3* expression. For instance, *Sox9*, *Pdx1*, and *Hnf6* activate *Ngn3* expression (Jacquemin et al., 2000; Lynn et al., 2007b; Oliver-Krasinski et al., 2009), while Notch signaling inhibits *Ngn3* expression (Apelqvist et al., 1999; Lee et al., 2001). It is also reported that *Ngn3* can repress its own expression and this self-inhibitory effect is proposed to explain the reduced expression of *Ngn3* after lineage commitment (Smith et al., 2004). However, this result comes from an *in vitro* cell transfection and overexpression experiment and may not reflect the *in vivo* situation.

Of all the factors that regulate *Ngn3* expression, Notch signaling is the predominant signaling pathway. In Notch signaling, binding of the membrane-bound ligand (Delta, Jagged, *etc*.) to the Notch receptor in adjacent cells triggers two successive proteolysis events of the Notch receptor and results in the release of the Notch intracellular domain (NICD domain). One of the two successive proteolysis events is mediated by the γ-secreatase. The γ-secretase is a protein complex composed of four proteins, including Presenilin 1 (Psen1), Nicastrin, Anterior Pharynx-defective 1 (APH-1), and Presenilin Enhancer 2 (PEN-2). The released NICD domain will translocate into the

nucleus and activate Notch downstream genes such as *Hes1*. Hes1 in turn represses the expression of *Ngn3*. In this way, Notch signaling can repress *Ngn3* expression (Edlund, 2001; Kim et al., 2010). Mice deficient for the NICD-binding coactivator gene *Rbp-jκ* or the Notch ligand gene *Dll1* display overexpression of *Ngn3* and precautious endocrine differentiation at the expense of exocrine lineages (Apelqvist et al., 1999). Repression of Notch signaling by expressing a dominant negative N3IC protein has the same effects (Apelqvist et al., 1999). In addition, deletion of *Hes1* causes significant pancreatic hypoplasia (Jensen et al., 2000). It is notable that activation of Notch signaling usually leads to Notch inactivation in the neighboring cells, a phenomenon called lateral inhibition (Edlund, 2001; Kim et al., 2010). The lateral inhibition model has been proposed to express *Ngn3* in selective cells and maintain the balance between endocrine and exocrine lineages. However, contrary to this model, careful immunostaining always shows that many *Ngn3*-expressing cells reside side-by-side (Fig 5.1), suggesting that a non-classical lateral inhibition mechanism exists to regulate *Ngn3* expression.

microRNAs (miRNAs) have also been implicated in regulating endocrine development. For instance, knockdown of miR375 with morpholino oligonucleotides causes pancreatic islet abnormalities in zebrafish (Kloosterman et al., 2007). Global *Dicer* deletion in mice is embryonically lethal (Bernstein et al., 2003) and mice deficient for *Dicer* specifically in the pancreas (*Pdx1*^{Cre}-driven) only survive to P3, along with severe pancreas hypoplasia, reduced Ngn3^{+} cells and endocrine cell mass, as well as altered endocrine cell type allocation (Lynn et al., 2007a). *Pdx1*^{Cre}-driven *Dicer* deletion does not affect *Pdx1* expression but cause *Hes1* overexpression, suggesting that the reduction of Ngn3^{+} cells is not due to the depletion of the Pdx1^{+} progenitor pool but because of enhanced Notch signaling. Interestingly, *Ngn3*^{Cre}-driven *Dicer*-knockout mice demonstrate normal islet development and hormone staining (Lynn et al., 2007a). This

suggests that the function of miRNAs is restricted within a narrow time window or that non-cell-autonomous effect exists. Combined, these findings lead us to explore whether miRNAs can regulate *Ngn3* expression.

In this chapter, we propose a modified Notch signaling model that regulates *Ngn3* expression to reconcile the reported discrepancies and to accommodate our preliminary data (Fig 5.8). In this model, Ngn3 activates the expression of several miRNA genes. These miRNAs can then target Notch signaling components Hes1 and Psen1 to tune down Notch inhibition on *Ngn3* expression. In this way, *Ngn3* can enhance its own expression indirectly. Because miRNAs can translocate to the neighboring cells, most likely via gap junctions (Valiunas et al., 2005; Wong et al., 2008), it allows pancreatic progenitors to escape lateral inhibition as multi-cell clusters. Below, I will present the preliminary data that support this model.

5.3 Results

Ngn3$^+$ cells reside side-by-side

Since Notch signaling inhibits *Ngn3* expression, the classical lateral inhibition model predicts that Ngn3$^+$ endocrine progenitors are scattered in the pancreatic epithelium as isolated single cells. However, by immunostaining I found that E14.5 pancreas demonstrated a side-by-side *Ngn3* expression pattern (Fig 5.1), casting doubt on the validity of the classical lateral inhibition model in regulating *Ngn3* expression. Observation of side-by-side Ngn3$^+$ progenitors is reported by others as well (Jensen, 2004). In addition, imaging *Ngn3*$^{EGFP/+}$ pancreas sections revealed neighboring or small clusters of EGFP$^+$ cells (data not shown). This side-by-side *Ngn3* expression pattern violates the prediction by the canonical Notch lateral inhibition model and prompts us to investigate other mechanisms that coexist to regulate *Ngn3* expression.

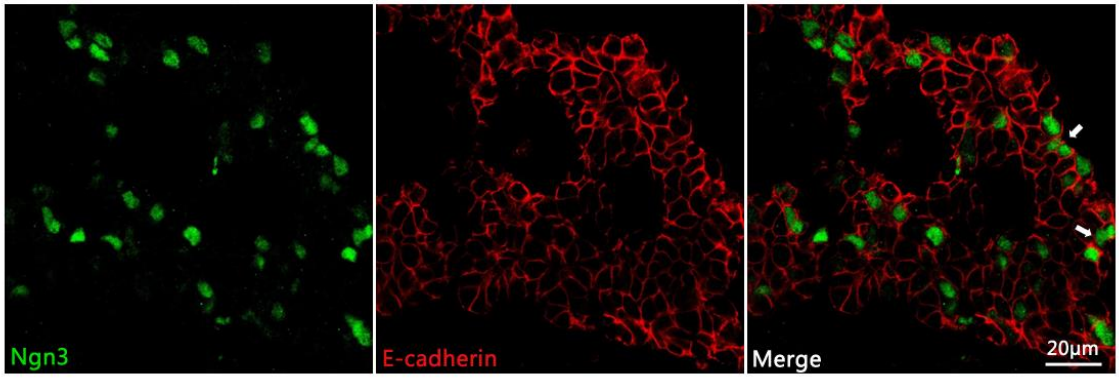


Fig 5.1 $Ngn3^+$ progenitors reside side-by-side or in small clusters. E14.5 pancreas is stained with anti-Ngn3 antibody and anti-E-cadherin antibody. White arrows indicate $Ngn3^+$ clusters that violate the prediction by the classical lateral inhibition model. Scale bar: 20μm.

Ngn3 activates its own expression

We utilized a knock-in *Ngn3*^{EGFP} allele (Lee et al., 2002) to study *Ngn3* expression regulation. This knock-in allele produces EGFP instead of Ngn3 protein. Dr. Gu dissected the E14.5 *Ngn3*$^{EGFP/+}$ and *Ngn3*$^{EGFP/EGFP}$ pancreas and compared their EGFP expression level under a fluorescent microscope. Surprisingly, *Ngn3*$^{EGFP/+}$ pancreas had higher EGFP intensity than the *Ngn3*$^{EGFP/EGFP}$ pancreas did, suggesting that *Ngn3* activated its own expression (Fig 5.2 a-b). Dr. Gu quantified the EGFP intensity with flow cytometry after digesting the embryonic pancreas into single cells. Flow cytometry revealed two EGFP^{+} cell populations: the EGFP^{hi} and EGFP^{low} population in the heterozygous pancreas, but only EGFP^{low} population in the null pancreas (Fig 5.2 e). The same result was observed and reported by the Sander group as well (Shih et al., 2012). These data suggest that *Ngn3* activates its own expression. The discrepancy between our data and the previously reported self-repressing property of *Ngn3* (Smith et al., 2004) may lie on the fact that the latter experiment was done in a cell transfection assay where *Ngn3* is overexpressed and did not reflect the *in vivo* situation.

Dicer is necessary for high *Ngn3* expression

Previous reports have shown that *Dicer* is essential for pancreatic cell survival and *Ngn3* expression (Bernstein et al., 2003; Lynn et al., 2007a), yet the regulation mechanism is not well known. To investigate how *Dicer* regulates *Ngn3* expression at a cellular level, Dr. Gu crossed and obtained the *Dicer*$^{F/F}$*; Pdx1*^{Cre}*; Ngn3*$^{EGFP/+}$ compound mice (*Pdx1*^{Cre}: Gu et al., 2002; *Dicer*$^{F/F}$: Harfe et al., 2005) and analyzed their EGFP expression in the developing pancreas. He found that pancreas-specific *Dicer* deletion did not eliminate *Ngn3* expression. Instead, *Ngn3* expression pattern in the *Dicer* conditional knockout pancreas phenocopied that of the *Ngn3*$^{EGFP/EGFP}$ mice (Fig 5.2 c-d), indicating that *Dicer* is not required to initiate *Ngn3* expression but can enhance *Ngn3*

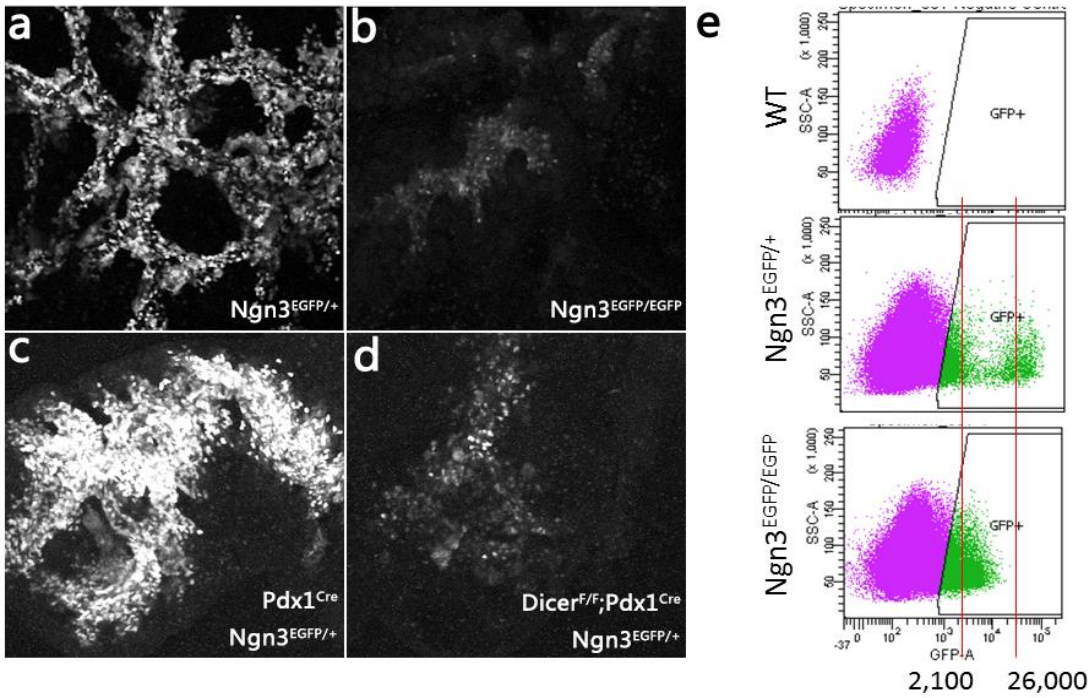


Fig 5.2 Ngn3 activates its own expression. (a-d) The left panel presents confocal images of E14.5 pancreas of different genotypes as labeled. Confocal images show higher EGFP expression in the heterozygous *Ngn3*$^{EGFP/+}$ pancreas (a) compared to the *Ngn3*$^{EGFP/EGFP}$ pancreas (b), suggesting that Ngn3 activates its own expression. Pancreas-specific deletion of *Dicer* (d) phenocopies *Ngn3*-null pancreas (b), indicating that *Dicer* could positively regulate *Ngn3* expression. (e) The right panel is the flow cytometry analysis result of WT, *Ngn3*$^{EGFP/+}$ and *Ngn3*$^{EGFP/EGFP}$ pancreas. Flow cytometry analysis reveals EGFP^{hi} and EGFP^{low} populations and a reduced EGFP^{hi} population is evident in the *Ngn3*$^{EGFP/EGFP}$ pancreas. Source: Dr. Guoqiang Gu.

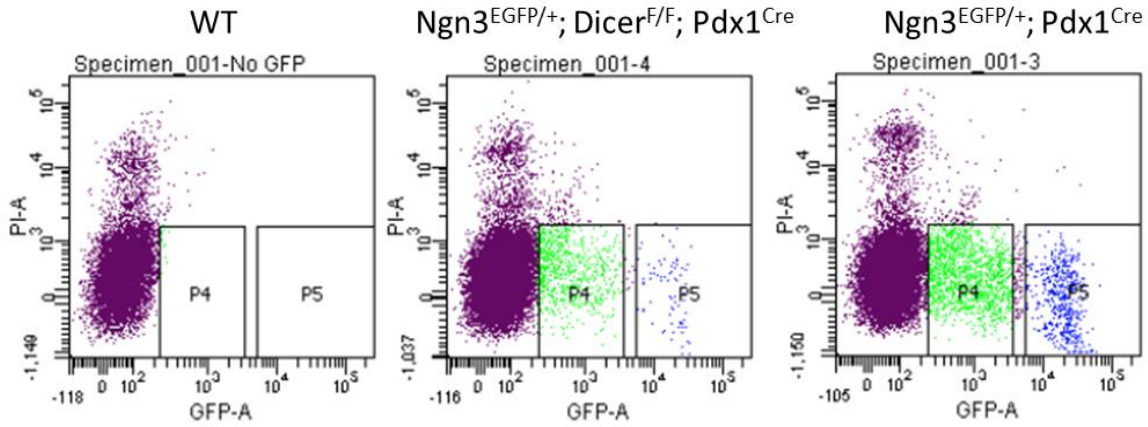


Fig 5.3 *Dicer* is necessary for enhancing *Ngn3* expression. This figure shows flow cytometry analysis results of pancreas of three genotypes as labeled. *Dicer*-deletion reduces $EGFP^{hi}$ cells ($EGFP^{low}$ to $EGFP^{hi}$ estimated ratio 10:1) in contrast to the control pancreas ($EGFP^{low}$ to $EGFP^{hi}$ estimated ratio 3:1). Source: Dr. Guoqiang Gu.

expression. Flow cytometry analysis confirmed a reduced number of $EGFP^{+}$ cells and lower $EGFP^{hi}$ to $EGFP^{low}$ ratio in the *Dicer* conditional knockout pancreas as compared to control pancreas (Fig 5.3), again suggesting that *Dicer*, or in other words, miRNAs, can augment *Ngn3* expression.

RNA-Seq reveals Ngn3-dependent miRNAs

In order to identify what miRNAs are involved in regulating *Ngn3* expression, Dr. Gu sorted out $EGFP^{+}$ cells from the E14.5 $Ngn3^{EGFP/+}$ and $Ngn3^{EGFP/EGFP}$ pancreas and conducted an RNA-Seq experiment. Among the ~500 well characterized miRNAs in the mouse genome (Chiang et al., 2010), only 20 members whose expression differed by 2-fold between the two groups were identified. Specifically, miR7, miR9, miR96 and miR182 were enriched in the $Ngn3^{EGFP/+}$ pancreas compared to the $Ngn3^{EGFP/EGFP}$ pancreas. These miRNAs were of particular interest because miR7 and miR9 were predicted to target Notch signaling component Psen1, whereas miR96 and miR182 were predicted to target Hes1. I used various web-based miRNA target prediction algorithms, including TargetScan.org, miRBase.org and microRNA.org, for cross validation. In addition, I also examined mRNA expression in these two cell populations (described in Chapter III) and did not detect significantly elevated Sox9 or Hnf6 transcripts in the $Ngn3^{EGFP/+}$ pancreas, ruling out the possibility of a positive feedback loop via Sox9 or Hnf6. These above findings suggest that Ngn3 activates the expression of the above several miRNA to enhance its own expression, pushing pancreatic progenitors towards endocrine cell fate.

In order to validate the predicted targets of miRNAs, I performed luciferase/GFP reporter assays (Fig 5.4 a). miRNAs were cloned from mouse genomic DNA and constructed into the SIBR vector for expression (Chung et al., 2006). The 3' untranslated region (UTR) sequences of predicted target genes, including Hes1 and Psen1, were

cloned into luciferase or mCherry reporter vectors. miRNA-expressing plasmids derived from the SIBR vector and the luciferase/mCherry reporter plasmids were co-transfected into 293T cells and analyzed with a luminometer or flow cytometry in the end. Meanwhile, I designed and constructed mutated miRNAs with directed mutagenesis PCR to introduce point mutations in the miRNAs' seed sequences (the conserved sequence usually from nucleotide 2 to 7 of a mature miRNA) (Horwich and Zamore, 2008). Mutation in the seed sequences of miRNAs abolishes miRNAs' capacity to knockdown their mRNA targets. These mutant miRNA-expressing plasmids were tested side-by-side with wild-type miRNA-expressing plasmids for validating sequence specificity (Table 5). I found that both miR96 and miR182 downregulated Hes1 level. Mutations of miR96 and miR182 partially rescued the phenotype but did not fully restore to the negative control level. miR9, however, quite unexpectedly upregulated Psen1 (Fig 5.4 b). It was not clear why miR9 upregulated Psen1 expression but there were technical issues and theoretical reasons to explain it. First of all, when I repeated the luciferase assay several times, the results were not always consistent. In theory, firefly luciferase and the normalization control Renilla luciferase are relatively stable during the time span of my assay (Thompson et al., 1991). It was not likely that loss of activity or degradation of luciferases caused the inconsistency. That being said, I constructed mCherry reporters and will try to use flow cytometry to analyze the inhibitory effects of miRNAs on their targets. Second, in the cell transfection assay, miRNAs and reporters are overexpressed as compared to the physiological expression level. The ratio between miRNAs and target mRNAs also affects the readout and thus titrating transfected plasmids to a higher miRNA/target ratio is necessary, otherwise the miRNA interference effect might not be obvious or even masked (Kuhn et al., 2008). Preparing stable cell lines harboring the reporters is also under consideration. Lastly, the cloned regions of 3'UTR sequences of *Psen1* and *Hes1* were chosen based on bioinformatics predictions. They were only part

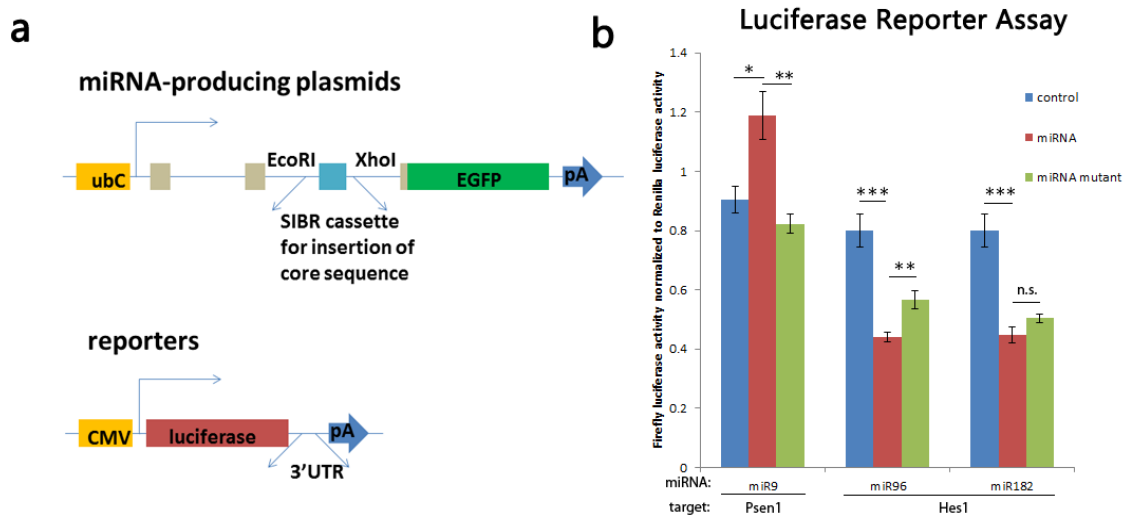


Fig 5.4 Validating miRNA targets with luciferase reporter assay. (a) Schematic representation of the design of miRNA-expression plasmids and luciferase reporters. miRNA genes are cloned from the mouse genome and are inserted in-between the EcoRI and XhoI sites of the SIBR vector. 3'UTR sequences of target genes are cloned from the mouse genome and are inserted downstream of a luciferase expression cassette. (b) Quantification of luciferase reporter assay. Firefly luciferase activity is normalized to Renilla luciferase activity. A control SIBR plasmid, a miRNA-expressing plasmid, and a mutant miRNA-expression plasmid are tested side-by-side for testing sequence specificity. Data are represented as mean±s.e.m. p-value: ***<0.001<**<0.01<*<0.05<n.s.

of the 3'UTR sequences. It is possible that miRNAs have additional target sites outside of the cloned region or even within the coding sequences (Forman et al., 2008).

Gja1 is expressed in embryonic pancreas

It is previously known that gap junctions exist in endodermal progenitors, which is necessary for their survival (Saund et al., 2012). It has come to our attention that miRNAs can freely transport through gap junctions and enable neighboring cells to communicate (Valiunas et al., 2005; Wong et al., 2008). We envision that communication among neighboring cells via miRNAs may explain the observation of side-by-side $Ngn3^{+}$ cells that the classical lateral inhibition model could not.

First of all, I detected the expression of a gap junction protein, Gja1 (connexin43), in embryonic pancreas (Fig 5.5). Gja1 forms relatively large gap junction channels that allow siRNAs to move between cells (Valiunas et al., 2005), fitting in our proposed models. Gja1, Gjc1 (connexin45), and Gjd2 (connexin36) are previously reported to be expressed in mouse pancreatic islets (Theis et al., 2004); they are also detected by our RNA-Seq analysis (data not shown).

To test the functional involvement of gap junctions in regulating *Ngn3* expression, Dr. Gu blocked gap junctions with inhibitors 18-α-glycyrrhetic acid or 18-β-glycyrrhetic acid in an *ex vivo* pancreas culture assay. The number of $EGFP^{+}$ cells in $Ngn3^{EGFP/+}$ pancreas decreased but more $EGFP^{Hi}$ cells surged (Fig 5.6). Since 18-α-glycyrrhetic acid or 18-β-glycyrrhetic acid blocked the free translocation of miRNAs, miRNAs could not enhance *Ngn3* expression by repressing Notch signaling in the neighboring cells, thus less $EGFP^{+}$ cells formed. Meanwhile, miRNAs were concentrated in selective cells because of the translocation restriction, leading to more $EGFP^{Hi}$ cells by reinforcing the expression from the *Ngn3* promoter.

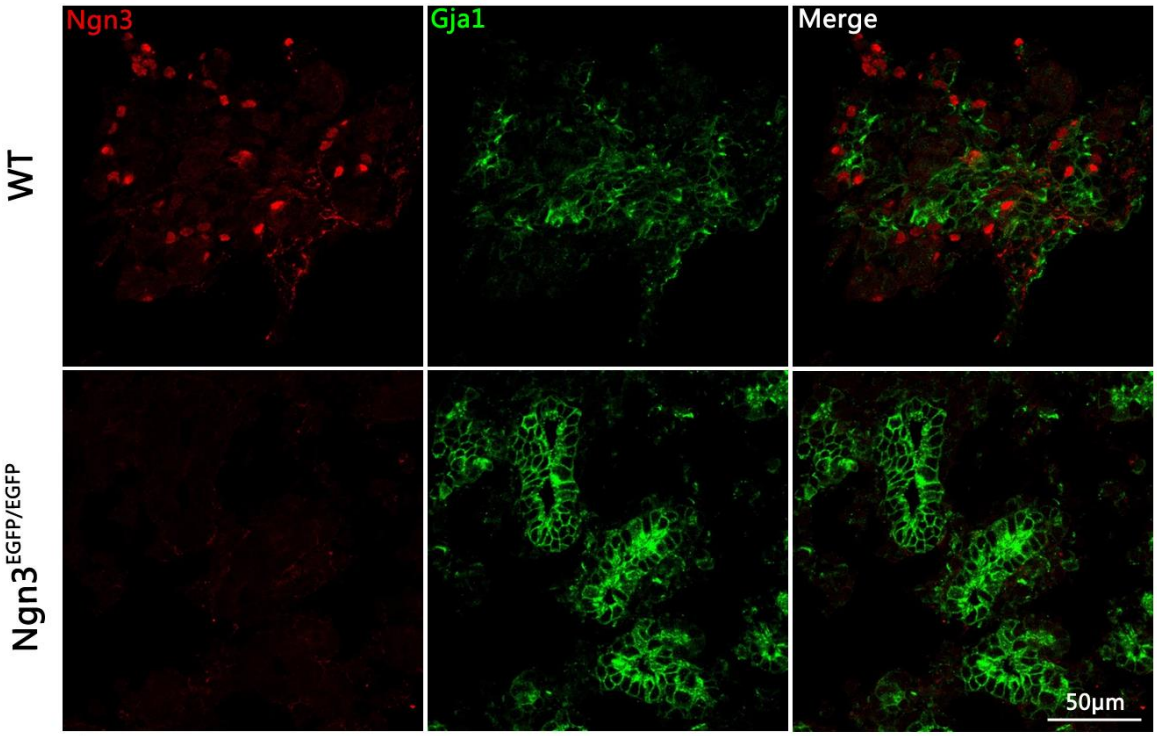


Fig 5.5 *Gja1* is expressed in embryonic pancreas. Wild-type and *Ngn3*$^{EGFP/EGFP}$ pancreas are co-stained with anti-Ngn3 and anti-Gja1 antibodies. *Gja1* is expressed in the developing pancreas and is *Ngn3*-independent. In fact, *Gja1* expression is higher in the *Ngn3*$^{EGFP/EGFP}$ pancreas, both semiquantitatively by confocal imaging and by RNA-Seq (see main text for details). Stage: E14.5. Scale bar: 50μm.

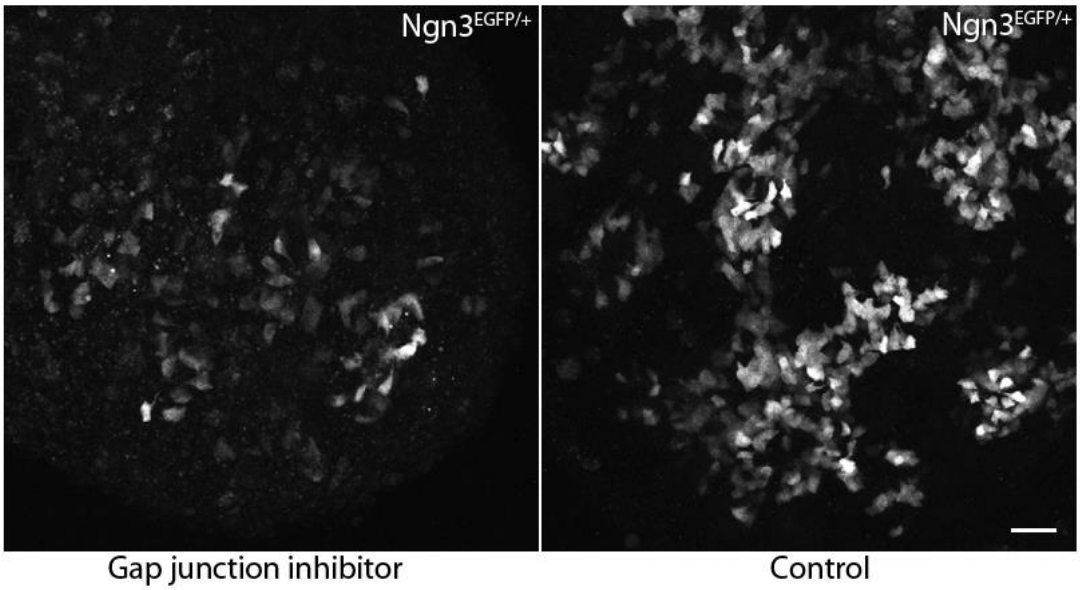


Fig 5.6 Blocking gap junctions in the *Ngn3*$^{EGFP/+}$ pancreas results in intensified EGFP expression per cell but loss of EGFP^{+} cells. In an *in vitro* pancreatic bud culture assay, dissected E8.5 *Ngn3*$^{EGFP/+}$ embryonic pancreas is cultured in medium in the presence or absence of a gap junction inhibitor, 18-α-glycyrrhetic acid (or 18-β-glycyrrhetic acid). Gap junction inhibition results in more EGFP^{Hi} cells but less EGFP^{+} cells, an effect probably due to the restriction of miRNA translocation across gap junctions. Scale bar: 20μm. Source: Dr. Guoqiang Gu.

5.4 Discussion

The pro-endocrine transcription factor Ngn3 plays a central role in pancreatic endocrine specification. *Ngn3* expression is initiated in the $Pdx1^+$ pancreatic epithelium domain and can be detected as early as E9.5. It reaches its peak of expression at around E14.5-E15.5 and then gradually declines during the perinatal stage. *Ngn3* expression is maintained at a very low level in adult islets and the Ngn3 protein is hardly detectable with regular immunostaining (Wang et al., 2009a). Although *Ngn3* expression almost diminishes in the adult pancreas, some researchers report that *Ngn3* expression can be reinitiated in the adult pancreas and repopulate β cells. (Al-Hasani et al., 2013; Xu et al., 2008). Ngn3 is also a key transcription factor in transdifferentiating acinar cells into β cells (Li et al., 2014b; Zhou et al., 2008). Knocking down or completely deleting *Ngn3* in mouse pancreatic progenitors or hESCs leads to the failure of massive endocrine cell differentiation (McGrath et al., 2015; Sugiyama et al., 2013). All these results underscore the central role of *Ngn3* in pancreatic endocrine differentiation.

Ngn3 expression is subject to Notch repression. The classical Notch lateral inhibition model prohibits adjacent cells to become Ngn3-positive simultaneously but this model could not explain our observation of side-by-side $Ngn3^+$ cells. It is likely that unidentified cell-cell communication allows neighboring cells to break this lateral inhibition constriction. This communication could be mediated by protein-receptor interaction, or by direct information exchange via junctional channels. Here I focused on gap junction communications. Indeed, we detected the expression of *Gja1* in E14.5 pancreas with immunostaining and other connexins with RNA-Seq. Furthermore, blockade of gap junctions with 18-α-glycyrrhetic acid or 18-β-glycyrrhetic acid reduced $EGFP^+$ cells in the *in vitro* cultured $Ngn3^{EGFP/+}$ pancreas. These findings allow us to propose that miRNAs can translocate to neighboring cells via gap junctions and execute

their function non-cell-autonomously, allowing the formation of $Ngn3^+$ clusters. This type of coordinated *Ngn3* expression in many pancreatic progenitor cells is necessary to warrant the production of enough β cells for normal physiology.

Functional involvement of miRNAs in pancreas development

To investigate the functional invovlement of miRNAs, it is desirable to perform loss-of-function or gain-of-function experiements. To this end, antagomirs can be administered in an *in vitro* assay to inhibit miRNA activity. Antagomirs are RNA analogs that can bind with and silence their target miRNAs (Krutzfeldt et al., 2007; Krutzfeldt et al., 2005). *In vitro* cultured pancreatic buds can be treated with antagomirs and their transcription profile can then be analyzed by immunostaining, qRT-PCR, and flow cytometry *etc*.

A primary advantage of antagmirs over genetic knockout is that antagomirs can block miRNAs of the same family and of the same seed sequence. Both miR7 and miR9 have multiple members in their family, and it is difficult to genetically knockout all miRNA genes because of this genetic redundancy, let alone the demanding time and cost. Moreover, the pharmaceutical potentials of antagomirs make the research promising for future clinical translation. There are concerns about antagomir efficiency and specificity though. To this end, antagomirs could be modified with a cholesterol moiety to enhance the delivery efficiency (Horwich and Zamore, 2008). Scrambled or mutated antogamirs should be utilized in parallel to serve as specificity controls.

To achieve *in vivo* knockdown, complementary miRNAs could be overexpressed as “miRNA sponges” to block the endogenous miRNAs of interest (Ebert et al., 2007; Ebert and Sharp, 2010; Kluiver et al., 2012). Like antagomirs, miRNA sponges bypass the redundancy problem of miRNA gene families. In a similar fashion, miRNAs can be overexpressed, often in targeted tissue and in a temporally controlled manner (Chen et

al., 2011; Chung et al., 2006), to analyze their functional involvement in pancreas development.

Does non-cell-autonomous effect exist?

Although we found that the blockade of gap junctions by chemical inhibitors interfered with *Ngn3* expression, it is not clear whether this is a direct result of blocked miRNA transportation or only a secondary effect. Although gap junctions are shown to be permeable to miRNAs in cell assays (Valiunas et al., 2005), a fundamental question of our model is whether miRNAs translocate among neighboring cells and function non-cell-autonomously in the developing pancreas. To this end, fluorophore moiety-conjugated or radioactive-labeled miRNAs, together with microinjection and high resolution microscopy, can be used to examine the movement of miRNAs within and among cells.

If non-cell-autonomous effect does exist, I expect that miRNAs produced in one cell will likely rescue *Dicer* defects in neighboring cells. To test this possibility, I can create mosaic *Dicer* deletion in the pancreatic progenitor cells and examine whether *Ngn3* expression in these cells can be rescued by neighboring Dicer^{+} cells (Fig 5.7). Mosaic *Dicer* deletion can be achieved by administering a low dose of tamoxifen to pregnant female mice and harvesting *Dicer*$^{F/F}$*;Pdx1*^{CreER}*;Ngn3*$^{EGFP/+}$ embryos. Then co-localization of Dicer and Ngn3/EGFP can be examined by immunostaining. Alternatively if no good anti-Dicer antibody is available, *Dicer* allele copy number in the EGFP^{+} cells can be quantified and compared between the TM-treated group and the control group. In brief, EGFP^{+} cells are collected by FACS. These cells are homogenized and DNA is extracted. The copy number of *Dicer* allele is then quantified with qRT-PCR. I will then measure the copy number of *Dicer* allele on a per cell basis by normalizing it to the EGFP^{+} cell number. This result will test the existence of non-cell-autonomous effects. If

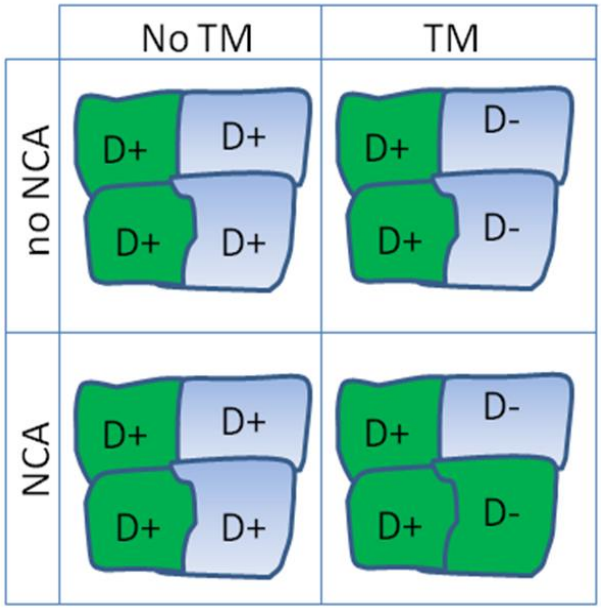


Fig 5.7 A proposed experiment for testing non-cell-autonomous effect. Mosaic *Dicer* deletion is achieved by administering timed pregnant female mice with a low dose of tamoxifen (TM). *Dicer*$^{F/F}$;*Pdx1*^{CreER};*Ngn3*$^{EGFP/+}$ embryos will be dissected and dissociated into single cells, which are subject to FACS and quantitative real-time PCR for measuring *Dicer* allele copy number. *Dicer* allele copy number is normalized to EGFP^{hi} cell (green cells) number. The normalized *Dicer* copy number is a measurement for testing the existence of non-cell-autonomous effect. TM: tamoxifen. NCA: non-cell-autonomous. D+: there is an effective Dicer gene. D-: Dicer null. Cells of green color represent EGFP^{hi} cells.

non-cell-autonomous effect does not exist, then the ratio of *Dicer* allele to $EGFP^+$ cell number will be approximately the same between TM-treated and non-treated pancreas. In contrast, if non-cell-autonomous effect exists, then the ratio of *Dicer* allele to $EGFP^+$ cell number will be lower in the TM-treated pancreas than the non-treated controls.

Gap junction genetic models and pancreas development

Gap junctions are passive diffusion channels between adjacent cells. A gap junction is composed of two connexons; each is a hexamer of connexins and is docked on the membrane of two adjacent cells. Gap junctions have selective permeability for small molecules, such as ions (Na^+, K^+, Cl^-, Ca^{2+}), second messengers (cAMP), amino acids and siRNAs (Valiunas et al., 2005; Wong et al., 2008). It is well-established that endodermal progenitor cells communicate through gap junctions (Saund et al., 2012).

In order to fully dissect the functional involvement of gap junctions in pancreas development, genetic models are preferred, complementary to the above chemical inhibition assay. To this end, we can take advantage of the existing connexin alleles, for example, *Gja1* knockout (Eloff et al., 2001), *Gjc1* knockout (Kumai et al., 2000), and *Gjd2* knockout mice (Guldenagel et al., 2001) *etc*., for the analysis of their role in pancreas development. Due to the wide expression of connexins, global knockout might cause growth or developmental defects in multiple organs, such as the heart and the eyes *etc*. (Guldenagel et al., 2001; Kumai et al., 2000). To circumvent this problem, pancreas- or β-cell-specific deletion of connexins can be achieved by using the floxed connexin alleles (Liao et al., 2001; Nishii et al., 2003). One disadvantage of genetic knockout lies on the connexin family redundancy, which should be taken into consideration when interpreting results. In addition, $Ngn3^+$ progenitor-specific overexpression of *Gja1*, *Gjb1*, and *Gjd2* could also be employed to investigate the

functional involvement of gap junction in pancreas development and endocrine lineage allocation (Klee et al., 2011).

How is Ngn3 downregulated?

Ngn3 is transiently expressed in the early pancreatic endocrine progenitor cells. Once the $Ngn3^{+}$ progenitors differentiate into endocrine cells, *Ngn3* expression declines. The previously reported self-repression of *Ngn3* expression helps to explain the transient *Ngn3* expression nature. However, if our model is correct and *Ngn3* activates its own expression, how could it be down-regulated during pancreas development to prevent the unlimited expansion of endocrine progenitors? One possiblity is that the affinity of Ngn3 protein for its targets differs and that Ngn3 binds to its own promoter and inhibits itself when Ngn3 protein level reaches a higher threshold. The other possiblity is that gap junctions could be shut down.

Indeed, I found that *Gja1* and *Ngn3* expression were negatively correlated. Gja1 level was higher in the *Ngn3*-null pancreas than in the wild-type pancreas. This discovery by immunohistochemsitry was consistent with our RNA-Seq data. It seems that there is a feedfack mechanism between *Ngn3* and *Gja1* to fine tune the expression level of *Ngn3*. At the early stage of development, gap junctions augment *Ngn3* expression and $Ngn3^{+}$ progenitor expansion by allowing the translocation of miRNAs to neighboring cells. As *Ngn3* expression increases, *Gja1* is downregulated to restrict the dilution of miRNAs, allowing the formation of $Ngn3^{Hi}$ cells and preventing $Ngn3^{+}$ progenitor expansion. Gradually, as $Ngn3^{Hi}$ cells differentiate into endocrine cells and *Ngn3* expression diminishes, *Gja1* expression is turned on again, possibly serve a function in endocrine maturation or functional maintenance. Indeed, gap junctions exist in the mature β cells and control insulin secretion synchronization, presumably through

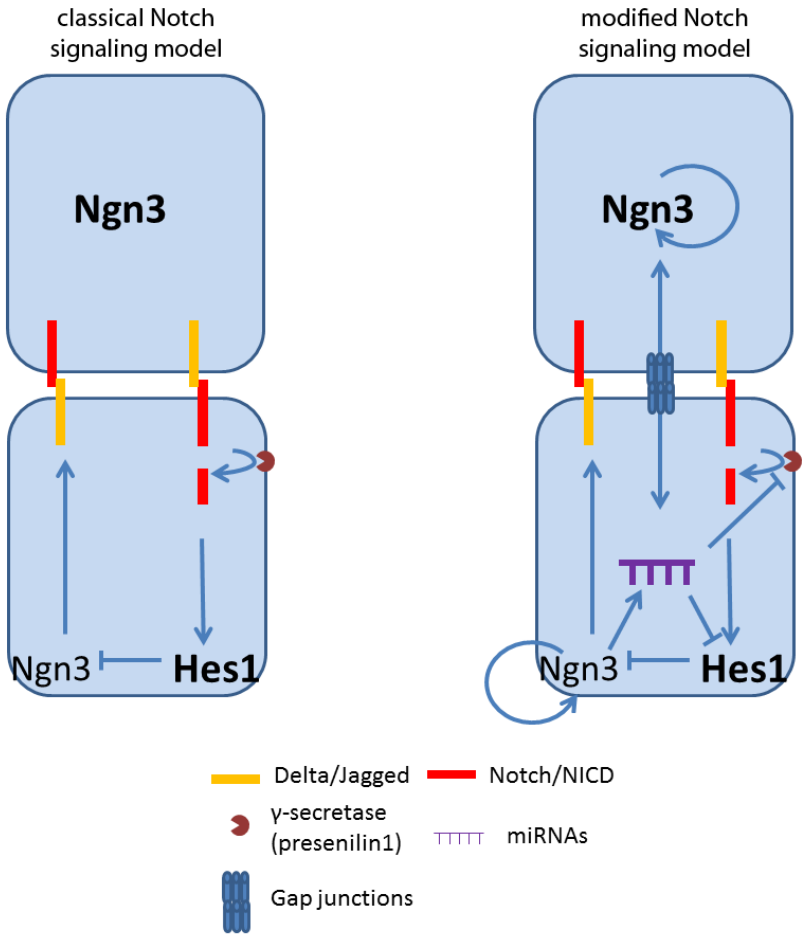


Fig 5.8 A modified Notch lateral inhibition model. In the modified model, inhibition of Notch signaling on *Ngn3* expression is relieved by miRNA-mediated repression of Hes1 and Psen1, both of which are Notch signaling components. These miRNAs also allow neighboring cells to communicate with each other, explaining the observation of side-by-side $Ngn3^{+}$ cells.

mediating Ca^{2+} flux (Benninger et al., 2011; Calabrese et al., 2003; Serre-Beinier et al., 2009).

In this chapter, we proposed a model (Fig 5.8) which integrates Notch signaling and miRNA-mediated gap junction communication. This new model resolves the discrepencies between lateral inhibition prediction and actual observation, and is supported by lines of evidences. Further investigation is requried to validate the functional involement of miRNAs and gap junctions as well as elucidate the molecular mechanisms. Understanding *Ngn3* expression regulation is central in pancreas organogenesis research and *in vitro* β cell differentiaiton.

5.5 Materials and methods

Mouse strains and care

Mouse strains *Dicer*$^{F/F}$ (Harfe et al., 2005) and *Ngn3*$^{EGFP/+}$ (Lee et al., 2002) are purchased from the Jackson Laboratory. *Pdx1*^{Cre} mouse was made by Dr. Gu (Gu et al., 2002). All mice were housed and cared in the Vanderbilt Division of Animal Care and in compliance to IACUC regulations.

FACS assay

Ngn3$^{EGFP/+}$ or *Ngn3*$^{EGFP/EGFP}$ pancreas was dissected out at E10.5, E15.5 and E18.5. Pancreas was then dissociated into single cells with trypsin and subject to flow cytometry analysis or cell sorting. *Ngn3*$^{+/+}$ embryos were used as negative controls. DAPI was added as a viability marker. Flow cytometry analysis and cell sorting experiments were performed by the Vanderbilt Flow Cytometry Shared Resources. RNA-Seq data was previously obtained (data unpublished). In general, $EGFP^{+}$ cells

collected by FACS were homogenized and total RNA was extracted. RNA samples were sent to the Vanderbilt Vantage Sequencing Core for sequencing and analysis.

Immunohistology

Primary antibody: guinea pig anti Ngn3 (1:100), mouse anti-E-cadherin (1:200), goat anti-Ngn3 (1:1000), rabbit anti-Gja1 (Abcam Ab11370, 1:1000). Secondary antibody: 1:500 FITC-conjugated donkey anti-guinea pig (1:500), Cy3-conjugated donkey anti-mouse (1:500), FITC-conjugated donkey anti-rabbit (1:500, all secondary antibodies are from Jackson Immunoresearch). Staining procedures followed previously published methods (Wang et al., 2008; Wang et al., 2010).

Plasmid construction

Plasmids were constructed with conventional molecular cloning methods. miRNA genes and 3'UTR of Hes1 and Psen1 sequences were cloned from mouse genomic DNA and inserted into pBluescript KS II vectors followed by sequencing. Correct sequences were then subcloned into SIBR plasmids (for miRNA expression; EcoRI and XhoI), or a pCS2-based luciferase reporter plasmid (a gift from Dr. Ethan Lee, Vanderbilt), or a pmCherry-C1 (Clontech) based fluorescent reporter plasmid.

Primers for cloning miR7, miR9, miR96 and miR182:

miR9-1s: 5'agaattcgagactacggaggtccag3', miR9-1a: 5'actcgagcgcgaggtggctcgggctg3'; miR7-2s: 5'agaattctagggaactgtatgagcag3', miR7-2a: 5'actcgagccttctgaggtttcctcaactg3'; miR96s: 5'agaattcataaacagagcagagacagatc3', miR96a: 5'actcgagccagctcggattgcccagctc3'; miR182s: 5'agaattcactggaacaggaccatacagg3', miR182a: 5'actcgagccttttcaccgagaagaggtc3'.

Primers for cloning 3'UTR sequences of Hes1 or Psen1:

Hes1s: 5’agaattctgactgagagcctcaggccactgc3’, Hes1a: 5’aggatccttcactcttttattatattttctca3’; Psenmir7-s: 5’gtgtcacaagacatggaccatcgt3’, Psenmir7-a: 5’gcaagtgagcctccttcatcga3’; Psen1mir9-s: 5’ctgacagcagacaaggcagctct3’, Psen1mir9-a: 5’cccagcattggacattactcgga3’; Psen1st pa-s: 5’aaccatagcctgctttgtagccat3’, Psen1st pa-a1: 5’ggcttgctctctgtttttgtgttt3’.

Primers used for introducing mutations in miRNA plasmids:

miR7_mu_1: 5’ttcaaacggggctggccc3’, miR7_mu_2: 5’cgtctagtgattttgttgttgt3’; miR7_mu_3: 5’aatcactagacgttcaaacgg3’, miR7_mu_4: 5’ttgttgttgtgtctctgtatcc3’; miR9_mu_1: 5’cacgataacaaccaaccccg3’, miR9_mu_2: 5’gattatctagctgtatgagtgg3’; miR96_mu_1: 5’caaaatcggccaagcagatg3’, miR96_mu_2: 5’aagctagcacatttttgcttgt3’; miR182_mu_1: 5’cgaaaatggtgggaggcct3’, miR182_mu_2: 5’acgatggtagaactcacacc3’.

All constructs were sequenced for confirmation.

Luciferase assay

293T cells were co-transfected with SIBR-miRNA plasmid, pCS2-luciferase-3’UTR reporter plasmid, and a Renilla luciferase reporter. 24 hours after cell culture, cells were washed with PBS and ready to be processed. The Dual-Luciferase Reporter Assay System from Promega (Cat # E1960) was used for luciferase assay. Luciferase activity was measured with an illuminator. 3 samples were prepared for each condition and each sample was measured twice. Firefly luciferase activity was normalized to a Renilla luciferase control.

Pancreatic bud culture and gap junction inhibition

E8.5 *Ngn3*$^{EGFP/+}$ pancreas was dissected and cultured in RPMI 1640 medium (Gibco, 11875093) with the addition of 18-α/β-glycyrrhetic acid (75μM) for 48hrs.

Medium and drug was changed every day. DMSO was used as a vehicle control. At the end of 48hrs, pancreas buds were retrieved for confocal imaging.

Confocal microscopy

All confocal images were taken with Leica TCS-SP5 scanning confocal microscopy or Olympus FV-1000 confocal microscopy

CHAPTER VI

CONCLUSIONS AND FUTURE DIRECTIONS

Summary

Diabetes is a worldwide health issue. In both type I and late stage type II diabetes, significant β-cell loss causes insulin deficiency and hyperglycemia. Replenishing β cells is a promising therapy and that requires either activating *in vivo* β-cell replication, reinitiating a β-cell neogenesis program or transplanting exogenous β cells. This thesis investigates mouse pancreatic endocrine cell differentiation during embryogenesis. With a better understanding of *bona fide* β-cell differentiation, my research will provide useful information for the development of gene and cell therapies for diabetes mellitus.

The transcription factor Ngn3 specifies endocrine pancreas (Gu et al., 2002). However, how $Ngn3^+$ pro-endocrine progenitors are specified to each endocrine cell type is not well understood, and this issue is the main focus of this thesis. In this thesis, I propose a model that combinatorial transcription factor expression specifies $Ngn3^+$ progenitors to specific endocrine cell fate(s). By using a novel bipartite Cre cell lineage tracing system, we were able to show that $Ngn3^+Myt1^+$ cells preferred β cells over α cells. Transcriptional and epigenetic analysis of $Ngn3^+$ progenitors from different embryonic stages revealed that gene expression and DNA methylation of endocrine genes, including *Myt1*, underwent dynamic changes along the developmental timeline, which may explain why endocrine cells are generated in a competence window-dependent manner. To dissect the differentiation potential of $Ngn3^+Myt1^+$ progenitors at different embryonic stages, I designed a tamoxifen-inducible bipartite CreERT2 construct and

characterized it in cell lines. It could be used to generate mouse models in the future to label Ngn3^{+}Myt1^{+} progenitors at specific embryonic stages and trace their differentiation preference. As part of the Cre technique, variability of Cre reporter sensitivity was observed and non-parallel recombination of floxed alleles in the same cell is reported here. In addition, we investigated the regulation of *Ngn3* expression by Notch signaling and miRNAs. We found that *Ngn3* augmented its own expression by a mechanism of miRNA-mediated inhibition of Notch signaling and explored the possibility of miRNA translocation through gap junctions to attenuate Notch signaling in a non-cell-autonomous manner. Yet, with new discoveries arise more questions. In the following, I will discuss some future directions we may pursue.

Future Directions

Ngn3^{+} progenitor heterogeneity revisited

With the bipartite Cre cell lineage tracing method, we found that the Ngn3^{+}Myt1^{+} progenitors favorably differentiated into β cells rather than α cells. It should be noted, however, that not all Ngn3^{+}Myt1^{+} cells became β cells. The Ngn3^{+}Myt1^{+} progenitors could differentiate into all four types of endocrine cells only with a preference for β cells. On the other hand, the lineage tracing reporter *Ai9* did not label all β cells, suggesting either incomplete labeling or that other Ngn3^{+} progenitor subtypes contribute to β-cell lineage in addition to the Ngn3^{+}Myt1^{+} progenitors. As a matter of fact, we also generated an *Nkx2.2^{cCre}* allele and found that the Ngn3^{+}Nkx2.2^{+} progenitors also favored β-cell fate. This leads me to think whether the Ngn3^{+}Myt1^{+} progenitors and the Ngn3^{+}Nkx2.2^{+} progenitors are the same population or how much these two populations overlap. When I did co-immunostaining of Ngn3, Myt1 and Nkx2.2 on E14.5 pancreas sections, I observed Ngn3^{+}Myt1^{+}Nkx2.2^{+}, Ngn3^{+}Myt1^{+}Nkx2.2^{-}, Ngn3^{+}Myt1^{-}Nkx2.2^{+}, and Ngn3^{+}Myt1^{-}

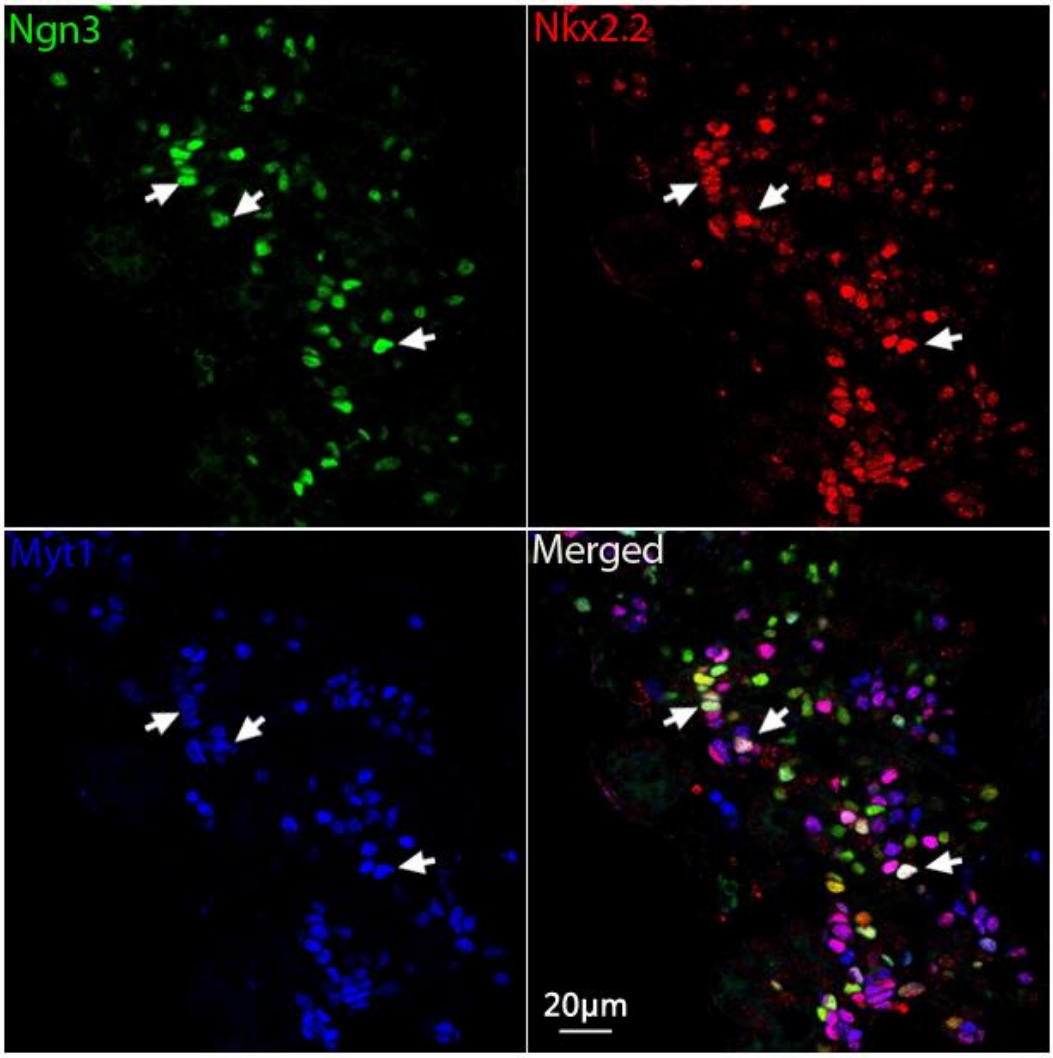


Fig 6.1 Ngn3, Myt1 and Nkx2.2 co-staining reveals a highly heterogeneous progenitor pool. Co-immunostaining of Ngn3, Myt1 and Nkx2.2 on E14.5 wild-type pancreas. White arrows indicate $Ngn3^{+}Myt1^{+}Nkx2.2^{+}$ cells. Scale bar: 20μm.

Nkx2.2$^-$ cells (Fig 6.1), suggesting that the Ngn3$^+$Myt1$^+$ progenitors can be further divided into subpopulations based on the expression of other endocrine transcription factors, e.g., Nkx2.2. How such mosaic expression pattern is formed, whether it is a pre-determined or stochastic event, is beyond our understanding at this moment, but this highly heterogeneous pattern may partially explain the mixed lineages differentiated from the Ngn3$^+$Myt1$^+$ progenitors on an intuitive level. That is, the Ngn3$^+$Myt1$^+$Nkx2.2$^+$ progenitors and the Ngn3$^+$Myt1$^+$Nkx2.2$^-$ progenitors have their preferences for certain endocrine lineage(s). Therefore, the Ngn3$^+$Myt1$^+$ progenitors can still give rise to all endocrine cell types only with an overall preference for β cells.

Ngn3$^+$ progenitors are more heterogeneous than we can image. As is shown in Fig 3.1, *Pax4* is also nonuniformly expressed in the Ngn3$^+$ progenitor pool. Due to the lack of proper antibody, I was unable to examine the expression of *Arx* in embryonic pancreas but published reports demonstrate its heterogeneous co-expression pattern with *Ngn3* (Collombat et al., 2003). Similarly, *Rfx6* (Soyer et al., 2010) and *Insm1* (Mellitzer et al., 2006; Osipovich et al., 2014) also shows heterogeneous co-expression pattern with *Ngn3* and is *Ngn3*-dependent (Soyer et al., 2010). Although no co-staining of all these transcription factors has been done yet, one can imagine the complexity and dynamics of Ngn3$^+$ progenitors during pancreas development. The recently developed MultiOmyx technique could be employed to analyze the heterogeneity of Ngn3$^+$ progenitors (Gerdes et al., 2013).

Then there comes a question: can we experimentally dissect the fate of Ngn3$^+$Myt1$^+$Nkx2.2$^+$ progenitors with cell lineage tracing tools similar to the bipartite Cre?

The answer may lie in the mechanism of Cre/loxP-mediated DNA recombination. During Cre/loxP-mediated recombination, a Holliday junction consisting of four Cre proteins and two loxP sites forms. A loxP site (ATAACTTCGTATAGCATACATTATACGAAGTTAT) contains two 13mer palindromic

sequences connected by an 8mer spacer (underlined). Each 13mer is occupied by one Cre recombinase, thus each loxP site is occupied by two Cre molecules. A functional homotetramer of Cre is formed on two loxP sites.

Researchers have developed engineered Cre proteins that recognize mutated loxP sequences (Gelato et al., 2008; Santoro and Schultz, 2002; Saraf-Levy et al., 2006). One loxP mutant, termed loxM7 (ATAACTCTATATAGCATACATTATATAGAGTTAT), is not recognizable by wild-type Cre but is readily recognized by a Cre mutant named C2(+/-)#4 (Santoro and Schultz, 2002). Furthermore, a hybrid of loxP and loxM7 by swapping a 13mer, termed loxP-M7 (or loxM7-P depending on which 13mer is exchanged), is bound by a heterodimer of wild-type Cre and C2(+/-)#4 mutant. The loxP-M7 is more efficiently recombined in the presence of both wild-type Cre and C2(+/-)#4 mutant than either one alone (Saraf-Levy et al., 2006). In a similar fashion, another group has developed Cre-FLP hybrid proteins, which were coined “Fre” and “Clp”, that can recognize loxP-FRT hybrid sequences and execute DNA recombination (Shaikh and Sadowski, 2000).

Theoretically, if the wild-type Cre and the C2(+/-)#4 mutant are expressed from two promoters respectively (e.g., Ngn3 and Myt1) and a reporter is constructed with the loxP-M7 instead of the wild-type loxP sites, this strategy could serve the same purpose of double-marker cell lineage tracing as our bipartite Cre system does. Hypothetically then, if we blend the heterotetramer with the bipartite Cre idea, i.e., splitting both the wild-type Cre and the C2(+/-)#4 mutant, we could possibly create a vast array of combinations, which could allow us to explore the heterogeneity of $Ngn3^{+}$ progenitors and increase the cellular resolution of lineage tracing from double-marked to triple- and quadruple-marked progenitors.

However, the biggest hurdle preventing us from applying this idea to cell lineage tracing is the specificity issue. The research of Cre mutation is initiated by the Schultz

group in an effort that will ultimately allow them to design customized Cre and edit mammalian genome (Santoro and Schultz, 2002), but this attempt is quickly overshadowed by the ZFN (zinc finger nuclease), TALEN (transcription activator-like effector nuclease) techniques and the emerging CRISPR/Cas9 technique due to the lack of specificity and versatility. Further mutation and screening in search of more specific Cre/loxP pair is necessary if we want to advance to a multipartite cell lineage tracing system but it could be difficult.

Attempt to sort Ngn3^{+}Myt1^{+} progenitors

To unveil the mechanisms of the biased cell fate determination, it is desirable to separate Ngn3^{+}Myt1^{+} cells from Ngn3^{+}Myt1^{-} cells and compare the differences between them. For this purpose, I have been trying to isolate the Ngn3^{+}Myt1^{+} and Ngn3^{+}Myt1^{-} subpopulations with two FACS strategies. My plan was to isolate these two cell populations and examine their transcription and epigenetic differences and correlate the differences with their endocrine fate. Yet overall, these methods did not work out for technical reasons.

The first method was to immunostain pancreatic cells directly with anti-Ngn3 and anti-Myt1 antibodies. However, this method was not successful mainly because the crude antibody serum I used was not suitable for FACS purpose. The Gu lab is now trying to purify the anti-Ngn3 and anti-Myt1 antibodies from the crude serum with antibody-antigen affinity-based purification method. Hopefully, the purified antibodies will give less background and allow me to sort out target cells with FACS. I followed a first-primary-antibody-then-secondary-antibody protocol, which complicated the process of antibody concentration titration for FACS. To solve this problem, I can chemically conjugate primary antibodies with fluorophores and then use these antibodies to stain cells in one step. It is doubtful whether this strategy will eventually help to overcome the

staining problem because cell surface marker is preferred in FACS assay. Immunostaining for nuclear transcription factors for FACS purpose is intrinsically difficult. Although different fixation regimes (4% paraformaldehyde or 1% formaldehyde, various length of time) were tested and various detergents (Triton X-100, Tween-20, Saponin) were tried, FACS results were not significantly improved.

The second method was to use endogenously produced fluorescent proteins as surrogate markers for Ngn3 and Myt1 (Fig 6.2 a). I crossed mice to get the *Ngn3*$^{nCre/EGFP}$*; Myt1*^{cCre}*; Ai9/+* embryos. After dissociation, pancreatic cells were stained with an anti-CD133 antibody and subjected to FACS. CD133 (prominin-1) is a cell surface maker for many types of stem cells and it is expressed on the apical surface of ductal epithelial cells in the pancreas (Sugiyama et al., 2007). CD133^{+}EGFP^{+} cells are Ngn3^{+} progenitor cells. CD133 staining is necessary in selecting Ngn3^{+} progenitors because EGFP does not only mark Ngn3^{+} progenitors due to its longevity. Out of the CD133^{+}EGFP^{+} cells, I reasoned that CD133^{+}EGFP^{+}tdTomato^{+} cells were expected to be Ngn3^{+}Myt1^{+} cells while CD133^{+}EGFP^{+}tdTomato^{-} cells represented Ngn3^{+}Myt1^{-} cells (Fig 6.2 b).

Based on my pilot study, CD133 staining gave a clean background and showed distinct negative and positive populations in FACS. I also obtained some target cell populations and executed qRT-PCR to validate the identity of these cell populations (Fig 6.3). Both CD133^{+}EGFP^{+}tdTomato^{+} and CD133^{+}EGFP^{+}tdTomato^{-} cells displayed higher *Ngn3* expression than the CD133^{-}EGFP^{-}tdTomato^{+} control cells, which were differentiated endocrine cells that lacked both *CD133* and *Ngn3*^{EGFP} expression. CD133^{+}EGFP^{+}tdTomato^{+} cells also had higher *Myt1* expression than the CD133^{+}EGFP^{+}tdTomato^{-} counterpart. In addition, both CD133^{+}EGFP^{+}tdTomato^{+} and CD133^{+}EGFP^{+}tdTomato^{-} cells lacked substantial *Ins1*, *Ins2*, or *Gcg* expression as compared to the CD133^{-}EGFP^{-}tdTomato^{+} control cells. In summary, qRT-PCR data implied that CD133^{+}EGFP^{+}tdTomato^{+} and CD133^{+}EGFP^{+}tdTomato^{-} cells were

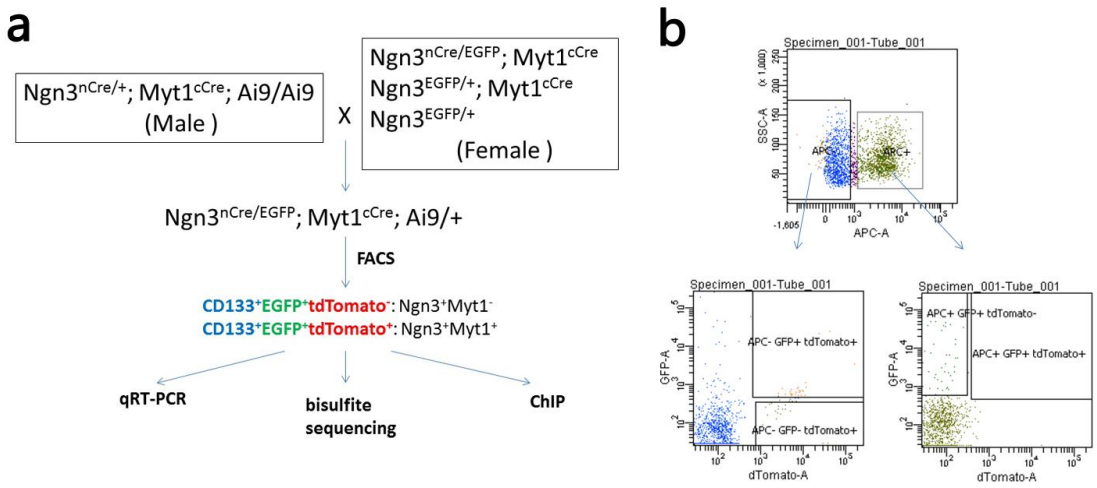


Fig 6.2 Experimental design for sorting Ngn3^{+}Myt1^{+} and Ngn3^{+}Myt1^{-} progenitors. (a) Mouse cross scheme was designed to maximize the chance of obtaining *Ngn3$^{nCre/EGFP}$;Myt1^{cCre};Ai9/+* embryos. E15.5 pancreata were dissected and dissociated into single cells. After staining with anti-CD133 antibody, cells are subject to FACS. CD133^{+}EGFP^{+}tdTomato^{-} and CD133^{+}EGFP^{+}tdTomato^{+} cells were collected. These two cell populations presumably represent Ngn3^{+}Myt1^{-} and Ngn3^{+}Myt1^{+} progenitors and their identity will be confirmed with qRT-PCR. CD133^{-}EGFP^{-}tdTomato^{+} cells were also collected as a differentiated endocrine cell control. Sorted cells were subject to downstream analysis, including qRT-PCR, bisulfite sequencing, and chromatin immunoprecipitation (ChIP) to interrogate the differences between Ngn3^{+}Myt1^{-} and Ngn3^{+}Myt1^{+} progenitors in their transcription profile, DNA methylation and histone modification. (b) One FACS result example showing the relative abundance of each cell population. There were very few CD133^{+}EGFP^{+}tdTomato^{+} cells harvested. (anti-CD133 antibody is APC-conjugated. CD133 and APC are used interchangeably in the description.)

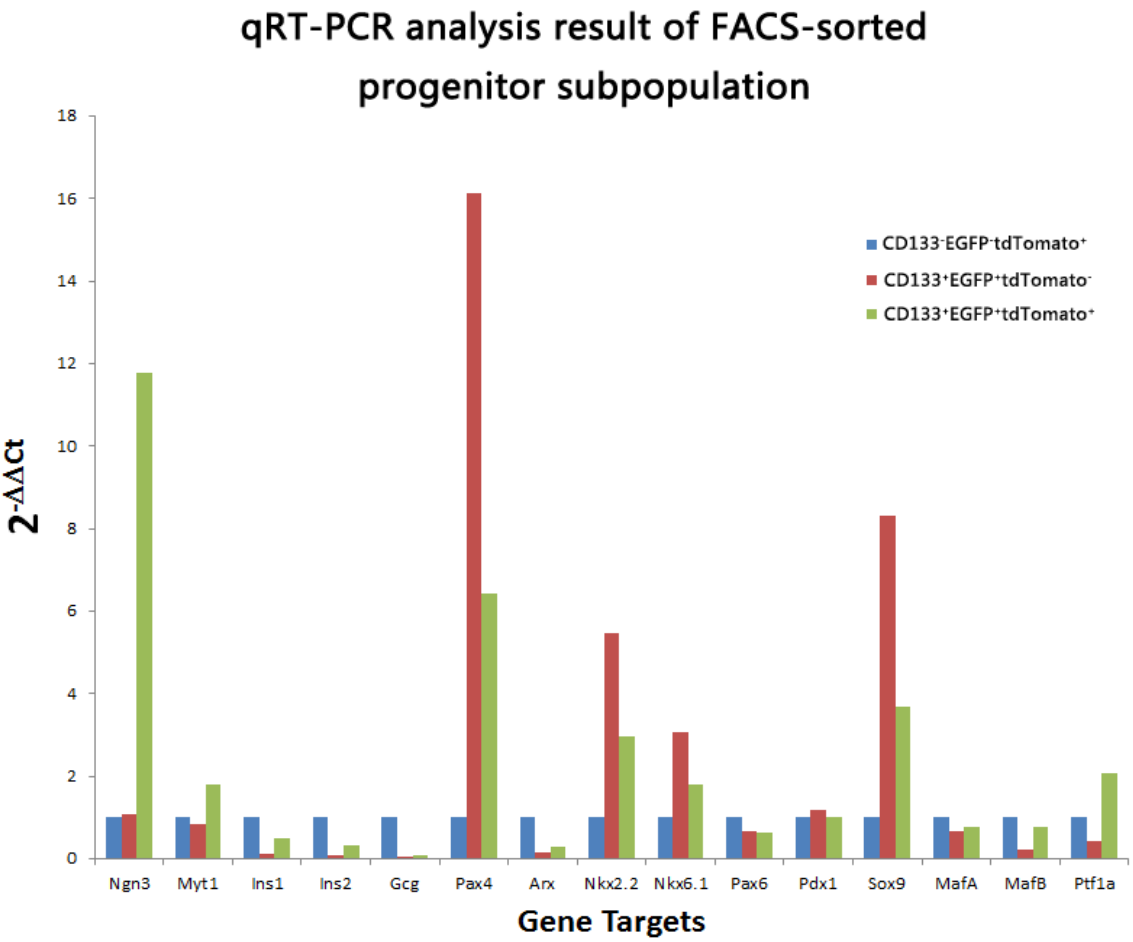


Fig 6.3 qRT-PCR analysis result of FACS-sorted progenitor subpopulation. Three cell populations were collected, the tdTomato single positive population, which are supposedly differentiated cells and serve as our normalization control, the $CD133^{+}EGFP^{+}tdTomato^{-}$ cells, and the $CD133^{+}EGFP^{+}tdTomato^{+}$ cells. The collected cell populations were each homogenized and their mRNA transcripts were quantified with qRT-PCR.

Ngn3^{+}Myt1^{+} and Ngn3^{+}Myt1^{-} progenitors, as I expected. Why CD133^{+}EGFP^{+}tdTomato^{-} cells would possess higher expression of *Pax4*, *Nkx2.2*, and *Nkx6.1* is difficult to interpret as CD133^{+}EGFP^{+}tdTomato^{+} cells are β-cell-prone progenitors based on our bipartite Cre cell lineage tracing (Fig 3.7).

The robustness of this qRT-PCR assay was dampened by the low quantity of cells I obtained from FACS. It was difficult to harvest enough CD133^{+}EGFP^{+}tdTomato^{+} cells for downstream assays and there were several possible explanations for this.

First, I examined the presence of CD133^{+}EGFP^{+}tdTomato^{+} cells. I examined the E14.5 pancreas tissue by immunostaining. Despite low image quality, I believed the existence of CD133^{+}EGFP^{+}tdTomato^{+} or Muc1^{+}EGFP^{+}tdTomato^{+} (Muc1 is another duct epithelium apical maker) cells in the tissue (data not shown). However, the number of CD133^{+}EGFP^{+}tdTomato^{+} cells is low, probably because CD133^{+}EGFP^{+}tdTomato^{+} is a very transient state. Once cells turn on tdTomato, they quickly delaminate from the duct epithelium and become CD133 negative.

Another explanation is that the *Ngn3^{nCre}* allele is hypomorphic. From the data in Chapter V, we knew that the *Ngn3$^{EGFP/+}$* pancreas had higher EGFP signal than the *Ngn3$^{EGFP/EGFP}$* pancreas, suggesting that *Ngn3* activates its own expression. This was bolstered by flow cytometry analysis, which showed both EGFP^{hi} and EGFP^{low} cell populations in the *Ngn3$^{EGFP/+}$* pancreas but mainly EGFP^{low} cell population in the *Ngn3$^{EGFP/EGFP}$* pancreas. Because of the feed-forward activation of *Ngn3*, I suspected that the *Ngn3^{nCre}* allele might be hypomorphic and thus decreased *Ngn3* and *EGFP* expression from the *Ngn3* promoter. When I genotyped embryos and examined their fluorescence intensity under the microscope, I found that almost all *Ngn3$^{EGFP/nCre}$* pancreata showed lower EGFP intensity than the *Ngn3$^{EGFP/+}$* pancreata, suggesting that the *Ngn3^{nCre}* allele was indeed hypomorphic. That could be another reason why I was

not able to obtain enough CD133^{+}EGFP^{+}tdTomato^{+} cells because EGFP^{low} cells were not picked up by the FACS cell sorter.

Why are Ngn3^{+}Myt1^{+} progenitors biased to β cells?

There are two possible mechanisms to explain the biased β-cell fate chosen by Ngn3^{+}Myt1^{+} progenitors. One mechanism is that Ngn3 and Myt1 cooperatively activate or enhance the expression level of a cascade of transcription factors that favor β-cell fate. To validate this hypothesis, we need to identify *Ngn3* and *Myt1* downstream targets, especially whether they share any common downstream targets. Some *Ngn3* downstream targets are reported, such as *Myt1*, *Nkx2.2*, *Nkx6.1*, *Pax4*, *Arx*, *Insm1*, and *Rfx6, etc.* (Collombat et al., 2003; Mellitzer et al., 2006; Smith et al., 2003; Smith et al., 2010; Soyer et al., 2010; Wang et al., 2008; Watada et al., 2003). Information regarding *Myt1* downstream targets is more limited though. To systematically discover *Ngn3* and *Myt1* downstream targets, ChIP-Seq assay will be invaluable. Meanwhile, it is useful to test whether Ngn3 and Myt1 proteins can physically interact and activate gene expression as a complex.

Because I showed that manipulating DNA methylation changed β-to-α cell ratio, a second and more plausible mechanism is that Ngn3 and Myt1 recruit epigenetic modifiers to specific genomic loci and alter the epigenetic landscape. It is reported that in β cells, Nkx2.2 forms a repressor complex on methylated *Arx* promoter by recruiting Hdac1, Grg3, and Dnmt3a and reinforces the repression of *Arx* by DNA methylation, preventing the acquisition of α-cell fate (Papizan et al., 2011). I speculate whether a similar mechanism exists for Ngn3 and Myt1. To this end, I learned that both Myt1 and Myt1L can bind to Sin3B, which in turn recruits Hdac1 and Hdac2 and represses a heterologous promoter (Romm et al., 2005). It is also reported that Myt1 physically interacts with lysine-specific demethylase 1(Lsd1) and represses the expression of *Pten*

(phosphatase and tensin homolog) gene (Yokoyama et al., 2014). Xenopus Neurogenin 1 (xNgn1), a member of the bHLH neurogenin family, is shown to recruit CREB/p300/PACF complex, all of which possess histone actyltransferase activity, and activate its downstream targets *xMyt1* and *xNeuroD* (Koyano-Nakagawa et al., 1999). Although these studies were done in systems other than mouse pancreas, they provide insightful information of how Ngn3 and Myt1 could potentially utilize the epigenetic machinery to distinguish $Ngn3^{+}Myt^{+}$ progenitors from the $Ngn3^{+}Myt1^{-}$ progenitors. Co-immunoprecipitation and mass spectrometry assays will aid the discovery of Ngn3/Myt1 binding partners, at least testing whether similar results can be reproduced in the embryonic pancreas.

What are the downstream targets of Ngn3 and Myt1?

As discussed above, another interesting question is identifying the direct downstream targets of *Myt1* and *Ngn3*. There are many occasional reports about the identification of *Myt1/Ngn3* downstream targets. For example, *Ngn3* controls the expression of *Arx* (Collombat et al., 2003), *Pax4* (Smith et al., 2003), *NeuroD* (Huang et al., 2000), *Myt1* (Wang et al., 2008), *Nkx2.2* (Watada et al., 2003), *Insm1* (Mellitzer et al., 2006; Osipovich et al., 2014), *Rfx6* (Smith et al., 2010; Soyer et al., 2010) *etc*. There are relatively less known targets of *Myt1*. One example is *Pten* (Yokoyama et al., 2014).

Despite that, there are caveats in the interpretation of these results. First, many results come from *in vitro* assays, such as gel shift or reporter assays. Whether these results truly recapitulate the *in vivo* situation is doubtful. Second, some reports use *Ngn3*-dependent expression as the evidence for being an *Ngn3* downstream target, but the criterion of *Ngn3*-dependent expression is insufficient to distinguish whether a gene is a direct or indirect target of *Ngn3*. In order to systematically discover *Ngn3* targets, one experiment used microarray-based method and compared the differential

expression between *Ngn3*$^{+/+}$ and *Ngn3*$^{-/-}$ developing pancreas (Petri et al., 2006). A similar but superior microarray was done by the Kaestner group by using an *Ngn3*^{EGFP} allele which allowed them to isolate EGFP^{+} progenitors instead of homogenized whole pancreas (White et al., 2008). Microarray facilitates the systematic and fast discovery of potential downstream targets but still it fails to discern the direct-target versus secondary-effect scenarios. Thus, it is desirable to design a ChIP-Seq assay to systematically discover the direct targets of *Myt1* and *Ngn3*. To this end, the Gu lab has been optimizing the ChIP assay protocol as well as comparing the efficiency and specificity of different antibodies. We are looking forward to the identification of direct downstream targets of *Myt1* and *Ngn3* in the future.

How are the heterogeneous co-expression pattern established?

Another interesting question is how the Ngn3^{+} progenitor heterogeneity is established. I have shown that *Myt1*, *Nkx2.2* and *Pax4* expression is diminished in the E14.5 *Ngn3*$^{EGFP/EGFP}$ pancreas, suggesting that these key endocrine transcription factors are controlled by *Ngn3*, which is consistent with previous reports (Smith et al., 2003; Wang et al., 2008; Watada et al., 2003). Given also that *Myt1* and *Ngn3* form a positive feed-back loop (Wang et al., 2008), it is difficult to understand how *Myt1* is only expressed in a portion of the Ngn3^{+} progenitors. So are *Nkx2.2* and *Pax4*, as they are also *Ngn3*-dependent.

Because *Ngn3* expression is transient while *Myt1* expression persists after it is turned on, we conjecture that *Myt1* is rapidly activated in some Ngn3^{+} progenitors but is delayed in other Ngn3^{+} progenitors, resulting in the heterogeneous co-expression pattern. But why does such a responsiveness difference exist? Why is *Myt1* (and *Nkx2.2*, *Pax4*, *etc*.) activated quicker in some Ngn3^{+} progenitors but slower in others? Is there an Ngn3 protein threshold to activate its downstream genes? Are all *Ngn3* downstream genes

equivalently sensitive to Ngn3 binding? Is the heterogeneity determined well before *Ngn3* expression initiates? These are all stimulating questions. Sampling developing pancreas at different time points and taking successive snapshots of the endocrine progenitors' transcriptional profile and epigenetic landscape will provide insightful information. In addition, single-cell mRNA sequencing and DNA methylation sequencing techniques are powerful tools to dissect cellular heterogeneity and they are also options worth considering (Shapiro et al., 2013; Smallwood et al., 2014).

Do Ngn3$^+$Myt$^+$ progenitors have the same differentiation potential across all embryonic stages?

Our bipartite Cre cell lineage tracing results show that Ngn3$^+$Myt1$^+$ progenitors preferentially give rise to β cells. We then asked whether ectopic overexpression of *Myt1* in the Ngn3$^+$ domain will promote the acquisition of β-cell fate. To address this question, Dr. Gu utilized an *Ngn3^{Myt1}* transgenic mouse model (data unpublished) to ectopically overexpress *Myt1* in the Ngn3$^+$ progenitors and analyzed the effect on endocrine lineage allocation. As expected, Ngn3$^+$ progenitors all turned into Ngn3$^+$Myt1$^+$ in the transgenic mouse pancreas and an increased β-to-α cell ratio was observed (data not shown), which is consistent with the bipartite Cre lineage tracing result that Ngn3$^+$Myt1$^+$ progenitors favor β cells over α cells.

The observed result that Ngn3$^+$Myt$^+$ progenitors favorably produced β cells only reflects the lineage tracing result of aggregated Ngn3$^+$Myt$^+$ progenitors across all embryonic stages but does not distinguish Ngn3$^+$Myt$^+$ progenitors from different stages. Because pancreatic endocrine cells are produced according to a competence window model (Johansson et al., 2007), whether Ngn3$^+$Myt$^+$ progenitors have the same differentiation potential at different embryonic stages is an intriguing question. On the other hand, since *Myt1* expression is not constant during the course of pancreatic

endocrine differentiation but increases from E10.5 to E18.5 (Fig 3.2), forced *Myt1* expression at a precocious stage might influence the differentiation potential of $Ngn3^+$ progenitors. Therefore, instead of constitutively overexpressing *Myt1* in $Ngn3^+$ progenitors, I would like to overexpress *Myt1* in specific time windows and examine how it could change the competence of $Ngn3^+$ progenitors. To this purpose, I can employ the TetON/OFF system to induce *Myt1* expression with doxycycline administration. The existing $Ngn3^{tTA/+}$ (Wang et al., 2009a), $Myt1b^{tet}$ (Wang et al., 2008), or $Ngn3^{Cre};Rosa26^{rtTA/+}$ (Belteki et al., 2005; Gu et al., 2002) mice could be crossed to obtain $Ngn3^{tTA/+};Myt1b^{tet}$ or $Ngn3^{Cre};Rosa26^{rtTA/+};Myt1b^{tet}$ mice and achieve temporally induced *Myt1* ectopic expression.

In addition, I would like to do a "pulse-chase" experiment in which only $Ngn3^+Myt1^+$ progenitors within a short time window are labeled and tracked. This "pulse-chase" experiment is feasible with a temporally inducible bipartite Cre technique. I have designed and created fusion proteins between bipartite Cre and an estrogen receptor, ERT2. I compared different constructs and identified one bipartite CreERT2 design that demonstrated desired characteristics for lineage tracing, including low background activity prior to induction, high recombination efficiency upon induction, tamoxifen dosage-dependent response, as well as easy immunostaining detection. The well-characterized inducible bipartite CreERT2 in Chapter IV is readily available to generate transgenic mouse models in the future and is able to answer the question of stage-specific differentiation capacity.

How are Dnmts recruited to specific loci?

In Chapter III, I examined the endocrine lineage allocation using a *Dnmt3b* overexpression model first created by Linhart et al. (Linhart et al., 2007). In the $Rosa26^{M2\text{-}rtTA/+};ColA1^{tetOP\text{-}Dnmt3b/+}$ mouse model, Linhart et al. found that *Dnmt3b*

overexpression does not cause global DNA methylation increase because no significant increase of 5'methylcytosine amount is observed. Instead, only selective loci show hypermethylation than the control group. Then how is Dnmt3b recruited to specific genomic loci? I speculate there are two possible mechanisms.

First, Dnmts could be recruited to specific loci through a large complex whose component recognizes the loci. To this end, I learned that Nkx2.2 forms a repressor complex on methylated *Arx* promoter by recruiting Hdac1, Grg3, and Dnmt3a, thus reinforcing the repression of *Arx* by DNA methylation and promoting β-cell fate (Papizan et al., 2011). It is reported that Myt1 and Myt1L could form a repressor complex with Sin3B, Hdac1 and Hdac2 (Romm et al., 2005), or with Lsd1 (Yokoyama et al., 2014), and that a close family member of Ngn3, Ngn1, interacts with histone actyltransferase CREB/p300/PACF (Koyano-Nakagawa et al., 1999). Thus, Dnmts could potentially be recruited by Ngn3 and Myt1 to specific targets through similar mechanisms to the above or even shared adaptor proteins. If such a physical interaction is proved and target loci are identified, it will explain why $Ngn3^{+}Myt1^{+}$ progenitors have β-cell biased differentiation preference and why *Dnmt3b* overexpression increases β-to-α ratio (Chapter III).

A second mechanism of Dnmt recruitment is through the recognition of histone markers or interaction with histone modification enzymes. Dnmt3L, a Dnmt-like protein without methyltransferase activity, recruits Dnmt3a and Dnmt3b to vacant H3 histone tail whereas H3K4 methylation inhibits the recruitment (Ooi et al., 2007). In addition, Dnmt3a can bind to protein arginine methyltransferase 5 (Prmt5)-mediated H4R3me2 site and subsequently methylates DNA in the nearby region (Zhao et al., 2009). Dnmt3a/b also physically interacts with various histone modifiers, including the PRC2 complex component histone methyltransferase Ezh2 (Vire et al., 2006), H3K9 methyltransferase G9a and Suv39h1(Epsztejn-Litman et al., 2008; Fuks et al., 2003) *etc*. All these results

suggest that histone methylation affects DNA methylation and could serve as “landing sites” for Dnmts. DNA methylation and histone modification are often co-dependent and the influence can be bi-directional (Cedar and Bergman, 2009; Rose and Klose, 2014). Specifically in the pancreas, Dnmt1-mediated methylation of *Arx* UR2 region leads to the binding of MeCP2, a methyl-DNA binding protein, to the *Arx* UR2 region, which in turn recruits Prmt6 to catalyze H3R2 methylation. H3R2me serves as a repressive marker and reinforces repression exerted by DNA methylation (Dhawan et al., 2011).

Can we deliberately recruit Dnmts to specific loci and thus manipulate DNA methylation and gene expression? For this purpose, the ZFN, TALEN and CRISPR/Cas9 techniques could be powerful tools. Besides genome editing, endonuclease activity-disabled ZFN/TALEN/Cas9 have been used to make synthesized transcription activator or repressor for directed gene expression regulation (Beerli et al., 2000; Cheng et al., 2013; Cong et al., 2012; Farzadfard et al., 2013; Gilbert et al., 2013; Kabadi et al., 2014; Konermann et al., 2013; Maeder et al., 2013; Mali et al., 2013; Miller et al., 2011; Perez-Pinera et al., 2013; Polstein and Gersbach, 2015; Qi et al., 2013; Zhang et al., 2011). In these synthesized transactivators or repressors, endonuclease-dead ZFN/TALEN/Cas9 are fused with effector proteins, such as VP16, VP64 (transactivator) or KRAB, ERD, SID (transrepressor) *etc*. But their DNA sequence-specific binding capacity is preserved, which allows them for targeted gene activation or repression. We could create fusion proteins between ZFN/TALEN/Cas9 and the epigenetic enzymes such as Dnmts, or *bona fide* transcription factors, such as Ngn3/Myt1, instead of generic effector proteins, and bring these effector proteins to specific loci. The simplicity of CRISPR/Cas9 system even allows us to recruit synthetic factors to multiple loci if two or more gRNAs are expressed simultaneously, making it a versatile platform for directed gene expression regulation.

What are the roles of miRNAs and gap junctions in pancreas development?

While our preliminary results support a non-canonical Notch lateral inhibition model in which neighboring cells communicate through the translocation of miRNAs across gap junctions and these miRNAs can enhance *Ngn3* expression by inhibiting Notch signaling, many informational gaps need to be filled to validate this model. For example, direct evidence showing miRNA translocation through gap junctions in developing pancreas is unavailable. If miRNAs do transverse plasma membrane, what type of connexins are involved? Is it an active transportation process or passive diffusion? What are the targets of these miRNAs? These questions are intriguing and could be addressed as outlined in Chapter V.

Conclusion remarks

This work investigates the pancreatic endocrine differentiation process. Through lineage tracing with an innovative bipartite Cre system, I found that combinatorial *Ngn3* and *Myt1* expression encourages the endocrine progenitors to adopt the β-cell differentiation pathway. It appears that epigenetics, especially DNA methylation, plays an important role in endocrine lineage allocation because the methylation of a *Myt1* regulatory element decreased from E10.5 to E15.5 and interfering with methylation with chemical inhibitors and *Dnmt3b* overexpression altered β-to-α cell ratio. Cellular variation of DNA methylation could be the reason behind the Ngn3/Myt1 heterogeneous co-expression pattern. Alternatively, Ngn3/Myt1 co-expression heterogeneity can cause the differential epigenetic marker deposition, leading to the divergence of endocrine cell specification. The involvement of other transcription factors, epigenetic modification enzymes, miRNAs, and lncRNAs in the process of endocrine lineage specification should also be examined. Sampling $Ngn3^{+}$ progenitors and $Ngn3^{+}$ progenitor subtypes

from successive embryonic stages and assessing their transcriptome and epigenome profile will elucidate many mysteries.

Understanding the pancreas development process not only is important from a basic research perspective but also has clinical relevance. *In vitro* β-cell reprogramming and transplantation is a promising therapy for diabetes. Delineating pancreatic organogenesis, especially the β-cell differentiation process, is central in designing optimal *in vitro* β-cell reprogramming protocol. I expect this work will provide useful information for the pancreas research community, as well as hope for diabetic patients.

REFERENCES

Ahren, B. (2009). Islet G protein-coupled receptors as potential targets for treatment of type 2 diabetes. Nature Reviews Drug Discovery *8*, 369-385.

Akiyama, H., Kim, J.E., Nakashima, K., Balmes, G., Iwai, N., Deng, J.M., Zhang, Z.P., Martin, J.F., Behringer, R.R., Nakamura, T.*, et al.* (2005). Osteo-chondroprogenitor cells are derived from Sox9 expressing precursors. Proceedings of the National Academy of Sciences of the United States of America *102*, 14665-14670.

Al-Hasani, K., Pfeifer, A., Courtney, M., Ben-Othman, N., Gjernes, E., Vieira, A., Druelle, N., Avolio, F., Ravassard, P., Leuckx, G.*, et al.* (2013). Adult Duct-Lining Cells Can Reprogram into beta-like Cells Able to Counter Repeated Cycles of Toxin-Induced Diabetes. Developmental Cell *26*, 86-100.

Anastassiadis, K., Fu, J., Patsch, C., Hu, S.B., Weidlich, S., Duerschke, K., Buchholz, F., Edenhofer, F., and Stewart, A.F. (2009). Dre recombinase, like Cre, is a highly efficient site-specific recombinase in E-coli, mammalian cells and mice. Disease Models & Mechanisms *2*, 508-515.

Apelqvist, A., Li, H., Sommer, L., Beatus, P., Anderson, D.J., Honjo, T., de Angelis, M.H., Lendahl, U., and Edlund, H. (1999). Notch signalling controls pancreatic cell differentiation. Nature *400*, 877-881.

Arnes, L., Hill, J.T., Gross, S., Magnuson, M.A., and Sussel, L. (2012). Ghrelin Expression in the Mouse Pancreas Defines a Unique Multipotent Progenitor Population. Plos One *7*.

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.

Ashcroft, F.M., and Rorsman, P. (2012). Diabetes Mellitus and the beta Cell: The Last Ten Years. Cell *148*, 1160-1171.

Atkinson, M.A. (2012). The Pathogenesis and Natural History of Type 1 Diabetes. Cold Spring Harbor Perspectives in Medicine *2*, 18.

Baeyens, L., Lemper, M., Leuckx, G., De Groef, S., Bonfanti, P., Stange, G., Shemer, R., Nord, C., Scheel, D.W., Pan, F.C.*, et al.* (2014). Transient cytokine treatment induces acinar cell reprogramming and regenerates functional beta cell mass in diabetic mice. Nature Biotechnology *32*, 76-+.

Barker, N., van Es, J.H., Kuipers, J., Kujala, P., van den Born, M., Cozijnsen, M., Haegebarth, A., Korving, J., Begthel, H., Peters, P.J.*, et al.* (2007). Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature *449*, 1003-U1001.

Basford, C.L., Prentice, K.J., Hardy, A.B., Sarangi, F., Micallef, S.J., Li, X., Guo, Q., Elefanty, A.G., Stanley, E.G., Keller, G.*, et al.* (2012). The functional and molecular characterisation of human embryonic stem cell-derived insulin-positive cells compared with adult pancreatic beta cells. Diabetologia *55*, 358-371.

Beckervordersandforth, R., Tripathi, P., Ninkovic, J., Bayam, E., Lepier, A., Stempfhuber, B., Kirchhoff, F., Hirrlinger, J., Haslinger, A., Lie, D.C.*, et al.* (2010). In Vivo Fate Mapping and Expression Analysis Reveals Molecular Hallmarks of Prospectively Isolated Adult Neural Stem Cells. Cell Stem Cell *7*, 744-758.

Beckman, J.A., Creager, M.A., and Libby, P. (2002). Diabetes and atherosclerosis - Epidemiology, pathophysiology, and management. Jama-Journal of the American Medical Association *287*, 2570-2581.

Beerli, R.R., Dreier, B., and Barbas, C.F. (2000). Positive and negative regulation of endogenous genes by designed transcription factors. Proceedings of the National Academy of Sciences of the United States of America *97*, 1495-1500.

Belteki, G., Haigh, J., Kabacs, N., Haigh, K., Sison, K., Costantini, F., Whitsett, J., Quaggin, S.E., and Nagy, A. (2005). Conditional and inducible transgene expression in mice through the combinatorial use of Cre-mediated recombination and tetracycline induction. Nucleic Acids Research *33*, 10.

Benninger, R.K.P., Head, W.S., Zhang, M., Satin, L.S., and Piston, D.W. (2011). Gap junctions and other mechanisms of cell-cell communication regulate basal insulin secretion in the pancreatic islet. Journal of Physiology-London *589*, 5453-5466.

Bernstein, E., Kim, S.Y., Carmell, M.A., Murchison, E.P., Alcorn, H., Li, M.Z., Mills, A.A., Elledge, S.J., Anderson, K.V., and Hannon, G.J. (2003). Dicer is essential for mouse development. Nature Genetics *35*, 215-217.

Bhushan, A., Itoh, N., Kato, S., Thiery, J.P., Czernichow, P., Bellusci, S., and Scharfmann, R. (2001). Fgf10 is essential for maintaining the proliferative capacity of epithelial progenitor cells during early pancreatic organogenesis. Development *128*, 5109-5117.

Blad, C.C., Tang, C., and Offermanns, S. (2012). G protein-coupled receptors for energy metabolites as new therapeutic targets. Nature Reviews Drug Discovery *11*, 603-619.

Bonner-Weir, S., and Weir, G.C. (2005). New sources of pancreatic beta-cells. Nature Biotechnology *23*, 857-861.

Bosco, D., Armanet, M., Morel, P., Niclauss, N., Sgroi, A., Muller, Y.D., Giovannoni, L., Parnaud, G., and Berney, T. (2010). Unique Arrangement of alpha- and beta-Cells in Human Islets of Langerhans. Diabetes *59*, 1202-1210.

Bramswig, N.C., Everett, L.J., Schug, J., Dorrell, C., Liu, C.Y., Luo, Y.P., Streeter, P.R., Naji, A., Grompe, M., and Kaestner, K.H. (2013). Epigenomic plasticity enables human pancreatic alpha to beta cell reprogramming. Journal of Clinical Investigation *123*, 1275-1284.

Branda, C.S., and Dymecki, S.M. (2004). Talking about a revolution: The impact of site-specific recombinases on genetic analyses in mice. Developmental Cell *6*, 7-28.

Brink, C., Chowdhury, K., and Gruss, P. (2001). Pax4 regulatory elements mediate beta cell specific expression in the pancreas. Mechanisms of Development *100*, 37-43.

Brink, C., and Gruss, P. (2003). DNA sequence motifs conserved in endocrine promoters are essential for Pax4 expression. Developmental Dynamics *228*, 617-622.

Brissova, M., Aamodt, K., Brahmachary, P., Prasad, N., Hong, J.-Y., Dai, C., Mellati, M., Shostak, A., Poffenberger, G., Aramandla, R.*, et al.* (2014). Islet Microenvironment, Modulated by Vascular Endothelial Growth Factor-A Signaling, Promotes beta Cell Regeneration. Cell Metabolism *19*, 498-511.

Bruin, J.E., Rezania, A., Xu, J., Narayan, K., Fox, J.K., O'Neil, J.J., and Kieffer, T.J. (2013). Maturation and function of human embryonic stem cell-derived pancreatic progenitors in macroencapsulation devices following transplant into mice. Diabetologia *56*, 1987-1998.

Burcelin, R., and Dejager, S. (2010). GLP-1: What is known, new and controversial in 2010? Diabetes & Metabolism *36*, 503-509.

Burcelin, R., Knauf, C., and Cani, P.D. (2008). Pancreatic alpha-cell dysfunction in diabetes. Diabetes & Metabolism *34*, S49-S55.

Caicedo, A. (2013). Paracrine and autocrine interactions in the human islet: More than meets the eye. Seminars in Cell & Developmental Biology *24*, 11-21.

Calabrese, A., Zhang, M., Serre-Beinier, W., Caton, D., Mas, C., Satin, L.S., and Meda, P. (2003). Connexin 36 controls synchronization of Ca2+ oscillations and insulin secretion in MIN6 cells. Diabetes *52*, 417-424.

Cedar, H., and Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. Nature Reviews Genetics *10*, 295-304.

Cepko, C.L., Austin, C.P., Yang, X.J., and Alexiades, M. (1996). Cell fate determination in the vertebrate retina. Proceedings of the National Academy of Sciences of the United States of America *93*, 589-595.

Chandra, R., and Liddle, R.A. (2009). Neural and hormonal regulation of pancreatic secretion. Current Opinion in Gastroenterology *25*, 441-446.

Chen, H., Gu, X., Su, I.h., Bottino, R., Contreras, J.L., Tarakhovsky, A., and Kim, S.K. (2009). Polycomb protein Ezh2 regulates pancreatic beta-cell Ink4a/Arf expression and regeneration in diabetes mellitus. Genes & Development *23*, 975-985.

Chen, S.C.Y., Stern, P., Guo, Z.Y., and Chen, J.Z. (2011). Expression of Multiple Artificial MicroRNAs from a Chicken miRNA126-Based Lentiviral Vector. Plos One *6*, 10.

Cheng, A.W., Wang, H., Yang, H., Shi, L., Katz, Y., Theunissen, T.W., Rangarajan, S., Shivalila, C.S., Dadon, D.B., and Jaenisch, R. (2013). Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. Cell Research *23*, 1163-1171.

Chiang, H.R., Schoenfeld, L.W., Ruby, J.G., Auyeung, V.C., Spies, N., Baek, D., Johnston, W.K., Russ, C., Luo, S., Babiarz, J.E.*, et al.* (2010). Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. Genes & Development *24*, 992-1009.

Choi, E., Kraus, M.R.C., Lemaire, L.A., Yoshimoto, M., Vemula, S., Potter, L.A., Manduchi, E., Stoeckert, C.J., Jr., Grapin-Botton, A., and Magnuson, M.A. (2012). Dual Lineage-Specific Expression of Sox17 During Mouse Embryogenesis. Stem Cells *30*, 2297-2308.

Christman, J.K. (2002). 5-Azacytidine and 5-aza-2 '-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. Oncogene *21*, 5483-5495.

Chung, K.-H., Hart, C.C., Al-Bassam, S., Avery, A., Taylor, J., Patel, P.D., Vojtek, A.B., and Turner, D.L. (2006). Polycistronic RNA polymerase II expression vectors for RNA interference based on BIC/miR-155. Nucleic Acids Research *34*.

Collins, E.C., Pannell, R., Simpson, E.M., Forster, A., and Rabbitts, T.H. (2000). Inter-chromosomal recombination of Mll and Af9 genes mediated by cre-loxP in mouse development. Embo Reports *1*, 127-132.

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., Hecksher-Sorensen, J., Krull, J., Berger, J., Riedel, D., Herrera, P.L., Serup, P., and Mansouri, A. (2007). Embryonic endocrine pancreas and mature beta cells acquire alpha and PP cell phenotypes upon Arx misexpression. Journal of Clinical Investigation *117*, 961-970.

Collombat, P., Hecksher-Sorensen, J., Serup, P., and Mansouri, A. (2006). Specifying pancreatic endocrine cell fates. Mechanisms of Development *123*, 501-512.

Collombat, P., Mansouri, A., Hecksher-Sorensen, J., Serup, P., Krull, J., Gradwohl, G., and Gruss, P. (2003). Opposing actions of Arx and Pax4 in endocrine pancreas development. Genes & Development *17*, 2591-2603.

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.

Cong, L., Zhou, R., Kuo, Y.-c., Cunniff, M., and Zhang, F. (2012). Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. Nature Communications *3*.

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 Genetics *9*, 18.

Crawford, L.A., Guney, M.A., Oh, Y.A., DeYoung, R.A., Valenzuela, D.M., Murphy, A.J., Yancopoulos, G.D., Lyons, K.M., Brigstock, D.R., Economides, A.*, et al.* (2009). Connective Tissue Growth Factor (CTGF) Inactivation Leads to Defects in Islet Cell Lineage Allocation and beta-Cell Proliferation during Embryogenesis. Molecular Endocrinology *23*, 324-336.

Criscimanna, A., Speicher, J.A., Houshmand, G., Shiota, C., Prasadan, K., Ji, B., Logsdon, C.D., Gittes, G.K., and Esni, F. (2011). Duct Cells Contribute to Regeneration of Endocrine and Acinar Cells Following Pancreatic Damage in Adult Mice. Gastroenterology *141*, 1451-U1447.

D'Alessio, D. (2011). The role of dysregulated glucagon secretion in type 2 diabetes. Diabetes Obesity & Metabolism *13*, 126-132.

D'Amour, K.A., Bang, A.G., Eliazer, S., Kelly, O.G., Agulnick, A.D., Smart, N.G., Moorman, M.A., Kroon, E., Carpenter, M.K., and Baetge, E.E. (2006). Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells. Nature Biotechnology *24*, 1392-1401.

Dai, C., Brissova, M., Hang, Y., Thompson, C., Poffenberger, G., Shostak, A., Chen, Z., Stein, R., and Powers, A.C. (2012). Islet-enriched gene expression and glucose-induced insulin secretion in human and mouse islets. Diabetologia *55*, 707-718.

Davidson, H.W. (2004). (Pro)Insulin processing: a historical perspective. Cell Biochem Biophys *40*, 143-158.

Del Prato, S., and Marchetti, P. (2004). Beta- and alpha-cell dysfunction in type 2 diabetes. Hormone and Metabolic Research *36*, 775-781.

Desgraz, R., Bonal, C., and Herrera, P.L. (2011). beta-Cell regeneration: the pancreatic intrinsic faculty. Trends in Endocrinology and Metabolism *22*, 34-43.

Dhawan, S., Georgia, S., Tschen, S.I., Fan, G.P., and Bhushan, A. (2011). Pancreatic beta Cell Identity Is Maintained by DNA Methylation-Mediated Repression of Arx. Developmental Cell *20*, 419-429.

Donnelly, M.L.L., Hughes, L.E., Luke, G., Mendoza, H., ten Dam, E., Gani, D., and Ryan, M.D. (2001). The 'cleavage' activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring '2A-like' sequences. Journal of General Virology *82*, 1027-1041.

Dor, Y., Brown, J., Martinez, O.I., and Melton, D.A. (2004). Adult pancreatic beta-cells are formed by self-duplication rather than stem-cell differentiation. Nature *429*, 41-46.

Doyle, M.J., and Sussel, L. (2007). Nkx2.2 regulates beta-cell function in the mature islet. Diabetes *56*, 1999-2007.

Dunn, M.F. (2005). Zinc-ligand interactions modulate assembly and stability of the insulin hexamer - a review. Biometals *18*, 295-303.

Dzierzak, E., and Speck, N.A. (2008). Of lineage and legacy: the development of mammalian hematopoietic stem cells. Nature Immunology *9*, 129-136.

Ebert, M.S., Neilson, J.R., and Sharp, P.A. (2007). MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. Nature Methods *4*, 721-726.

Ebert, M.S., and Sharp, P.A. (2010). MicroRNA sponges: Progress and possibilities. Rna-a Publication of the Rna Society *16*, 2043-2050.

Edlund, H. (2001). Factors controlling pancreatic cell differentiation and function. Diabetologia *44*, 1071-1079.

Edlund, H. (2002). Pancreatic organogenesis - Developmental mechanisms and implications for therapy. Nature Reviews Genetics *3*, 524-532.

Efrat, S. (2013). Recent progress in generation of human surrogate beta cells. Current Opinion in Endocrinology Diabetes and Obesity *20*, 259-264.

Elliott, A.D., Ustione, A., and Piston, D.W. (2015). Somatostatin and insulin mediate glucose-inhibited glucagon secretion in the pancreatic alpha-cell by lowering cAMP. American Journal of Physiology-Endocrinology and Metabolism *308*, E130-E143.

Eloff, B.C., Lerner, D.L., Yamada, K.A., Schuessler, R.B., Saffitz, J.E., and Rosenbaum, D.S. (2001). High resolution optical mapping reveals conduction slowing in connexin43 deficient mice. Cardiovascular Research *51*, 681-690.

Epsztejn-Litman, S., Feldman, N., Abu-Remaileh, M., Shufaro, Y., Gerson, A., Ueda, J., Deplus, R., Fuks, F., Shinkai, Y., Cedar, H.*, et al.* (2008). De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. Nature Structural & Molecular Biology *15*, 1176-1183.

Espes, D., Lau, J., and Carlsson, P.-O. (2014). Increased circulating levels of betatrophin in individuals with long-standing type 1 diabetes. Diabetologia *57*, 50-53.

Farzadfard, F., Perli, S.D., and Lu, T.K. (2013). Tunable and Multifunctional Eukaryotic Transcription Factors Based on CRISPR/Cas. Acs Synthetic Biology *2*, 604-613.

Feil, R., Brocard, J., Mascrez, B., LeMeur, M., Metzger, D., and Chambon, P. (1996). Ligand-activated site-specific recombination in mice. Proceedings of the National Academy of Sciences of the United States of America *93*, 10887-10890.

Feil, R., Wagner, J., Metzger, D., and Chambon, P. (1997). Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains. Biochemical and Biophysical Research Communications *237*, 752-757.

Forman, J.J., Legesse-Miller, A., and Coller, H.A. (2008). A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence. Proceedings of the National Academy of Sciences of the United States of America *105*, 14879-14884.

Friedrich, G., and Soriano, P. (1991). Promoter traps in embryonic stem-cells - a genetic screen to identify and mutate developmental genes in mice. Genes & Development *5*, 1513-1523.

Fu, Z., Gilbert, E.R., and Liu, D. (2013). Regulation of insulin synthesis and secretion and pancreatic Beta-cell dysfunction in diabetes. Curr Diabetes Rev *9*, 25-53.

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.E. (2006). Targeted deletion of a cis-regulatory region reveals differential gene dosage requirements for Pdx1 in foregut organ differentiation and pancreas formation. Genes & Development *20*, 253-266.

Fuks, F., Hurd, P.J., Deplus, R., and Kouzarides, T. (2003). The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. Nucleic Acids Research *31*, 2305-2312.

Furuyama, K., Kawaguchi, Y., Akiyama, H., Horiguchi, M., Kodama, S., Kuhara, T., Hosokawa, S., Elbahrawy, A., Soeda, T., Koizumi, M.*, et al.* (2011). Continuous cell supply from a Sox9-expressing progenitor zone in adult liver, exocrine pancreas and intestine. Nature Genetics *43*, 34-U52.

Gannon, M., Ables, E.T., Crawford, L., Lowe, D., Offield, M.F., Magnuson, M.A., and Wright, C.V.E. (2008). pdx-1 function is specifically required in embryonic cells to generate appropriate numbers of endocrine cell types and maintain glucose homeostasis. Developmental Biology *314*, 406-417.

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 Metabolism *19*, 259-271.

Gelato, K.A., Martin, S.S., Liu, P.H., Saunders, A.A., and Baldwin, E.P. (2008). Spatially directed assembly of a heterotetrameric Cre-Lox synapse restricts recombination specificity. Journal of Molecular Biology *378*, 653-665.

Gerdes, M.J., Sevinsky, C.J., Sood, A., Adak, S., Bello, M.O., Bordwell, A., Can, A., Corwin, A., Dinn, S., Filkins, R.J.*, et al.* (2013). Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. Proceedings of the

National Academy of Sciences of the United States of America *110*, 11982-11987.

German, M.S. (2013). Anonymous sources: where do adult beta cells come from? Journal of Clinical Investigation *123*, 1936-1938.

Gibb, B., Gupta, K., Ghosh, K., Sharp, R., Chen, J., and Van Duyne, G.D. (2010). Requirements for catalysis in the Cre recombinase active site. Nucleic Acids Research *38*, 5817-5832.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A.*, et al.* (2013). CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. Cell *154*, 442-451.

Gittes, G.K. (2009). Developmental biology of the pancreas: A comprehensive review. Developmental Biology *326*, 4-35.

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. Proceedings of the National Academy of Sciences of the United States of America *97*, 1607-1611.

Grundy, S.M., Cleeman, J.I., Daniels, S.R., Donato, K.A., Eckel, R.H., Franklin, B.A., Gordon, D.J., Krauss, R.M., Savage, P.J., Smith, S.C.*, et al.* (2005). Diagnosis and management of the metabolic syndrome - An American Heart Association/National Heart, Lung, and Blood Institute Scientific Statement. Circulation *112*, 2735-2752.

Gu, G.Q., Brown, J.R., and Melton, D.A. (2003). Direct lineage tracing reveals the ontogeny of pancreatic cell fates during mouse embryogenesis. Mechanisms of Development *120*, 35-43.

Gu, G.Q., 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.

Guldenagel, M., Ammermuller, J., Feigenspan, A., Teubner, B., Degen, J., Sohl, G., Willecke, K., and Weiler, R. (2001). Visual transmission deficits in mice with targeted disruption of the gap junction gene connexin36. Journal of Neuroscience *21*, 6036-6044.

Gunawardana, S.C., and Piston, D.W. (2012). Reversal of Type 1 Diabetes in Mice by Brown Adipose Tissue Transplant. Diabetes *61*, 674-682.

Gunawardana, S.C., and Piston, D.W. (2015). Insulin-independent reversal of type 1 diabetes in nonobese diabetic mice with brown adipose tissue transplant. American Journal of Physiology-Endocrinology and Metabolism *308*, E1043-E1055.

Guo, F., Gopaul, D.N., and VanDuyne, G.D. (1997). Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. Nature *389*, 40-46.

Hafner, M., Wenk, J., Nenci, A., Pasparakis, M., Scharffetter-Kochanek, K., Smyth, N., Peters, T., Kess, D., Holtkotter, O., Shephard, P.*, et al.* (2004). Keratin 14 Cre transgenic mice authenticate keratin 14 as an oocyte-expressed protein. Genesis *38*, 176-181.

Hale, M.A., Kagami, H., Shi, L., Holland, A.M., Elsasser, H.P., Hammer, R.E., and MacDonald, R.J. (2005). The homeodomain protein PDX1 is required at mid-pancreatic development for the formation of the exocrine pancreas. Developmental Biology *286*, 225-237.

Hameyer, D., Loonstra, A., Eshkind, L., Schmitt, S., Antunes, C., Groen, A., Bindels, E., Jonkers, J., Krimpenfort, P., Meuwissen, R.*, et al.* (2007). Toxicity of ligand-dependent Cre recombinases and generation of a conditional Cre deleter mouse allowing mosaic recombination in peripheral tissues. Physiological Genomics *31*, 32-41.

Harfe, B.D., McManus, M.T., Mansfield, J.H., Hornstein, E., and Tabin, C.J. (2005). The RNaseIII enzyme Dicer is required for morphogenesis but not patterning of the vertebrate limb. Proceedings of the National Academy of Sciences of the United States of America *102*, 10898-10903.

Hayashi, S., and McMahon, A.P. (2002). Efficient recombination in diverse tissues by a tamoxifen-inducible form of Cre: A tool for temporally regulated gene activation/inactivation in the mouse. Developmental Biology *244*, 305-318.

Heckman, K.L., and Pease, L.R. (2007). Gene splicing and mutagenesis by PCR-driven overlap extension. Nature Protocols *2*, 924-932.

Heffner, C.S., Pratt, C.H., Babiuk, R.P., Sharma, Y., Rockwood, S.F., Donahue, L.R., Eppig, J.T., and Murray, S.A. (2012). Supporting conditional mouse mutagenesis with a comprehensive cre characterization resource. Nature Communications *3*.

Helin, K., and Dhanak, D. (2013). Chromatin proteins and modifications as drug targets. Nature *502*, 480-488.

Heremans, Y., Van De Casteele, M., Veld, P.I., 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. Journal of Cell Biology *159*, 303-311.

Herrera, P.L., Orci, L., and Vassalli, J.D. (1998). Two transgenic approaches to define the cell lineages in endocrine pancreas development. Molecular and Cellular Endocrinology *140*, 45-50.

Hirrlinger, J., Requardt, R.P., Winkler, U., Wilhelm, F., Schulze, C., and Hirrlinger, P.G. (2009a). Split-CreERT2: Temporal Control of DNA Recombination Mediated by Split-Cre Protein Fragment Complementation. Plos One *4*, 8.

Hirrlinger, J., Scheller, A., Hirrlinger, P.G., Kellert, B., Tang, W.N., Wehr, M.C., Goebbels, S., Reichenbach, A., Sprengel, R., Rossner, M.J.*, et al.* (2009b). Split-Cre Complementation Indicates Coincident Activity of Different Genes In Vivo. Plos One *4*, 10.

Horwich, M.D., and Zamore, P.D. (2008). Design and delivery of antisense oligonucleotides to block microRNA function in cultured Drosophila and human cells. Nature Protocols *3*, 1537-1549.

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. Molecular and Cellular Biology *20*, 3292-3307.

Hudson, L.D., Romm, E., Berndt, J.A., and Nielsen, J.A. (2011). A tool for examining the role of the zinc finger myelin transcription factor 1 (Myt1) in neural development: Myt1 knock-in mice. Transgenic Research *20*, 951-961.

Iber, D., and Menshykau, D. (2013). The control of branching morphogenesis. Open Biology *3*.

Ichise, H., Ichise, T., Sasanuma, H., and Yoshida, N. (2014). The Cd6 Gene as a Permissive Locus for Targeted Transgenesis in the Mouse. Genesis *52*, 440-450.

Imayoshi, I., Hirano, K., Sakamoto, M., Miyoshi, G., Imura, T., Kitano, S., Miyachi, H., and Kageyama, R. (2012). A multifunctional teal-fluorescent Rosa26 reporter mouse line for Cre- and Flp-mediated recombination. Neuroscience Research *73*, 85-91.

Indra, A.K., Warot, X., Brocard, J., Bornert, J.M., Xiao, J.H., Chambon, P., and Metzger, D. (1999). Temporally-controlled site-specific mutagenesis in the basal layer of the epidermis: comparison of the recombinase activity of the tamoxifen-inducible Cre-ERT and Cre-ERT2 recombinases. Nucleic Acids Research *27*, 4324-4327.

Ivanova, A., Signore, M., Caro, N., Greene, N.D.E., Copp, A.J., and Martinez-Barbera, J.P. (2005). In vivo genetic ablation by Cre-mediated expression of diphtheria toxin fragment A. Genesis *43*, 129-135.

Iype, T., Taylor, D.G., Ziesmann, S.M., Garmey, J.C., Watada, H., and Mirmira, R.G. (2004). The transcriptional repressor Nkx6.1 also functions as a deoxyribonucleic acid context-dependent transcriptional activator during pancreatic beta-cell differentiation: Evidence for feedback activation of the nkx6.1 gene by Nkx6.1. Molecular Endocrinology *18*, 1363-1375.

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. Molecular and Cellular Biology *20*, 4445-4454.

Jensen, J. (2004). Gene regulatory factors in pancreatic development. Developmental Dynamics *229*, 176-200.

Jensen, J., Pedersen, E.E., Galante, P., Hald, J., Heller, R.S., Ishibashi, M., Kageyama, R., Guillemot, F., Serup, P., and Madsen, O.D. (2000). Control sf endodermal endocrine development by Hes-1. Nature Genetics *24*, 36-44.

Jiao, Y., Le Lay, J., Yu, M., Naji, A., and Kaestner, K.H. (2014). Elevated Mouse Hepatic Betatrophin Expression Does Not Increase Human beta-Cell Replication in the Transplant Setting. Diabetes *63*, 1283-1288.

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. Developmental Cell *12*, 457-465.

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. Endocrine Reviews *28*, 685-705.

Kabadi, A.M., Ousterout, D.G., Hilton, I.B., and Gersbach, C.A. (2014). Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector. Nucleic Acids Research *42*.

Kawaguchi, Y., Cooper, B., Gannon, M., Ray, M., MacDonald, R.J., and Wright, C.V.E. (2002). The role of the transcriptional regulator Ptf1a in converting intestinal to pancreatic progenitors. Nature Genetics *32*, 128-134.

Kellendonk, C., Tronche, F., Monaghan, A.P., Angrand, P.O., Stewart, F., and Schutz, G. (1996). Regulation of cre recombinase activity by the synthetic steroid RU 486. Nucleic Acids Research *24*, 1404-1411.

Kim, W., Shin, Y.K., Kim, B.J., and Egan, J.M. (2010). Notch signaling in pancreatic endocrine cell and diabetes. Biochemical and Biophysical Research Communications *392*, 247-251.

Klee, P., Allagnat, F., Pontes, H., Cederroth, M., Charollais, A., Caille, D., Britan, A., Haefliger, J.A., and Meda, P. (2011). Connexins protect mouse pancreatic beta cells against apoptosis. Journal of Clinical Investigation *121*, 4870-4879.

Kloosterman, W.P., Lagendijk, A.K., Ketting, R.F., Moulton, J.D., and Plasterk, R.H.A. (2007). Targeted inhibition of miRNA maturation with morpholinos reveals a role for miR-375 in pancreatic islet development. Plos Biology *5*, 1738-1749.

Kluiver, J., Gibcus, J.H., Hettinga, C., Adema, A., Richter, M.K.S., Halsema, N., Slezak-Prochazka, I., Ding, Y., Kroesen, B.J., and van den Berg, A. (2012). Rapid Generation of MicroRNA Sponges for MicroRNA Inhibition. Plos One *7*, 8.

Knip, M., Veijola, W., Virtanen, S.M., Hyoty, H., Vaarala, O., and Akerblom, H.K. (2005). Environmental triggers and determinants of type 1 diabetes. Diabetes *54*, S125-S136.

Koch, M., and Horvath, T.L. (2014). Molecular and cellular regulation of hypothalamic melanocortin neurons controlling food intake and energy metabolism. Molecular Psychiatry *19*, 752-761.

Koike, H., Horie, K., Fukuyama, H., Kondoh, G., Nagata, S., and Takeda, J. (2002). Efficient biallelic mutagenesis with Cre/loxP-mediated inter-chromosomal recombination. Embo Reports *3*, 433-437.

Konermann, S., Brigham, M.D., Trevino, A.E., Hsu, P.D., Heidenreich, M., Cong, L., Platt, R.J., Scott, D.A., Church, G.M., and Zhang, F. (2013). Optical control of mammalian endogenous transcription and epigenetic states. Nature *500*, 472-+.

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.

Kordowich, S., Collombat, P., Mansouri, A., and Serup, P. (2011). Arx and Nkx2.2 compound deficiency redirects pancreatic alpha- and beta-cell differentiation to a somatostatin/ghrelin co-expressing cell lineage. Bmc Developmental Biology *11*.

Koyano-Nakagawa, N., Wettstein, D., and Kintner, C. (1999). Activation of Xenopus genes required for lateral inhibition and neuronal differentiation during primary neurogenesis. Molecular and Cellular Neuroscience *14*, 327-339.

Kretzschmar, K., and Watt, F.M. (2012). Lineage Tracing. Cell *148*, 33-45.

Kroon, E., Martinson, L.A., Kadoya, K., Bang, A.G., Kelly, O.G., Eliazer, S., Young, H., Richardson, M., Smart, N.G., Cunningham, J.*, et al.* (2008). Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells in vivo. Nature Biotechnology *26*, 443-452.

Krutzfeldt, J., Kuwajima, S., Braich, R., Rajeev, K.G., Pena, J., Tuschl, T., Manoharan, M., and Stoffel, M. (2007). Specificity, duplex degradation and subcellular localization of antagomirs. Nucleic Acids Research *35*, 2885-2892.

Krutzfeldt, J., Rajewsky, N., Braich, R., Rajeev, K.G., Tuschl, T., Manoharan, M., and Stoffel, M. (2005). Silencing of microRNAs in vivo with 'antagomirs'. Nature *438*, 685-689.

Kubicek, S., O'Sullivan, R.J., August, E.M., Hickey, E.R., Zhan, Q., Teodoro, M.L., Rea, S., Mechtler, K., Kowalski, J.A., Homon, C.A.*, et al.* (2007). Reversal of H3K9me2 by a small-molecule inhibitor for the G9a histone methyltransferase. Molecular Cell *25*, 473-481.

Kuhn, D.E., Martin, M.M., Feldman, D.S., Terry, A.V., Jr., Nuovo, G.J., and Elton, T.S. (2008). Experimental validation of miRNA targets. Methods *44*, 47-54.

Kumai, M., Nishii, K., Nakamura, K., Takeda, N., Suzuki, M., and Shibata, Y. (2000). Loss of connexin45 causes a cushion defect in early cardiogenesis. Development *127*, 3501-3512.

Lao, Z.M., Raju, G.P., Bai, C.B., and Joyner, A.L. (2012). MASTR: A Technique for Mosaic Mutant Analysis with Spatial and Temporal Control of Recombination Using Conditional Floxed Alleles in Mice. Cell Reports *2*, 386-396.

Lee, C.S., Perreault, N., Brestelli, J.E., and Kaestner, K.H. (2002). Neurogenin 3 is essential for the proper specification of gastric enteroendocrine cells and the maintenance of gastric epithelial cell identity. Genes & Development *16*, 1488-1497.

Lee, J., Sugiyama, T., Liu, Y., Wang, J., Gu, X., Lei, J., Markmann, J.F., Miyazaki, S., Miyazaki, J.-i., Szot, G.L.*, et al.* (2013). Expansion and conversion of human pancreatic ductal cells into insulin-secreting endocrine cells. Elife *2*.

Lee, J.C., Smith, S.B., Watada, H., Lin, J., Scheel, D., Wang, J.H., Mirmira, R.G., and German, M.S. (2001). Regulation of the pancreatic pro-endocrine gene neurogenin3. Diabetes *50*, 928-936.

Lee, J.Y., Ristow, M., Lin, X.Y., White, M.F., Magnuson, M.A., and Hennighausen, L. (2006). RIP-Cre revisited, evidence for impairments of pancreatic beta-cell function. Journal of Biological Chemistry *281*, 2649-2653.

Lenoir, O., Flosseau, K., Ma, F.X., Blondeau, B., Mai, A., Bassel-Duby, R., Ravassard, P., Olson, E.N., Haumaitre, C., and Scharfmann, R. (2011). Specific Control of Pancreatic Endocrine beta- and delta-Cell Mass by Class IIa Histone Deacetylases HDAC4, HDAC5, and HDAC9. Diabetes *60*, 2861-2871.

Lewandoski, M. (2001). Conditional control of gene expression in the mouse. Nature Reviews Genetics *2*, 743-755.

Li, K., Zhu, S., Russ, H.A., Xu, S., Xu, T., Zhang, Y., Ma, T., Hebrok, M., and Ding, S. (2014a). Small Molecules Facilitate the Reprogramming of Mouse Fibroblasts into Pancreatic Lineages. Cell Stem Cell *14*, 228-236.

Li, L.C., and Dahiya, R. (2002). MethPrimer: designing primers for methylation PCRs. Bioinformatics *18*, 1427-1431.

Li, W., Nakanishi, M., Zumsteg, A., Shear, M., Wright, C., Melton, D.A., and Zhou, Q. (2014b). In vivo reprogramming of pancreatic acinar cells to three islet endocrine subtypes. Elife *3*.

Liao, Y., Day, K.H., Damon, D.N., and Duling, B.R. (2001). Endothelial cell-specific knockout of connexin 43 causes hypotension and bradycardia in mice. Proceedings of the National Academy of Sciences of the United States of America *98*, 9989-9994.

Linhart, H.G., Lin, H., Yamada, Y., Moran, E., Steine, E.J., Gokhale, S., Lo, G., Cantu, E., Ehrich, M., He, T.*, et al.* (2007). Dnmt3b promotes tumorigenesis in vivo by gene-specific de novo methylation and transcriptional silencing. Genes & Development *21*, 3110-3122.

Liu, J., Willet, S.G., Bankaitis, E.D., Xu, Y.W., Wright, C.V.E., and Gu, G.Q. (2013). Non-parallel recombination limits cre-loxP-based reporters as precise indicators of conditional genetic manipulation. Genesis *51*, 436-442.

Livesey, F.J., and Cepko, C.L. (2001). Vertebrate neural cell-fate determination: Lessons from the retina. Nature Reviews Neuroscience *2*, 109-118.

Livet, J., Weissman, T.A., Kang, H.N., Draft, R.W., Lu, J., Bennis, R.A., Sanes, J.R., and Lichtman, J.W. (2007). Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. Nature *450*, 56-+.

Lobe, C.G., Koop, K.E., Kreppner, W., Lomeli, H., Gertsenstein, M., and Nagy, A. (1999). Z/AP, a double reporter for Cre-mediated recombination. Developmental Biology *208*, 281-292.

Lobocka, M.B., Rose, D.J., Plunkett, G., Rusin, M., Samojedny, A., Lehnherr, H., Yarmolinsky, M.B., and Blattner, F.R. (2004). Genome of bacteriophage P1. Journal of Bacteriology *186*, 7032-7068.

Long, M.A., and Rossi, F.M.V. (2009). Silencing Inhibits Cre-Mediated Recombination of the Z/AP and Z/EG Reporters in Adult Cells. Plos One *4*, 8.

Lynn, F.C., Skewes-Cox, P., Kosaka, Y., McManus, M.T., Harfe, B.D., and German, M.S. (2007a). MicroRNA expression is required for pancreatic islet cell genesis in the mouse. Diabetes *56*, 2938-2945.

Lynn, F.C., Smith, S.B., Wilson, M.E., Yang, K.Y., Nekrep, N., and German, M.S. (2007b). Sox9 coordinates a transcriptional network in pancreatic progenitor cells. Proceedings of the National Academy of Sciences of the United States of America *104*, 10500-10505.

Madisen, L., Zwingman, T.A., Sunkin, S.M., Oh, S.W., Zariwala, H.A., Gu, H., Ng, L.L., Palmiter, R.D., Hawrylycz, M.J., Jones, A.R.*, et al.* (2010). A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. Nature Neuroscience *13*, 133-U311.

Maeder, M.L., Linder, S.J., Cascio, V.M., Fu, Y., Ho, Q.H., and Joung, J.K. (2013). CRISPR RNA-guided activation of endogenous human genes. Nature Methods *10*, 977-+.

Magenheim, J., Klein, A.M., Stanger, B.Z., Ashery-Padan, R., Sosa-Pineda, B., Gu, G., and Dor, Y. (2011). Ngn(3+) endocrine progenitor cells control the fate and morphogenesis of pancreatic ductal epithelium. Developmental Biology *359*, 26-36.

Magnuson, M.A., and Osipovich, A.B. (2013). Pancreas-Specific Cre Driver Lines and Considerations for Their Prudent Use. Cell Metabolism *18*, 9-20.

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., and Church, G.M. (2013). CAS9 transcriptional activators for target specificity

screening and paired nickases for cooperative genome engineering. Nature Biotechnology *31*, 833-+.

Mao, X.H., Fujiwara, Y., Chapdelaine, A., Yang, H.D., and Orkin, S.H. (2001). Activation of EGFP expression by Cre-mediated excision in a new ROSA26 reporter mouse strain. Blood *97*, 324-326.

Mastracci, T.L., Wilcox, C.L., Arnes, L., Panea, C., Golden, J.A., May, C.L., and Sussel, L. (2011). Nkx2.2 and Arx genetically interact to regulate pancreatic endocrine cell development and endocrine hormone expression. Developmental Biology *359*, 1-11.

McGrath, P.S., Watson, C.L., Ingram, C., Helmrath, M.A., and Wells, J.M. (2015). The Basic Helix-Loop-Helix Transcription Factor NEUROG3 Is Required for Development of the Human Endocrine Pancreas. Diabetes *64*, 2497-2505.

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 Journal *25*, 1344-1352.

Merani, S., and Shapiro, A.M.J. (2006). Current status of pancreatic islet transplantation. Clinical Science *110*, 611-625.

Metzger, D., Clifford, J., Chiba, H., and Chambon, P. (1995). Conditional site-specific recombination in mammalian-cells using a ligand-dependent chimeric Cre recombinase. Proceedings of the National Academy of Sciences of the United States of America *92*, 6991-6995.

Mighiu, P.I., Yue, J.T.Y., Filippi, B.M., Abraham, M.A., Chari, M., Lam, C.K.L., Yang, C.S., Christian, N.R., Charron, M.J., and Lam, T.K.T. (2013). Hypothalamic glucagon signaling inhibits hepatic glucose production. Nature Medicine *19*, 766-+.

Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J.*, et al.* (2011). A TALE nuclease architecture for efficient genome editing. Nature Biotechnology *29*, 143-U149.

Mojibian, M., Lam, A.W.Y., Fujita, Y., Asadi, A., Grassl, G.A., Dickie, P., Tan, R.S., Cheung, A.T., and Kieffer, T.J. (2014). Insulin-Producing Intestinal K Cells Protect Nonobese Diabetic Mice From Autoimmune Diabetes. Gastroenterology *147*, 162-U275.

Moller, D.E. (2001). New drug targets for type 2 diabetes and the metabolic syndrome. Nature *414*, 821-827.

Moore, L.D., Le, T., and Fan, G. (2013). DNA Methylation and Its Basic Function. Neuropsychopharmacology *38*, 23-38.

Muoio, D.M., and Newgard, C.B. (2008). Molecular and metabolic mechanisms of insulin resistance and beta-cell failure in type 2 diabetes. Nature Reviews Molecular Cell Biology *9*, 193-205.

Murtaugh, L.C. (2007). Pancreas and beta-cell development: from the actual to the possible. Development *134*, 427-438.

Muzumdar, M.D., Tasic, B., Miyamichi, K., Li, L., and Luo, L.Q. (2007). A global double-fluorescent cre reporter mouse. Genesis *45*, 593-605.

Naya, F.J., Huang, H.P., Qiu, Y.H., 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 & Development *11*, 2323-2334.

Nishii, K., Kumai, M., Egashira, K., Miwa, T., Hashizume, K., Miyano, Y., and Shibata, Y. (2003). Mice lacking connexin45 conditionally in cardiac myocytes display embryonic lethality similar to that of germline knockout mice without endocardial cushion defect. Cell Communication and Adhesion *10*, 365-369.

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. Developmental Biology *293*, 526-539.

Norgaard, G.A., Jensen, J.N., and Jensen, J. (2003). FGF10 signaling maintains the pancreatic progenitor cell state revealing a novel role of Notch in organ development. Developmental Biology *264*, 323-338.

Novak, A., Guo, C., Yang, W., Nagy, A., and Lobe, C.G. (2000). Z/EG, a double reporter mouse line that expresses enhanced green fluorescent protein upon Cre-mediated excision. Genesis *28*, 147-155.

Offield, M.F., Jetton, T.L., Labosky, P.A., Ray, M., Stein, R.W., Magnuson, M.A., Hogan, B.L.M., and Wright, C.V.E. (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. Journal of Clinical Investigation *119*.

Ooi, S.K.T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C.D.*, et al.* (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature *448*, 714-U713.

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.

Pagliuca, F.W., and Melton, D.A. (2013). How to make a functional beta-cell. Development *140*, 2472-2483.

Pagliuca, F.W., Millman, J.R., Guertler, 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., Bankaitis, E.D., Boyer, D., Xu, X.B., Van de Casteele, M., Magnuson, M.A., Heimberg, H., and Wright, C.V.E. (2013). Spatiotemporal patterns of multipotentiality in Ptf1a-expressing cells during pancreas organogenesis and injury-induced facultative restoration. Development *140*, 751-764.

Pan, F.C., and Wright, C. (2011). Pancreas organogenesis: from bud to plexus to gland. Dev Dyn *240*, 530-565.

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 & Development *25*, 2291-2305.

Park, C.W., Kim, H.W., Ko, S.H., Lim, J.H., Ryu, G.R., Chung, H.W., Han, S.W., Shin, S.J., Bang, B.K., Breyer, M.D.*, et al.* (2007). Long-term treatment of glucagon-like peptide-1 analog exendin-4 ameliorates diabetic nephropathy through improving metabolic anomalies in db/db mice. Journal of the American Society of Nephrology *18*, 1227-1238.

Passegue, E., Wagner, E.F., and Weissman, I.L. (2004). JunB deficiency leads to a myeloproliferative disorder arising from hematopoietic stem cells. Cell *119*, 431-443.

Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., Kabadi, A.M., Polstein, L.R., Thakore, P.I., Glass, K.A., Ousterout, D.G., Leong, K.W.*, et al.* (2013). RNA-guided gene activation by CRISPR-Cas9-based transcription factors. Nature Methods *10*, 973-+.

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. Journal of Molecular Endocrinology *37*, 301-316.

Polstein, L.R., and Gersbach, C.A. (2015). A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. Nature Chemical Biology *11*, 198-U179.

Postic, C., and Magnuson, M.A. (2000). DNA excision in liver by an albumin-Cre transgene occurs progressively with age. Genesis *26*, 149-150.

Postic, C., Shiota, M., Niswender, K.D., Jetton, T.L., Chen, Y.J., Moates, J.M., Shelton, K.D., Lindner, J., Cherrington, A.D., and Magnuson, M.A. (1999). Dual roles for glucokinase in glucose homeostasis as determined by liver and pancreatic beta cell-specific gene knock-outs using Cre recombinase. Journal of Biological Chemistry *274*, 305-315.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. Cell *152*, 1173-1183.

Quesada, I., Tudurí, E., Ripoll, C., and Nadal, A. (2008). Physiology of the pancreatic alpha-cell and glucagon secretion: role in glucose homeostasis and diabetes. J Endocrinol *199*, 5-19.

Ramirez, A., Page, A., Gandarillas, A., Zanet, J., Pibre, S., Vidal, M., Tusell, L., Genesca, A., Whitaker, D.A., Melton, D.W*., et al.* (2004). A keratin K5Cre transgenic line appropriate for tissue-specific or generalized Cre-mediated recombination. Genesis *39*, 52-57.

Rankin, M.M., Wilbur, C.J., Rak, K., Shields, E.J., Granger, A., and Kushner, J.A. (2013). beta-Cells Are Not Generated in Pancreatic Duct Ligation-Induced Injury in Adult Mice. Diabetes *62*, 1634-1645.

Reinert, R.B., Brissova, M., Shostak, A., Pan, F.C., Poffenberger, G., Cai, Q., Hundemer, G.L., Kantz, J., Thompson, C.S., Dai, C*., et al.* (2013). Vascular Endothelial Growth Factor-A and Islet Vascularization Are Necessary in Developing, but Not Adult, Pancreatic Islets. Diabetes *62*, 4154-4164.

Rezania, A., Bruin, J.E., Riedel, M.J., Mojibian, M., Asadi, A., Xu, J., Gauvin, R., Narayan, K., Karanu, F., O'Neil, J.J*., et al.* (2012). Maturation of Human Embryonic Stem Cell-Derived Pancreatic Progenitors Into Functional Islets Capable of Treating Pre-existing Diabetes in Mice. Diabetes *61*, 2016-2029.

Rieck, S., Bankaitis, E.D., and Wright, C.V.E. (2012). Lineage determinants in early endocrine development. Seminars in Cell & Developmental Biology *23*, 673-684.

Riley, K.G., Pasek, R.C., Maulis, M.F., Peek, J., Thorel, F., Brigstock, D.R., Herrera, P.L., and Gannon, M. (2015). Connective Tissue Growth Factor Modulates Adult beta-Cell Maturity and Proliferation to Promote beta-Cell Regeneration in Mice. Diabetes *64*, 1284-1298.

Rodriguez-Diaz, R., Dando, R., Jacques-Silva, M.C., Fachado, A., Molina, J., Abdulreda, M.H., Ricordi, C., Roper, S.D., Berggren, P.-O., and Caicedo, A. (2011). Alpha cells secrete acetylcholine as a non-neuronal paracrine signal priming beta cell function in humans. Nature Medicine *17*, 888-U258.

Romer, A.I., and Sussel, L. (2015). Pancreatic islet cell development and regeneration. Current Opinion in Endocrinology Diabetes and Obesity *22*, 255-264.

Romm, E., Nielsen, J.A., Kim, J.G., and Hudson, L.D. (2005). Myt1 family recruits histone deacetylase to regulate neural transcription. Journal of Neurochemistry *93*, 1444-1453.

Rorsman, P., Eliasson, L., Renstrom, E., Gromada, J., Barg, S., and Gopel, S. (2000). The cell physiology of biphasic insulin secretion. News in Physiological Sciences *15*, 72-77.

Rose, N.R., and Klose, R.J. (2014). Understanding the relationship between DNA methylation and histone lysine methylation. Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms *1839*, 1362-1372.

Rother, K.I., and Harlan, D.M. (2004). Challenges facing islet transplantation for the treatment of type 1 diabetes mellitus. Journal of Clinical Investigation *114*, 877-883.

Saltiel, A.R., and Kahn, C.R. (2001). Insulin signalling and the regulation of glucose and lipid metabolism. Nature *414*, 799-806.

Samuel, V.T., and Shulman, G.I. (2012). Mechanisms for Insulin Resistance: Common Threads and Missing Links. Cell *148*, 852-871.

Santoro, S.W., and Schultz, P.G. (2002). Directed evolution of the site specificity of Cre recombinase. Proceedings of the National Academy of Sciences of the United States of America *99*, 4185-4190.

Saraf-Levy, T., Santoro, S.W., Volpin, H., Kushnirsky, T., Eyal, Y., Schultz, P.G., Gidoni, D., and Carmi, N. (2006). Site-specific recombination of asymmetric lox sites mediated by a heterotetrameric Cre recombinase complex. Bioorganic & Medicinal Chemistry *14*, 3081-3089.

Saund, R.S., Kanai-Azuma, M., Kanai, Y., Kim, I., Lucero, M.T., and Saijoh, Y. (2012). Gut endoderm is involved in the transfer of left-right asymmetry from the node to the lateral plate mesoderm in the mouse embryo. Development *139*, 2426-2435.

Schaffer, A.E., Freude, K.K., Nelson, S.B., and Sander, M. (2010). Nkx6 Transcription Factors and Ptf1a Function as Antagonistic Lineage Determinants in Multipotent Pancreatic Progenitors. Developmental Cell *18*, 1022-1029.

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 Genetics *9*.

Schmidt, E.E., Taylor, D.S., Prigge, J.R., Barnett, S., and Capecchi, M.R. (2000). Illegitimate Cre-dependent chromosome rearrangements in transgenic mouse spermatids. Proceedings of the National Academy of Sciences of the United States of America *97*, 13702-13707.

Schmidt-Supprian, M., and Rajewsky, K. (2007). Vagaries of conditional gene targeting. Nature Immunology *8*, 665-668.

Schonhoff, S.E., Giel-Moloney, M., and Leiter, A.B. (2004). Neurogenin 3-expressing progenitor cells in the gastrointestinal tract differentiate into both endocrine and non-endocrine cell types. Developmental Biology *270*, 443-454.

Schwetz, T.A., Ustione, A., and Piston, D.W. (2013). Neuropeptide Y and somatostatin inhibit insulin secretion through different mechanisms. American Journal of Physiology-Endocrinology and Metabolism *304*, E211-E221.

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.

Serre-Beinier, V., Bosco, D., Zulianello, L., Charollais, A., Caille, D., Charpantier, E., Gauthier, B.R., Diaferia, G.R., Giepmans, B.N., Lupi, R.*, et al.* (2009). Cx36 makes channels coupling human pancreatic beta-cells, and correlates with insulin expression. Human Molecular Genetics *18*, 428-439.

Seymour, P.A., Freude, K.K., Dubois, C.L., Shih, H.-P., Patel, N.A., and Sander, M. (2008). A dosage-dependent requirement for Sox9 in pancreatic endocrine cell formation. Developmental Biology *323*, 19-30.

Seymour, P.A., Freude, K.K., Tran, M.N., Mayes, E.E., Jensen, J., Kist, R., Scherer, G., and Sander, M. (2007). SOX9 is required for maintenance of the pancreatic progenitor cell pool. Proceedings of the National Academy of Sciences of the United States of America *104*, 1865-1870.

Shaikh, A.C., and Sadowski, P.D. (2000). Chimeras of the Flp and Cre recombinases: Tests of the mode of cleavage by Flp and Cre. Journal of Molecular Biology *302*, 27-48.

Shapiro, A.M.J., Lakey, J.R.T., Ryan, E.A., Korbutt, G.S., Toth, E., Warnock, G.L., Kneteman, N.M., and Rajotte, R.V. (2000). Islet transplantation in seven patients with type 1 diabetes mellitus using a glucocorticoid-free immunosuppressive regimen. New England Journal of Medicine *343*, 230-238.

Shapiro, A.M.J., Ricordi, C., Hering, B.J., Auchincloss, H., Lindblad, R., Robertson, P., Secchi, A., Brendel, M.D., Berney, T., Brennan, D.C.*, et al.* (2006). International trial of the edmonton protocol for islet transplantation. New England Journal of Medicine *355*, 1318-1330.

Shapiro, E., Biezuner, T., and Linnarsson, S. (2013). Single-cell sequencing-based technologies will revolutionize whole-organism science. Nature Reviews Genetics *14*, 618-630.

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

Shih, H.P., Wang, A., and Sander, M. (2013). Pancreas Organogenesis: From Lineage Determination to Morphogenesis. Annual Review of Cell and Developmental Biology, Vol 29 *29*, 81-105.

Slack, J.M.W. (1995). DEVELOPMENTAL BIOLOGY OF THE PANCREAS. Development *121*, 1569-1580.

Smallwood, S.A., Lee, H.J., Angermueller, C., Krueger, F., Saadeh, H., Peet, J., Andrews, S.R., Stegle, O., Reik, W., and Kelsey, G. (2014). Single-cell genome-

wide bisulfite sequencing for assessing epigenetic heterogeneity. Nature Methods *11*, 817-820.

Smith, S.B., Gasa, R., Watada, H., Wang, J.H., Griffen, S.C., and German, M.S. (2003). Neurogenin3 and hepatic nuclear factor 1 cooperate in activating pancreatic expression of Pax4. Journal of Biological Chemistry *278*, 38254-38259.

Smith, S.B., Qu, H.-Q., Taleb, N., Kishimoto, N.Y., Scheel, D.W., Lu, Y., Patch, A.-M., Grabs, R., Wang, J., Lynn, F.C*., et al.* (2010). Rfx6 directs islet formation and insulin production in mice and humans. Nature *463*, 775-780.

Smith, S.B., Watada, H., and German, M.S. (2004). Neurogenin3 activates the islet differentiation program while repressing its own expression. Molecular Endocrinology *18*, 142-149.

Solar, M., Cardalda, C., Houbracken, I., Martin, M., Angel Maestro, M., De Medts, N., Xu, X., Grau, V., Heimberg, H., Bouwens, L*., et al.* (2009). Pancreatic Exocrine Duct Cells Give Rise to Insulin-Producing beta Cells during Embryogenesis but Not after Birth. Developmental Cell *17*, 849-860.

Soriano, P. (1999). Generalized lacZ expression with the ROSA26 Cre reporter strain. Nature Genetics *21*, 70-71.

Sosa-Pineda, B. (2004). The gene pax4 is an essential regulator of pancreatic beta-cell development. Molecules and Cells *18*, 289-294.

Sousa-Ferreira, L., de Almeida, L.P., and Cavadas, C. (2014). Role of hypothalamic neurogenesis in feeding regulation. Trends in Endocrinology and Metabolism *25*, 80-88.

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.

Spence, J.R., Lange, A.W., Lin, S.-C.J., Kaestner, K.H., Lowy, A.M., Kim, I., Whitsett, J.A., and Wells, J.M. (2009). Sox17 Regulates Organ Lineage Segregation of Ventral Foregut Progenitor Cells. Developmental Cell *17*, 62-74.

Srinivas, S., Watanabe, T., Lin, C.-S., William, C.M., Tanabe, Y., Jessell, T.M., and Costantini, F. (2001). Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. BMC Developmental Biology *1*, 1-8.

Steiner, D.J., Kim, A., Miller, K., and Hara, M. (2010). Pancreatic islet plasticity Interspecies comparison of islet architecture and composition. Islets *2*, 135-145.

Sternberg, N., and Hamilton, D. (1981). Bacteriophage-p1 site-specific recombination .1. Recombination between loxp sites. Journal of Molecular Biology *150*, 467-486.

Stoffers, D.A., Kieffer, T.J., Hussain, M.A., Drucker, D.J., Bonner-Weir, S., Habener, J.F., and Egan, J.M. (2000). Insulinotropic glucagon-like peptide 1 agonists stimulate

expression of homeodomain protein IDX-1 and increase islet size in mouse pancreas. Diabetes *49*, 741-748.

Strowski, M.Z., and Blake, A.D. (2008). Function and expression of somatostatin receptors of the endocrine pancreas. Molecular and Cellular Endocrinology *286*, 169-179.

Sugiyama, T., Benitez, C.M., Ghodasara, A., Liu, L., McLean, G.W., Lee, J., Blauwkamp, T.A., Nusse, R., Wright, C.V.E., Gu, G.*, et al.* (2013). Reconstituting pancreas development from purified progenitor cells reveals genes essential for islet differentiation. Proceedings of the National Academy of Sciences of the United States of America *110*, 12691-12696.

Sugiyama, T., Rodriguez, R.T., McLean, G.W., and Kim, S.K. (2007). Conserved markers of fetal pancreatic epithelium permit prospective isolation of islet progenitor cells by FACS. Proceedings of the National Academy of Sciences of the United States of America *104*, 175-180.

Sulston, J.E., Schierenberg, E., White, J.G., and Thomson, J.N. (1983). The embryonic-cell lineage of the nematode Caenorhabditis elegans. Developmental Biology *100*, 64-119.

Sussel, L., Kalamaras, J., Hartigan-O'Connor, D.J., Meneses, J.J., Pedersen, R.A., Rubenstein, J.L.R., 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.

Taniguchi, C.M., Emanuelli, B., and Kahn, C.R. (2006). Critical nodes in signalling pathways: insights into insulin action. Nature Reviews Molecular Cell Biology *7*, 85-96.

Tasic, B., Hippenmeyer, S., Wang, C., Gamboa, M., Zong, H., Chen-Tsai, Y., and Luo, L. (2011). Site-specific integrase-mediated transgenesis in mice via pronuclear injection. Proceedings of the National Academy of Sciences of the United States of America *108*, 7902-7907.

Tasic, B., Miyamichi, K., Hippenmeyer, S., Dani, V.S., Zeng, H., Joo, W., Zong, H., Chen-Tsai, Y., and Luo, L. (2012). Extensions of MADM (Mosaic Analysis with Double Markers) in Mice. Plos One *7*.

Theis, M., Mas, C., Doring, B., Degen, J., Brink, C., Caille, D., Charollais, A., Kruger, O., Plum, A., Nepote, V.*, et al.* (2004). Replacement by a lacZ reporter gene assigns mouse connexin36, 45 and 43 to distinct cell types in pancreatic islets. Experimental Cell Research *294*, 18-29.

Thompson, J.F., Hayes, L.S., and Lloyd, D.B. (1991). Modulation of firefly luciferase stability and impact on studies of gene-regulation. Gene *103*, 171-177.

Thorel, F., Nepote, V., Avril, I., Kohno, K., Desgraz, R., Chera, S., and Herrera, P.L. (2010). Conversion of adult pancreatic alpha-cells to beta-cells after extreme beta-cell loss. Nature *464*, 1149-1154.

Thyagarajan, B., Guimaraes, M.J., Groth, A.C., and Calos, M.P. (2000). Mammalian genomes contain active recombinase recognition sites. Gene *244*, 47-54.

Urlinger, S., Baron, U., Thellmann, M., Hasan, M.T., Bujard, H., and Hillen, W. (2000). Exploring the sequence space for tetracycline-dependent transcriptional activators: Novel mutations yield expanded range and sensitivity. Proceedings of the National Academy of Sciences of the United States of America *97*, 7963-7968.

Vaes, B.L.T., Lute, C., van der Woning, S.P., Piek, E., Vermeer, J., Blom, H.J., Mathers, J.C., Mueller, M., de Groot, L.C.P.G.M., and Steegenga, W.T. (2010). Inhibition of methylation decreases osteoblast differentiation via a non-DNA-dependent methylation mechanism. Bone *46*, 514-523.

Valiunas, V., Polosina, Y.Y., Miller, H., Potapova, I.A., Valiuniene, L., Doronin, S., Mathias, R.T., Robinson, R.B., Rosen, M.R., Cohen, I.S*., et al.* (2005). Connexin-specific cell-to-cell transfer of short interfering RNA by gap junctions. Journal of Physiology-London *568*, 459-468.

Van Belle, T.L., Coppieters, K.T., and Von Herrath, M.G. (2011). Type 1 Diabetes: Etiology, Immunology, and Therapeutic Strategies. Physiological Reviews *91*, 79-118.

van der Meulen, T., Donaldson, C.J., Caceres, E., Hunter, A.E., Cowing-Zitron, C., Pound, L.D., Adams, M.W., Zembrzycki, A., Grove, K.L., and Huising, M.O. (2015). Urocortin3 mediates somatostatin-dependent negative feedback control of insulin secretion. Nature Medicine *21*, 769-+.

Van Duyne, G.D. (2001). A structural view of Cre-loxP site-specific recombination. Annual Review of Biophysics and Biomolecular Structure *30*, 87-104.

Varshavsky, A. (1996). The N-end rule: Functions, mysteries, uses. Proceedings of the National Academy of Sciences of the United States of America *93*, 12142-12149.

Vedadi, M., Barsyte-Lovejoy, D., Liu, F., Rival-Gervier, S., Allali-Hassani, A., Labrie, V., Wigle, T.J., DiMaggio, P.A., Wasney, G.A., Siarheyeva, A*., et al.* (2011). A chemical probe selectively inhibits G9a and GLP methyltransferase activity in cells. Nature Chemical Biology *7*, 566-574.

Villasenor, A., Chong, D.C., Henkemeyer, M., and Cleaver, O. (2010). Epithelial dynamics of pancreatic branching morphogenesis. Development *137*, 4295-4305.

Viollet, B., and Foretz, M. (2013). Revisiting the mechanisms of metformin action in the liver. Annales D Endocrinologie *74*, 123-129.

Vire, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J.M*., et al.* (2006). The Polycomb group protein EZH2 directly controls DNA methylation. Nature *439*, 871-874.

Vooijs, M., Jonkers, J., and Berns, A. (2001). A highly efficient ligand-regulated Cre recombinase mouse line shows that LoxP recombination is position dependent. Embo Reports *2*, 292-297.

Wang, J.F., Elghazi, L., Parker, S.E., Kizilocak, H., Asano, M., Sussel, L., and Sosa-Pineda, B. (2004). The concerted activities of Pax4 and Nkx2.2 are essential to initiate pancreatic beta-cell differentiation. Developmental Biology *266*, 178-189.

Wang, P., Alvarez-Perez, J.-C., Felsenfeld, D.P., Liu, H., Sivendran, S., Bender, A., Kumar, A., Sanchez, R., Scott, D.K., Garcia-Ocana, A.*, et al.* (2015). A high-throughput chemical screen reveals that harmine-mediated inhibition of DYRK1A increases human pancreatic beta cell replication. Nature Medicine *21*, 383-+.

Wang, S., Hecksher-Sorensen, J., Xu, Y., Zhao, A., Dor, Y., Rosenberg, L., Serup, P., and Gu, G. (2008). Myt1 and Ngn3 form a feed-forward expression loop to promote endocrine islet cell differentiation. Developmental Biology *317*, 531-540.

Wang, S., Jensen, J.N., Seymour, P.A., Hsu, W., Dor, Y., Sander, M., Magnuson, M.A., Serup, P., and Gu, G. (2009a). 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.E., and Gu, G. (2010). Neurog3 gene dosage regulates allocation of endocrine and exocrine cell fates in the developing mouse pancreas. Developmental Biology *339*, 26-37.

Wang, S., Zhang, J., Zhao, A., Hipkens, S., Magnuson, M.A., and Gu, G. (2007). Loss of Myt1 function partially compromises endocrine islet cell differentiation and pancreatic physiological function in the mouse. Mechanisms of Development *124*, 898-910.

Wang, W.D., Walker, J.R., Wang, X., Tremblay, M.S., Lee, J.W., Wu, X., and Schultz, P.G. (2009b). Identification of small-molecule inducers of pancreatic beta-cell expansion. Proceedings of the National Academy of Sciences of the United States of America *106*, 1427-1432.

Wang, Y., Quagliarini, F., Gusarova, V., Gromada, J., Valenzuela, D.M., Cohen, J.C., and Hobbs, H.H. (2013). Mice lacking ANGPTL8 (Betatrophin) manifest disrupted triglyceride metabolism without impaired glucose homeostasis. Proceedings of the National Academy of Sciences of the United States of America *110*, 16109-16114.

Wang, Z., and Thurmond, D.C. (2009). Mechanisms of biphasic insulin-granule exocytosis - roles of the cytoskeleton, small GTPases and SNARE proteins. Journal of Cell Science *122*, 893-903.

Watada, H., Mirmira, R.G., Leung, J., and German, M.S. (2000). Transcriptional and translational regulation of beta-cell differentiation factor Nkx6.1. Journal of Biological Chemistry *275*, 34224-34230.

Watada, H., Scheel, D.W., Leung, J., and German, M.S. (2003). Distinct gene expression programs function in progenitor and mature islet cells. Journal of Biological Chemistry *278*, 17130-17140.

White, P., May, C.L., Lamounier, R.N., Brestelli, J.E., and Kaestner, K.H. (2008). Defining pancreatic endocrine precursors and their descendants. Diabetes *57*, 654-668.

Wicksteed, B., Brissova, M., Yan, W.B., Opland, D.M., Plank, J.L., Reinert, R.B., Dickson, L.M., Tamarina, N.A., Philipson, L.H., Shostak, A.*, et al.* (2010). Conditional Gene Targeting in Mouse Pancreatic beta-Cells Analysis of Ectopic Cre Transgene Expression in the Brain. Diabetes *59*, 3090-3098.

Wilson, M.E., Scheel, D., and German, M.S. (2003). Gene expression cascades in pancreatic development. Mechanisms of Development *120*, 65-80.

Wong, R.C.B., Pera, M.F., and Pebay, A. (2008). Role of Gap Junctions in Embryonic and Somatic Stem Cells. Stem Cell Reviews *4*, 283-292.

Woodhead, G.J., Mutch, C.A., Olson, E.C., and Chenn, A. (2006). Cell-autonomous beta-catenin signaling regulates cortical precursor proliferation. Journal of Neuroscience *26*, 12620-12630.

Xiao, X., Guo, P., Shiota, C., Prasadan, K., El-Gohary, Y., Wiersch, J., Gaffar, I., and Gittes, G.K. (2013). Neurogenin3 Activation Is Not Sufficient to Direct Duct-to-Beta Cell Transdifferentiation in the Adult Pancreas. Journal of Biological Chemistry *288*, 25297-25308.

Xu, G., Stoffers, D.A., Habener, J.F., and Bonner-Weir, S. (1999). Exendin-4 stimulates both beta-cell replication and neogenesis, resulting in increased beta-cell mass and improved glucose tolerance in diabetic rats. Diabetes *48*, 2270-2276.

Xu, X., D'Hoker, J., Stange, G., Bonne, S., De Leu, N., Xiao, X., De Casteele, M.V., Mellitzer, G., Ling, Z., Pipeleers, D.*, et al.* (2008). beta cells can be generated from endogenous progenitors in injured adult mouse pancreas. Cell *132*, 197-207.

Xu, Y.W., Xu, G., Liu, B.D., and Gu, G.Q. (2007). Cre reconstitution allows for DNA recombination selectively in dual-marker-expressing cells in transgenic mice. Nucleic Acids Research *35*, 9.

Yamamoto, M., Shook, N.A., Kanisicak, O., Yamamoto, S., Wosczyna, M.N., Camp, J.R., and Goldhamer, D.J. (2009). A Multifunctional Reporter Mouse Line for Cre- and FLP-Dependent Lineage Analysis. Genesis *47*, 107-114.

Yang, Y.-P., Thorel, F., Boyer, D.F., Herrera, P.L., and Wright, C.V.E. (2011). Context-specific alpha-to-beta-cell reprogramming by forced Pdx1 expression. Genes & Development *25*, 1680-1685.

Yi, P., Park, J.-S., and Melton, D.A. (2013). Betatrophin: A Hormone that Controls Pancreatic beta Cell Proliferation. Cell *153*, 747-758.

Yokoyama, A., Igarashi, K., Sato, T., Takagi, K., Otsuka, M.I., Shishido, Y., Baba, T., Ito, R., Kanno, J., Ohkawa, Y.*, et al.* (2014). Identification of Myelin Transcription Factor 1 (MyT1) as a Subunit of the Neural Cell Type-specific Lysine-specific Demethylase 1 (LSD1) Complex. Journal of Biological Chemistry *289*, 18152-18162.

Zalzman, M., Anker-Kitai, L., and Efrat, S. (2005). Differentiation of human liver-derived, insulin-producing cells toward the beta-cell phenotype. Diabetes *54*, 2568-2575.

Zambrowicz, B.P., Imamoto, A., Fiering, S., Herzenberg, L.A., Kerr, W.G., and Soriano, P. (1997). Disruption of overlapping transcripts in the ROSA beta geo 26 gene trap strain leads to widespread expression of beta-galactosidase in mouse embryos and hematopoietic cells. Proceedings of the National Academy of Sciences of the United States of America *94*, 3789-3794.

Zhang, F., Cong, L., Lodato, S., Kosuri, S., Church, G.M., and Arlotta, P. (2011). Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nature Biotechnology *29*, 149-U190.

Zhang, Z.W., and Lutz, B. (2002). Cre recombinase-mediated inversion using lox66 and lox71: method to introduce conditional point mutations into the CREB-binding protein. Nucleic Acids Research *30*, 5.

Zhao, Q., Rank, G., Tan, Y.T., Li, H., Moritz, R.L., Simpson, R.J., Cerruti, L., Curtis, D.J., Patel, D.J., Allis, C.D.*, et al.* (2009). PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing. Nature Structural & Molecular Biology *16*, 304-311.

Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J., and Melton, D.A. (2008). In vivo reprogramming of adult pancreatic exocrine cells to beta-cells. Nature *455*, 627-U630.

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. Developmental Cell *13*, 103-114.

Zimmet, P., Alberti, K., and Shaw, J. (2001). Global and societal implications of the diabetes epidemic. Nature *414*, 782-787.

Ziv, O., Glaser, B., and Dor, Y. (2013). The Plastic Pancreas. Developmental Cell *26*, 3-7.

Zong, H., Espinosa, S., Su, H.H., Muzumdar, M.D., and Luo, L.Q. (2005). Mosaic analysis with double markers in mice. Cell *121*, 479-492.