

TRANSCRIPTIONAL PROFILING OF PANCREATIC PROGENITOR CELLS

By

Leah Ashley Potter

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

Molecular Physiology and Biophysics

December, 2011

Nashville, Tennessee

Approved:

Professor Maureen Gannon

Professor Alyssa Hasty

Professor Patricia Labosky

Professor David Piston

Professor Christopher V.E. Wright

MOLECULAR PHYSIOLOGY AND BIOPHYSICS

---

TRANSCRIPTIONAL PROFILING OF PANCREATIC PROGENITOR CELLS

LEAH ASHLEY POTTER

Dissertation under the direction of Mark A. Magnuson, M.D.

Pluripotent stem cells, because of their ability to differentiate into any cell type, have been widely advocated as a means of producing a nearly unlimited source of new insulin-producing  $\beta$  cells for the treatment of diabetic diseases. However, while there has been remarkable progress in learning how to direct the differentiation of human embryonic stem (hES) cells towards pancreatic endocrine cell fates, insulin-expressing cells made in this manner are often polyhormonal and lack a normal response to glucose, thereby suggesting a need for a deeper understanding of the gene regulatory networks that are established in a stepwise manner during pancreas development.

My thesis studies explored three main topics, each of which holds potential for the development of improved hES cell directed pancreatic differentiation protocols and the discovery of genes that may specifically affect  $\beta$  cell development. First, we used mice that contained a fluorescent reporter allele and fluorescence-activated cell sorting (FACS) to isolate several discrete pancreatic cell populations which were then analyzed using whole transcriptome sequencing (RNA-Seq). By doing so, we were able to examine the genetic requirement and temporal changes of cells expressing *pancreas specific transcription factor 1a* (*Ptf1a*), a marker of the pancreatic multipotent progenitor cells (MPCs) and of acinar-specified cells, during pancreas development. By comparing the transcriptional profiles, we identified five gene clusters, each of which provides insights into the dynamics of gene

expression during specific aspects of pancreas development. Second, my studies revealed that *Nephrocan*, an inhibitor of the TGF $\beta$  signaling pathway, was expressed in pancreatic MPCs. Thus, to explore the role of *Nepn* further, we generated mice containing a single copy insertion of a *Nepn-Cherry* transgene. Finally, to facilitate the combinatorial sorting of *Pdx1*- and *Ptf1a*-expressing cells during early pancreas development, we generated a mouse line expressing a cyan fluorescent protein under control of the endogenous *pancreatic and duodenal homeobox 1 (Pdx1)* gene.

The research that I have performed is part of a larger project focused on generating and characterizing a series of high quality transcriptional profiles representing key stages in the generation of pancreatic endocrine cells that occur naturally in the mouse. We anticipate that further analysis of the datasets I have generated for specific developmental stages, in combination with similarly generated datasets at other developmental stages, will facilitate identification of signaling pathways and gene clusters essential for formation of functional pancreatic  $\beta$  cells in the mouse, thereby stimulating new hypotheses for identifying pro- $\beta$  cell signals necessary to direct the differentiation of pluripotent stem cells into pancreatic  $\beta$  cells.

## ACKNOWLEDGEMENTS

Chinese fortunes have always been a fundamental means of communication in the Magnuson laboratory. As I looked through my collection of fortunes on the corner of my desk, the one that stood out the most was, “A solid challenge will bring forth your finest abilities.” By far, graduate school has been my most challenging endeavor, and I would be remiss not to acknowledge the people who have imparted their knowledge and encouraged me along the way.

First, I would like to thank the members of the Magnuson laboratory, both past and present. I am forever indebted to Jill Lindner, Susan Hipkens and Pamela Uttz. They have been with me since the very beginning, and I cannot fathom what my time here at Vanderbilt would have been like without them. They encouraged me every step of the way, “taught me the ropes,” and were always there to listen when I needed an ear. I am grateful to Eunyoung Choi, Anna Osipovich, Judsen Schneider, Madhi Saranadasa, Anil Laxman, Jared Burlison, Yanyun Gu, and Weiping Yuan. No lab can function without a scientific ‘sounding board,’ and I am thankful to them for providing an environment that stimulates scientific discussions. I would also like to thank, Rama Gangula, Changqing Zhang, Kathy Shelton, and Jody Peters, whose technical expertise and assistance were invaluable.

I would like to thank the members of the Beta Cell Biology Consortium, specifically Chris Stoeckert, Elisabetta Manduchi, Guoqiang Gu, Lori Sussel, and Ray MacDonald, as well as the BCBC coordinating center affiliates, Jean-Philippe Cartailler, Thomas Houfek, Suzannah Green and Kim Cornwell. Within the Vanderbilt Center for Stem Cell Biology, I would like to thank members of the Labosky and Huppert laboratories, specifically Stacey Huppert, Kari Huppert, Erin Sparks, Jennifer Plank and Elise Pfaltzgraff. The work presented

within this dissertation would not have been possible without the expertise and assistance from numerous shared resource facilities and their personnel, including Kevin Weller, Dave Flaherty, and Brittany Matlock from the Flow Cytometry Core and Shawn Levy, Vicky Amann, Chelsea Baker, Travis Clark, Christian Shaffer, David Sexton, and Robert Woodall from the Genome Sciences Resource. Additionally, a number of individuals at Vanderbilt have been generous in their assistance and in sharing their scientific expertise and reagents, including Anna Means, Al Powers, Michelle Southard-Smith, Jeffery Raum, Yan Hang, Michelle Guney, Chunhua Dai, Ashley Cantrell, Stephanie Byers, Clay Spencer, and Rebecca McWhirter. I am grateful to Roland Stein for providing constructive criticism, for cheering me on from the front row, and for simply making me smile even in the midst of difficult moments.

I am especially grateful to the members of my dissertation committee who have guided me through this entire process. To my chair, Maureen Gannon, who has gone above and beyond what is required, and to Alyssa Hasty, Trish Labosky, David Piston and Chris Wright, who have provided much needed encouragement at exactly the right moments, demonstrated what it means to be a successful scientist, provided constructive criticism and critiques, and who have seen me at my worst and still believed I had the potential to succeed.

While a great amount of support came from within my scientific community, a significant amount of support came from those outside this realm. First, I want to thank my family, Anthony, Charlotte, Rachel and Joshua. From the very beginning, they instilled in me the importance of hard work and nurtured my inquisitive mind. They have supported me every step of the way, from late night calls to weekend visits; they have been my one constant source of encouragement and for that I am grateful. To my extended family, Papa, Mrs. Doris, Nanny, JD, Dickey, KK, Jeff, Bob, Michelle, and Trey who were there

encouraging me at every step. To my close girlfriends, Amanda, Jessica, Mariruth, Mariam, Michelle, Stephanie, Katie, Heather and Stephanie, who were there to listen to every story and offer advice, an escape, a shoulder to cry on, or a glass of wine. To Eric, who has proved to be a wonderful distraction in the midst of what have been the most trying months of graduate school. I am blessed to call him my own. To Jennifer, who was one of the reasons I stayed, who was the solution to my FACS crisis, who was the one to practice my dissertation talk with, who was the shoulder to cry on, who was the one cheering when I succeeded, and who was the one has encouraged me at every point. For her, I am truly grateful.

Last, but certainly not least, I would like to thank Mark. Scientifically, he has afforded me the luxury of working in a laboratory outfitted with the best resources to perform my experiments. The colleagues he has surrounded himself with and the collaborations he has engaged in have furthered my scientific development and permitted my project to move beyond what was capable in my hands alone. He has given me the opportunity to pursue a project, that while difficult, has been intellectually rewarding and forced me to ‘blaze a trail.’ He has provided constructive criticism and encouraged me along the way, which is evident by the numerous fortunes and quotes that are displayed on my desk. While he once told me, “You won’t get a Ph.D. if you listen to me every day,” I must admit that it has been by his example, guidance, criticism, and encouragement that I have accomplished what I have.

## TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS .....	iv
LIST OF TABLES .....	xi
LIST OF FIGURES .....	xii
LIST OF ABBREVIATIONS .....	xiv
Chapter	
I. INTRODUCTION.....	1
Pancreas development, morphology and function .....	1
Diabetes mellitus and $\beta$ -cell dysfunction.....	3
Current treatments and limitations of islet transplantation .....	3
Emerging therapies for diabetes: Human embryonic stem (hES) cells as a source of insulin-producing cells.....	5
Embryonic stem (ES) cells, adult stem cells and progenitor cells .....	8
Signaling pathways and genetic regulatory factors directing pancreas development.....	9
Pancreatic multipotent progenitor cells .....	15
Genetic manipulations in mice and cells .....	16
Transcriptomics .....	19
Whole transcriptome profiling: RNA-Sequencing technology.....	21
Overview and Aims of Dissertation.....	23
II. DIRECTED DIFFERENTIATION OF HUMAN EMBRYONIC STEM CELLS AND GENERATION OF REPORTER ALLELES .....	27
Introduction.....	27
Directed differentiation of human ES cells towards pancreatic fates .....	27
Reporter alleles in human ES cells .....	29
Synopsis .....	30
Materials and Methods.....	31
Human ES cell culture and differentiation to pancreatic fates .....	31
TaqMan Low Density Array Design .....	33
RNA isolation and quantitative reverse transcription PCR (qRT-PCR).....	33
Generation of a <i>Pdx1</i> -driven GFP/Puro <sup>R</sup> -expressing reporter for RMCE .....	36
Generation of a <i>Pdx1</i> -driven CFP-expressing reporter for RMCE.....	40
Generation of targeting vectors for <i>Pdx1</i> gene targeting.....	42
Recombinase-mediated cassette exchange in human ES cells .....	43

Gene targeting in human ES cells.....	44
Results.....	45
Transition of human ES cells through definitive endoderm intermediate .....	45
Transition of human ES cells through the primitive gut tube and posterior foregut intermediates.....	48
Transition of human ES cells through an endocrine precursor intermediate and into hormone-positive cells.....	49
Generation of a reporter allele for <i>Pdx1</i> in human ES cells .....	50
Gene targeting in human ES cells.....	51
Discussion.....	52
 III. TRANSCRIPTIONAL PROFILING OF <i>PTF1A</i> <sup>YFP</sup> -EXPRESSING PANCREATIC CELLS .....	56
Introduction.....	56
Materials and Methods.....	58
Tissue isolation, embryonic genotyping, and flow cytometry .....	58
RNA amplification, library construction, RNA-Sequencing and analysis .....	59
Quantitative reverse transcription PCR (qRT-PCR) validation.....	60
Results.....	73
Isolation of three distinct <i>Ptf1a</i> <sup>YFP</sup> -expressing cell populations .....	73
Whole transcriptome profiling.....	73
Ptf1a-dependent gene expression at E11.5 .....	75
Temporal changes as MPCs are acinar-specified .....	77
Validation of specific genes and temporal profiling by qRT-PCR.....	77
Discussion.....	86
Gastrointestinal specification of pre-pancreatic endoderm occurs in the absence of Ptf1a.....	86
Developmental arrest of pre-pancreatic endoderm occurs in the absence of Ptf1a.....	87
Ptf1a <sup>YFP</sup> perdurance allows for the transient isolation of MPC progeny.....	89
Ptf1a promotes differentiation to acinar cell fates .....	90
Conclusions.....	90
 IV. CHARACTERIZATION OF NEPHROCAN DURING PANCREAS DEVELOPMENT .....	92
Introduction.....	92
Nephrocan, a unique member of the SLRP family .....	92
Nephrocan, an inhibitor of TGF $\beta$ signaling.....	93
Role of TGF $\beta$ signaling in endoderm and pancreas development.....	94
Expression of <i>Nephrocan</i> during mouse development .....	95
Materials and Methods.....	97
Generation of a ROSA26 <sup>LCA</sup> targeting vector .....	97
Generation of a ROSA26 <sup>Nepn-Cherry</sup> exchange cassette.....	98
Generation of a Nepn <sup>Cherry</sup> targeting vector .....	100
Mouse ES cell culture and electroporation .....	100

Recombinase-mediated cassette exchange .....	105
Blastocyst injections, mouse husbandry and genotyping .....	105
Immunolabeling.....	106
Fluorescence-activated cell sorting.....	107
Whole mount <i>in situ</i> hybridization .....	108
Embryo explants .....	109
Results.....	109
Expression pattern of <i>Nepn</i> in <i>Ptf1a<sup>YFP</sup></i> -expressing cells.....	109
Generation of a <i>ROSA26<sup>Nepn-Cherry</sup></i> reporter allele.....	110
Expression of <i>ROSA26<sup>Nepn-Cherry</sup></i> transgene .....	111
FACS analysis of <i>ROSA26<sup>Nepn-Cherry</sup></i> cells .....	114
Analysis of <i>Nephrocan</i> expression by <i>in situ</i> hybridization .....	116
Analysis of TGF $\beta$ signaling in embryos.....	116
Discussion.....	119
Generation of a reporter allele for <i>Nephrocan</i> by RMCE .....	119
<i>ROSA26<sup>Nepn-Cherry</sup></i> is expressed in a subset of pancreatic MPCs.....	120
<i>ROSA26<sup>Nepn-Cherry</sup></i> expression is downregulated as MPCs differentiate .....	122
<i>ROSA26<sup>Nepn-Cherry</sup></i> future directions.....	122
Potential role of <i>Nephrocan</i> in pancreatic progenitor development .....	123
V. FLUORESCENT PROTEIN REPORTER ALLELES FOR PDX1.....	126
Introduction.....	126
Materials and Methods.....	128
Generation of a loxed cassette acceptor allele for <i>Pdx1</i> .....	128
Generation of a GFP/Puro <sup>R</sup> -expressing reporter for RMCE.....	130
Generation of a CFP-expressing reporter for RMCE .....	132
Recombinase-mediated cassette exchange for <i>Pdx1</i> .....	136
Immunolabeling and imaging .....	138
Fluorescence-activated cell sorting.....	138
RNA isolation and semi-quantitative RT-PCR.....	139
Results.....	139
Generation of a <i>Pdx1<sup>LCA</sup></i> allele.....	139
Derivation of a <i>Pdx1<sup>GFP/PuroR</sup></i> allele by RMCE .....	139
Derivation of a <i>Pdx1<sup>CFP</sup></i> allele by RMCE .....	140
Expression pattern of <i>Pdx1<sup>CFP</sup></i> by whole mount fluorescence microscopy .....	141
Expression pattern of <i>Pdx1<sup>CFP</sup></i> by immunohistochemical analysis.....	143
FACS analysis of <i>Pdx1<sup>CFP</sup></i> cells.....	143
Discussion.....	146
VI. CONCLUSIONS AND FUTURE DIRECTIONS .....	148
Synopsis .....	148
Transcriptional profiling of pancreatic developmental intermediates .....	152
Examining complex gene regulatory networks and elucidating signaling pathways through transcriptional profiling of developmental intermediates .....	155
Analysis of cell populations using combinatorial sorting.....	157

Improving the directed pancreatic differentiation of human ES cells .....	160
VII. REFERENCES .....	164

## LIST OF TABLES

Table	Page
2.1 Summary of transcripts profiled on human ES cell TLDA .....	34
3.1 Summary of RNA-Seq results .....	61
3.2 Summary of transcripts profiled on 384- and 96-format TLDAs.....	63
4.1 Southern blot hybridization probes for <i>Nephrocan</i> gene targeting .....	101
4.2 Southern blot hybridization probes for <i>ROSA26</i> gene targeting .....	103
5.1 Southern blot hybridization probes for <i>Pdx1</i> gene targeting .....	131

## LIST OF FIGURES

Figure	Page
1.1 Pancreas anatomy and histology .....	2
1.2 Schematic of developmental intermediates <i>in vivo</i> and corresponding derivation of endocrine cells from ES cells <i>in vitro</i> .....	7
1.3 Pancreatic organogenesis: endoderm patterning and branching morphogenesis .....	11
1.4 <i>Ptf1a</i> and <i>Pdx1</i> expression during pancreas development .....	13
1.5 Scheme for gene targeting and recombinase-mediated cassette exchange (RMCE).....	18
1.6 Scheme of Illumina RNA-Sequencing .....	22
1.7 Developmental scheme of $\beta$ cell development and target cell populations for transcriptional profiling .....	24
2.1 Schematic diagram of various cell intermediates, transcripts expressed during development towards the endocrine lineage, and outline of differentiation procedure including media, growth factors and times of exposure .....	32
2.2 Directed differentiation of human ES cells towards pancreatic fates.....	37
2.3 Gene targeting of the <i>ROSA26</i> locus in human ES cells .....	39
2.4 Recombinase-mediated cassette exchange (RMCE) in human ES cells .....	41
2.5 Gene targeting of <i>Nanog</i> in human ES cells .....	46
3.1 Genetic and temporal comparisons of cells isolated from pancreatic MPCs, <i>Ptf1a</i> -deficient progenitor cells, and nascent acinar-specified cells.....	74
3.2 Genetic and temporal comparisons of differentially expressed transcripts as determined by RNA-Seq .....	76
3.3 Correlation of gene expression between RNA-Seq and qRT-PCR .....	79
3.4 Analysis of detectable transcripts by qRT-PCR relative to RPKM value .....	80
3.5 Quantitative RT-PCR profiling of differentially expressed transcripts.....	81
3.6 Genetic and temporal comparisons of differentially expressed transcripts as determined by qRT-PCR profiling .....	83

3.7 Temporal analysis of qRT-PCR profiling of differentially expressed transcripts.....	85
4.1 Gene expression profile of <i>Nepn</i> by qRT-PCR and RNA-Seq.....	96
4.2 VISTA plot displaying homology of <i>Nepn</i> locus among species.....	99
4.3 Generation of a <i>ROSA26<sup>Nepn-Cherry</sup></i> reporter allele by RMCE .....	104
4.4 Brightfield and fluorescent imaging of <i>ROSA26<sup>Nepn-Cherry</sup></i> embryos .....	112
4.5 Immunolabeling of <i>ROSA26<sup>Nepn-Cherry</sup></i> embryos .....	113
4.6 FACS analysis of <i>ROSA26<sup>Nepn-Cherry/+</sup></i> ; <i>Ptf1a<sup>YFP/+</sup></i> embryos .....	115
4.7 <i>In situ</i> analysis of endogenous <i>Nephrocan</i> .....	117
4.8 Embryo explants to study TGF $\beta$ signaling in specification of progenitor cells .....	118
4.9 Model of potential role of Nepn in pancreas development .....	124
5.1 Generation of a <i>Pdx1</i> loxed cassette acceptor (LCA) allele.....	129
5.2 Insertion of GFP/Puro <sup>R</sup> into a <i>Pdx1</i> loxed cassette acceptor (LCA) allele by RMCE ...	133
5.3 Validation of the GFP/Puro <sup>R</sup> -expressing construct in COS7 cells .....	134
5.4 Insertion of CFP into a <i>Pdx1</i> loxed cassette acceptor (LCA) allele by RMCE .....	135
5.5 Cerulean fluorescence in <i>Pdx1<sup>CFP/+</sup></i> embryos .....	142
5.6 Immunolabeling in <i>Pdx1<sup>CFP/+</sup></i> embryos.....	144
5.7 FACS analysis of <i>Pdx1<sup>CFP/+</sup></i> embryos .....	145
6.1 Combinatorial analysis of <i>Ptf1a<sup>YFP/+</sup></i> ; <i>Pdx1<sup>CFP/+</sup></i> embryos by fluorescence-activated cell sorting (FACS) .....	159
6.2 Spectral profile for cyan fluorescent protein (ECFP), green fluorescent protein (EGFP), yellow fluorescent protein (EYFP) and red fluorescent protein (DsRED).....	161

## LIST OF ABBREVIATIONS

6FAM:	6-carboxyfluorescein
7AAD:	7-Amino-actinomycin D
AIP:	anterior intestinal portal
ABI:	Applied Biosystems Inc.
BAC:	bacterial artificial chromosome
BMP:	bone morphogenic protein
bp:	base pair
BSA:	bovine serum albumin
cDNA:	complementary DNA
CFP:	cyan fluorescent protein
CHAPS:	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
ChIP-Seq:	Chromatin immunoprecipitation-Sequencing
CMV:	cytomegalovirus
Ct:	cycle threshold
CTD:	C-terminal domain
DAPI:	4',6-diamidino-2-phenylindole
DAPT:	N-[(3,5-Difluorophenyl)acetyl]-L-alanyl-2-phenyl]glycine-1,1-dimethylethyl ester
DE:	definitive endoderm
DIG:	digoxigenin
DMEM:	Dulbecco's Modified Eagle's Medium
DMSO:	Dimethyl sulfoxide

DNA:	deoxyribonucleic acid
DR4:	drug resistant 4 (neomycin, hygromycin, puromycin and 6-thioguanine)
DTA:	diphtheria toxin A
eGFP:	enhanced green fluorescent protein
EDTA:	ethylenediaminetetraacetic acid
EMT:	epithelial-mesenchymal transition
ES:	embryonic stem
ESP:	endoderm signature panel
EST:	expressed sequence tag
FACS:	fluorescence-activated cell sorting
FBS:	fetal bovine serum
FGF:	fibroblast growth factor
FLPe:	flippase
FP:	fluorescent protein
FRT:	flippase recognition target
GFP:	green fluorescent protein
GLP-1:	glucagon-like peptide-1
HA-TAG:	hemagglutinin tag
HEPES:	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hES:	human embryonic stem
HGF:	hepatocyte growth factor
HMG:	high mobility group
HSV:	herpes simplex virus
Hygro <sup>R</sup> :	hygromycin resistance gene (Hygromycin B phosphotransferase)

IGF:	insulin-like growth factor
iPS:	induced pluripotent stem
IRES:	internal ribosome entry site
kb:	kilobase
LCA:	loxed cassette acceptor
LM-PCR:	ligation mediated polymerase chain reaction
LP:	long pass
LRR:	leucine rich repeat
MCS:	multiple cloning site
MEF:	mouse embryonic fibroblast
mES:	mouse embryonic stem
MGBNFQ:	minor groove binding non-fluorescent quencher
MIP:	mouse insulin promoter
MPC:	multipotent progenitor cell
mRNA:	messenger ribonucleic acid
NCBI:	National Center for Biotechnology Information
NDS:	normal donkey serum
Neo <sup>R</sup> :	neomycin resistance gene (neomycin phosphotransferase II)
NICD:	Notch intracellular domain
NIH:	National Institutes of Health
NLS:	nuclear localization signal
nt:	nucleotide
OCT:	optimal cutting temperature
PI3K:	Phosphoinositide 3-kinase

PBS:	phosphate buffered saline
PCR:	polymerase chain reaction
pgk:	phosphoglycerol kinase
P/S:	Penicillin/Streptomycin
PTF1:	Pancreatic transcription factor 1
qRT-PCR:	quantitative reverse transcription PCR
RA:	retinoic acid
Rbpj:	recombination signal binding protein for immunoglobulin kappa J region
Rbpjl:	recombination signal binding protein for immunoglobulin kappa J region-like
RFP:	red fluorescent protein
RMCE:	recombinase-mediated cassette exchange
RNA:	ribonucleic acid
RNA-Seq:	RNA-Sequencing
ROCK:	p160-Rho-associated coiled-coil kinase
RPKM:	reads per kilobase of exon model per million mapped reads
RT-PCR:	reverse transcription PCR
SAGE:	serial analysis of gene expression
Shh:	Sonic hedgehog
SLRP:	small leucine rich repeat
SPIA:	single primer isothermal amplification
SSC:	saline-sodium citrate
SV40:	Simian virus 40
TBE:	Tris/borate/EDTA
TBS:	Tris buffered saline

tdRFP:	tandem dimer red fluorescent protein
TGF $\beta$ :	transforming growth factor $\beta$
tk:	thymidine kinase
TLDA:	TaqMan low density array
tRNA:	transfer RNA
UTR:	untranslated region
WA-09:	Wisconsin alumni 09
YFP:	yellow fluorescent protein

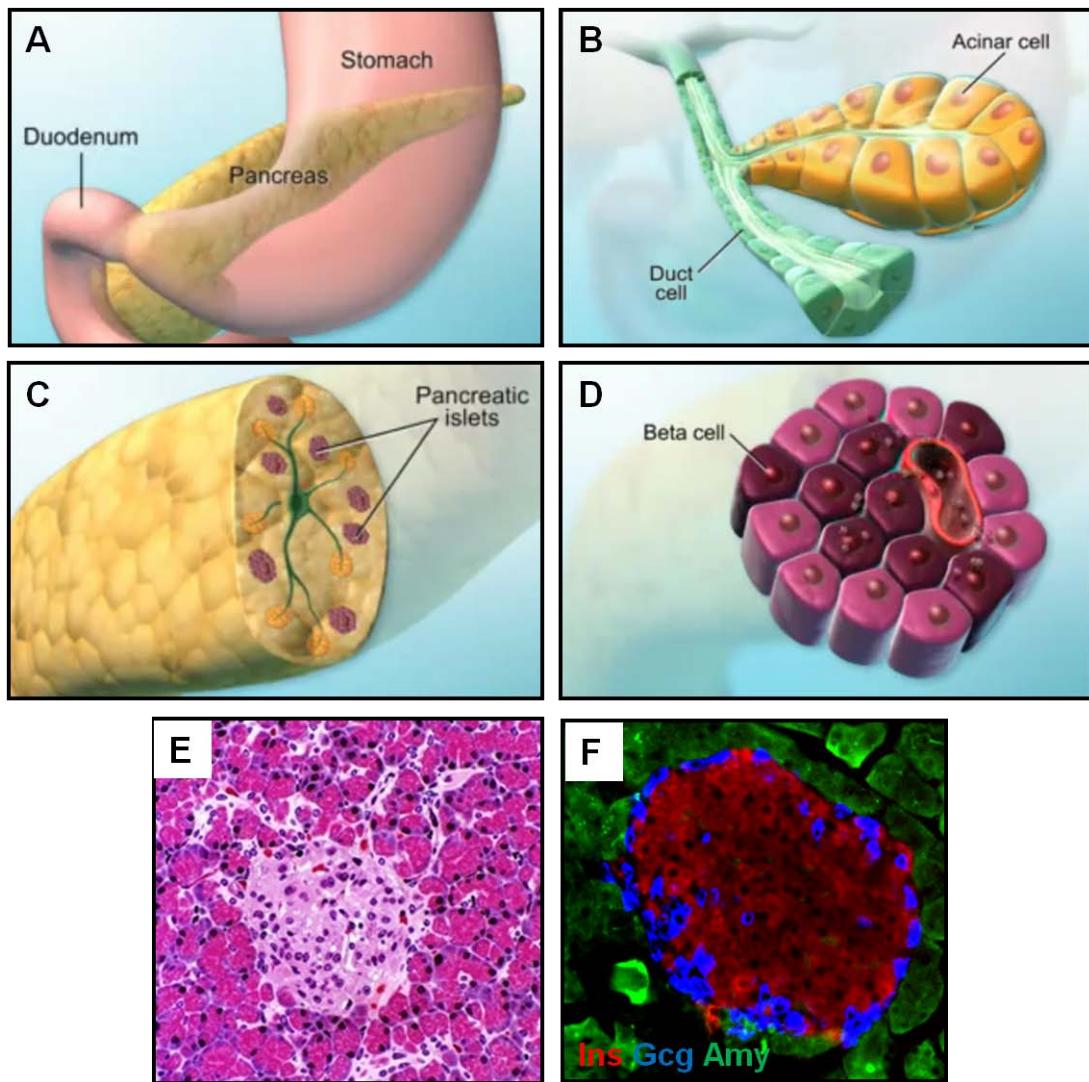
## CHAPTER I

### INTRODUCTION

#### Pancreas development, morphology and function

The pancreas is a dual-function organ located directly adjacent to the stomach and is an integral part of the digestive system (**Figure 1.1**). The exocrine cells comprise approximately 98% of the organ mass and produce digestive enzymes that are secreted into the small intestine through a system of ducts. Conversely, the endocrine cells, which are organized into the islets of Langerhans, comprise only 1 – 2% of the entire organ and produce a variety of hormones that regulate levels of glucose in the bloodstream. While the cellular architecture of the Islets of Langerhans differs among species, in the mouse pancreatic islets are primarily comprised of five hormone-secreting cell types arranged in a stereotypical manner: insulin-producing  $\beta$  cells lie in the core of the islet (comprising 60 – 80% of the islet), and glucagon-producing  $\alpha$  cells, somatostatin-producing  $\delta$  cells, pancreatic polypeptide-producing PP cells, and ghrelin-producing  $\epsilon$  cells lie in the periphery.

One of the critical roles of the pancreas is to maintain blood glucose levels within a narrow range. Primarily, insulin and glucagon secreted by the  $\beta$  cells and  $\alpha$  cells, respectively, are responsible for regulating glucose homeostasis. Glucagon, a catabolic hormone, is produced and secreted in response to a low blood glucose concentration and acts upon the liver to release stored glycogen, thus increasing circulating glucose. Conversely, insulin, an anabolic hormone, is secreted in response to high levels of blood glucose. Once secreted, insulin stimulates glucose uptake by adipose and muscle tissue and acts on the liver to increase glycogen synthesis and inhibit gluconeogenesis (Berne and Levy, 1993), thereby lowering blood glucose levels.



**Figure 1.1 Pancreas anatomy and histology.** A) The pancreas is located in the upper abdomen adjacent to the stomach. B) The exocrine acinar cells and the duct cells comprise 98% of the pancreatic mass. The acinar cells produce digestive enzymes that are carried to the small intestine through a system of ducts. C) The endocrine cells are organized into the pancreatic islets, termed the Islets of Langerhans, which contain five hormone-secreting cell types. D) In the mouse, the insulin-producing  $\beta$  cells (dark purple) primarily lie in the core of the islet and comprise 60 – 80% of the islet, while the glucagon-producing  $\alpha$  cells lie in the periphery and comprise 15 – 20% of the islet. E) Representative H&E stain of the pancreas. The endocrine cells (lighter staining) lie within the dense network of acini (darker staining) and form cell clusters that are highly vascularized. F) Representative immunolabeling of a murine pancreatic islet. The  $\beta$  cells and  $\alpha$  cells are labeled by insulin (red) and glucagon (blue), respectively. The digestive enzyme amylase (green) identifies the acinar cells surrounding the islet. Pancreas schematics adapted from: [www.youtube.com-Dual Role of the Pancreas](http://www.youtube.com-Dual Role of the Pancreas); Pancreas histology adapted from: <http://education.vetmed.vt.edu/>.

## **Diabetes mellitus and β-cell dysfunction**

The regulation of glucose is of particular medical importance as diabetes mellitus is one of the most common metabolic diseases worldwide. The World Health Organization estimates more than 180 million people are afflicted with diabetes worldwide, an estimate likely to double by 2030. In addition to hyperglycemia, a condition in which excessive amounts of glucose circulate in the blood plasma, diabetes is associated with a number of other conditions such as diabetic retinopathy, neuropathy, kidney failure, cardiovascular disease and stroke. As a result, the efforts of numerous investigators focus on elucidating the disease and discovering promising treatments.

Diabetes mellitus can be classified as type 1, type 2, or maturity onset diabetes of the young (MODY) which accounts for a number of hereditary forms of diabetes caused by genetic mutations. Type 1 diabetes, also referred to as juvenile or insulin-dependent diabetes, is a result of an autoimmune destruction of β cells which leads to a loss of insulin production. Alternatively, type 2 diabetes, termed insulin-independent, is the most common form of diabetes, comprising 90 – 95% of those diagnosed. This type is caused by a combination of genetic predisposition and environmental pressures that result in the progressive desensitization of peripheral tissues to insulin (Berne and Levy, 1993). This reduction in sensitivity triggers the existing β cells to produce more insulin, leading to increased desensitization of the β cells, and this cycle ultimately leads to decreased β cell function and β cell death, which results in hyperglycemia.

## **Current treatments and limitations of islet transplantation**

A wide variety of pharmacological agents, including sulfonylureas, meglitinides, biguanides, thiazolidinediones, and alpha-glucosidase inhibitors, are marketed towards type 2 diabetics. These current treatments are fairly effective for maintaining normal blood glucose

levels and controlling acute and chronic complications, such as diabetic ketoacidosis, retinopathy, and neuropathies. Alternatively, exogenous insulin injections are an obligatory requirement for type 1 patients; thus, islet transplantation has been viewed as a promising treatment that can lead to euglycemia in type 1 diabetics (Ryan et al., 2001; Ryan et al., 2005; Shapiro et al., 2000).

The transplantation of pancreatic islets as a treatment for diabetes first gained interest in the early 1970s when Paul Lacy and colleagues revealed that hyperglycemia in diabetic rats could be alleviated by the transplantation of donor islets (Ballinger and Lacy, 1972). In 1990 that work led to the first successful islet cell transplant in humans (Scharp et al., 1990) in which endocrine islets isolated from cadaveric donors were transplanted into the portal vein of type 1 diabetic patients. Throughout the 1990s, successful transplants were documented; however, the Islet Transplant Registry estimated that approximately only six percent of these transplantations were successful (Hering et al., 1996). A significant advance came in 2000, when Shapiro and colleagues reported sustained insulin independence of seven patients with type 1 diabetes following transplantation of donor islets (Shapiro et al., 2000), demonstrating that islet transplantation can reproducibly lead to insulin independence for at least one year.

While the use of islet transplantation may hold promise, there are a number of limitations and complications that arise from this treatment. First, the efficacy of the treatment is dependent upon the survival of engrafted islets which can be affected by both the donor's age and health. However, due to the shortage of organ donors, and thus transplantable pancreatic islets, the ability to discriminate based on donor qualifications is limited. Second, to prevent a host-immune response and reduce transplant rejection, patients are subjected to life-long immunosuppressive therapy. These immunosuppressants can cause a variety of complications, including mouth ulcers, hypertension, diarrhea, anemia, weight

loss, ovarian cysts, and insulin resistance (Hirshberg et al., 2003). These complications highlight the need to assess the risks that diabetic patients face and whether the benefits of a cell therapy outweigh the given risks. Third, in Shapiro's report, the diabetic recipients required islets from at least two donor pancreata to achieve insulin independence (Shapiro et al., 2000). Once again, due to the low rate of organ donation, the shortage of human cadaveric donor islets presents a major limitation, and overcoming the supply and demand issue is a significant determinant concerning the accessibility of this therapy. Lastly, in a later report by Shapiro and colleagues, it was reported that less than ten percent of patients receiving islet transplants remained insulin independent after five years (Ryan et al., 2005). These results, taken with other limitations and complications, highlight the need for improvements if islet transplantation will prove to be an effective therapeutic treatment.

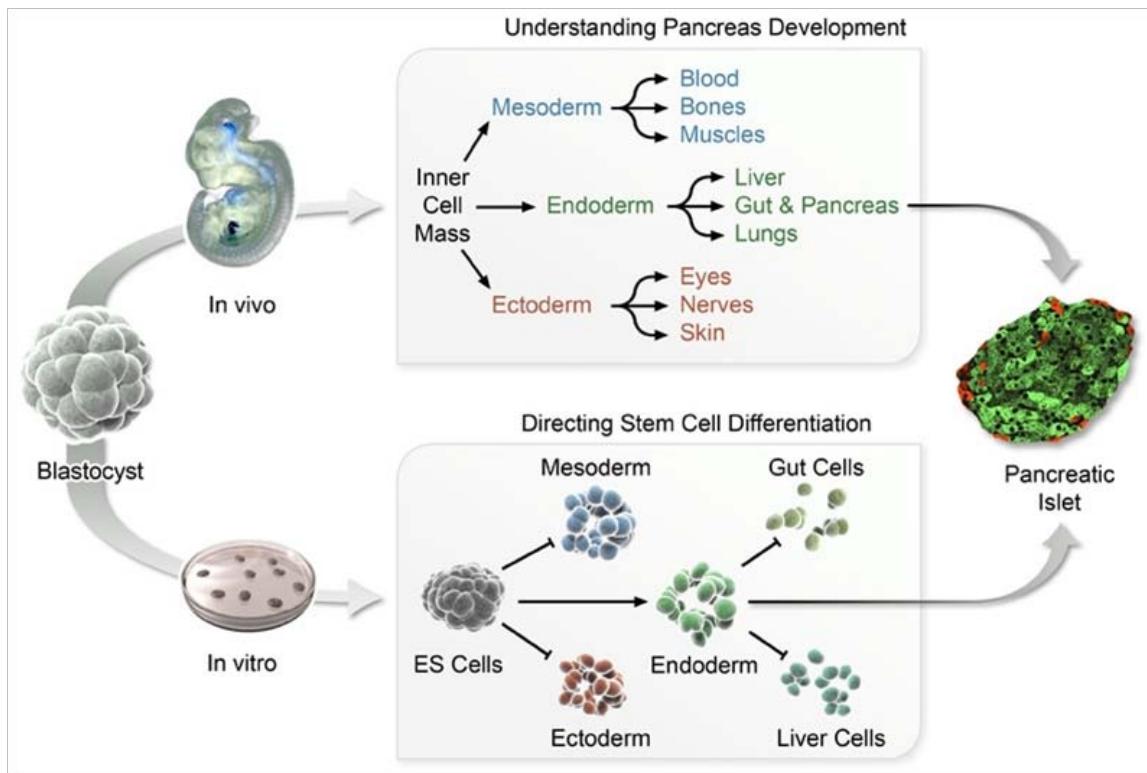
### **Emerging therapies for diabetes: Human embryonic stem (hES) cells as a source of insulin-producing cells**

Due to the limitations and complications of islet transplantation from human cadaveric donors, especially as pertains to the limited supply of donor tissue, there has been much interest in alternative methods for generating insulin-producing  $\beta$  cells. There are numerous alternatives for potential sources of pancreatic insulin-producing cells, including but not limited to self-replication of existing  $\beta$  cells, direct transdifferentiation of committed cells types, and the differentiation of hES cells towards pancreatic fates. Given the ability of human ES cells to differentiate into any cell type and their unlimited supply, the generation of insulin-producing cells from hES cells has been viewed as an appealing alternative.

Human ES cells were first isolated in 1998 by James Thomson at the University of Wisconsin-Madison (Thomson et al., 1998). ES cells are derived from the inner cell mass of a developing blastocyst and are characterized by both their ability to self-renew and their

pluripotency. ES cells can develop into any derivative of the three primary germ layers: ectoderm, endoderm and mesoderm, and because this plasticity, hES cells have become an attractive alternative for regenerative medicine and tissue replacement following disease or injury. Following the initial report regarding their isolation, numerous hES cell lines have been generated and used in a variety of studies. While there has been much enthusiasm concerning the potential use of hES cells in therapeutic applications, there has also been equal opposition arising from the ethical issues introduced by the usage and destruction of fertilized human embryos (Jain, 2005; Robertson, 2010). In spite of this opposition, significant progress has been made in understanding the characteristics of hES cells and the mechanisms by which they can differentiate to various cell types.

Recently, there has been remarkable progress in the directed differentiation of human ES cells towards pancreatic hormone-producing cell fates (D'Amour et al., 2006; Kubo et al., 2004; Phillips et al., 2007). Most notably, Baetge and colleagues at Novocell Inc. reported a five-stage protocol to direct hES cells through a process that mimics normal pancreatic development, taking into account numerous years of research in vertebrate endoderm development and pancreatic differentiation (**Figure 1.2**). The induction protocol led to the production of a number of hormone-expressing endocrine cells, with approximately 7% of final endocrine-specified cells expressing insulin. However, there were some significant deficiencies in these insulin-expressing endocrine cells, such as the lack of glucose-stimulated insulin secretion and improper maturation of the  $\beta$ -like cells. Therefore, in a later publication, the authors chose to assess whether the hES cell-derived pancreatic progenitors, apparent during the fourth stage of the induction, are competent to mature *in vivo* into endocrine cells (Kroon et al., 2008). The gene expression pattern and functionality of the hES cell-derived insulin-positive cells were consistent with mature, functional *Insulin*-expressing cells. These findings present compelling evidence supporting the therapeutic potential of



**Figure 1.2 Schematic of developmental intermediates *in vivo* and corresponding derivation of endocrine cells from ES cells *in vitro*.** The key *in vivo* developmental intermediates of pancreatic endocrine development are shown in the upper panel. To efficiently direct the differentiation of ES cells towards the endocrine lineage, investigators have shown that it is advantageous to mimic normal developmental events. The lower panel depicts the stepwise differentiation of ES cells towards hormone-producing cell fates. Figure from Jean-Philippe Cartailler.

stem cell research in treatments for diabetes and highlight the potential that hES cells hold in generating an abundant and renewable source of glucose-responsive, insulin-producing cells.

### **Embryonic stem (ES) cells, adult stem cells and progenitor cells**

While stem and progenitor cells both possess the ability to differentiate into various specialized cell types that are committed to given fates, they differ in some of their functional characteristics. ES cells are isolated from the inner cell mass of a developing blastocyst and are characterized by their ability to self-renew while maintaining their undifferentiated state, as well as producing daughter cells that can differentiate into committed cell types (McCulloch and Till, 2005). The capacity for stem cells to differentiate into specialized cell types requires them to be pluripotent, in which the cell can give rise to derivatives of the three germ layers but does not contribute to the extraembryonic lineage (Beddington and Robertson, 1989).

While ES cells are characterized by their pluripotency, adult stem cells can only differentiate into a limited number of cell types. Similar to ES cells, adult stem cells possess the ability to self-renew; however, they are typically multipotent or unipotent and can give rise to only a particular lineage or a family of closely related cells. These adult stem cells are evident in the hematopoietic system where multipotent adult stem cells give rise to cells that can replace damaged and aging blood cells or replenish low volumes (Ema and Nakauchi, 2003). Additionally, multipotent adult stem cells in the liver possess the ability to regenerate the majority of the liver's mass following injury or disease (Fausto, 2000; Stanger et al., 2007).

Similar to adult stem cells, progenitor cells are also characterized by their multipotency or unipotency; however, in comparison to a stem cell's ability to self-renew, progenitor cells can divide only a limited number of times before exiting the cell cycle and

committing to a terminal fate (Fuchs et al., 2004; Li and Xie, 2005; Scheres, 2007). Within the context of embryogenesis, numerous organs contain multipotent progenitor cells that are capable to respond to developmental cues and inductive signals which dictate their commitment towards particular fates. While the majority of these tissue-specific progenitor cells are specified during embryogenesis, some progenitor cells persist in adult tissues, such as the epidermal transient amplifying cell population (Clayton et al., 2007).

### **Signaling pathways and genetic regulatory factors directing pancreas development**

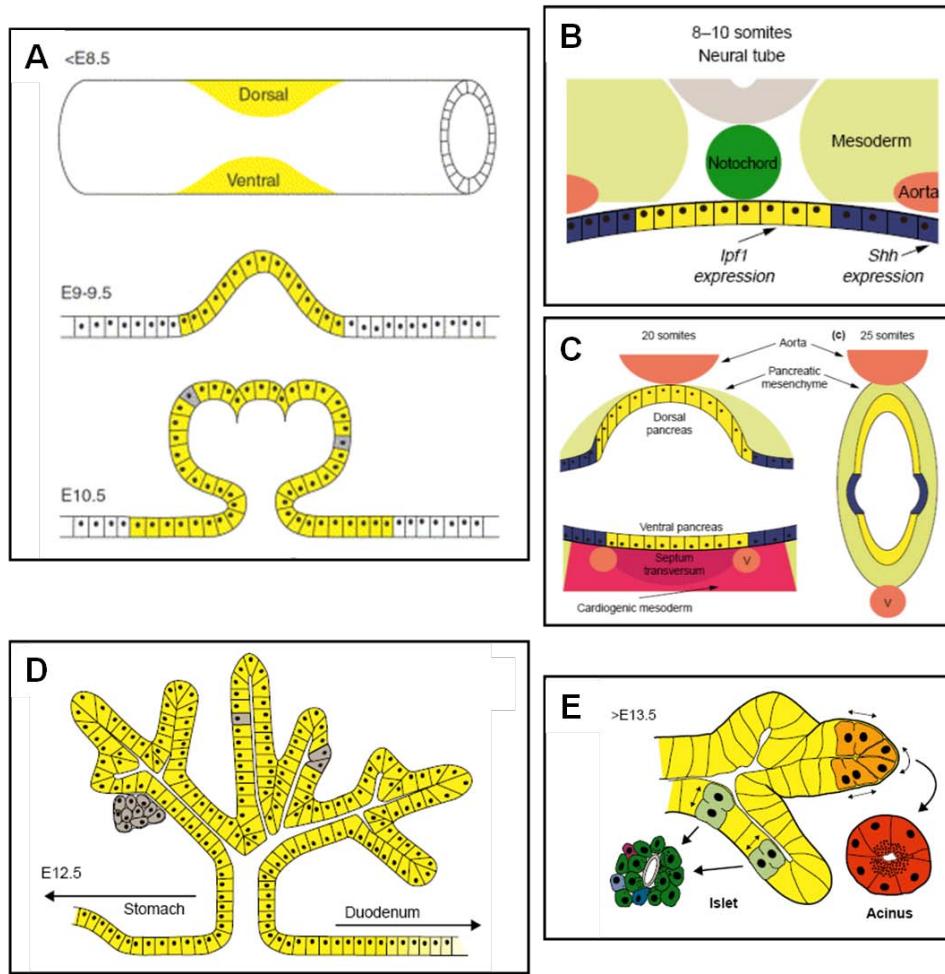
During a process termed gastrulation, the pluripotent epiblast cells become specified to three principal germ layers: the ectoderm, mesoderm, and endoderm. As the epiblast cells migrate through the anterior region of the primitive streak and the node, they undergo an epithelial-to-mesenchymal transition (EMT) and subsequently become either mesoderm or definitive endoderm. The definitive endoderm emerges from this primitive streak and forms a single epithelial sheet, resembling a cup, of approximately 500 – 1,000 cells (Kwon et al., 2008; Wells and Melton, 2000). It has been postulated that the mesoderm and definitive endoderm arise from a bi-potent mesendoderm precursor (Kimelman and Griffin, 2000; Rodaway and Patient, 2001) which is typically characterized by the expression of *Goosecoid* (*Gsc*) or *Brachyury* (*T*). As the cells migrate through the primitive streak, members of the Nodal family of proteins, a subclass of the TGF $\beta$  superfamily, play a critical role inducing and patterning the endoderm and mesoderm.

There are many known transcription factors involved in vertebrate endoderm development, such as factors in the Gata, Sox and Forkhead families. Sox genes encode transcriptional regulators, belonging to a superfamily of proteins characterized by a high mobility group (HMG) DNA-binding domain. Of these factors, Sox17 was first identified in endoderm formation from work in *Xenopus* in which Sox17 was shown to mediate activin-

induced endoderm differentiation in animal caps (Hudson et al., 1997). Later studies characterized the role of Sox17 in mouse endoderm development and revealed its expression in the visceral endoderm as well as the definitive endoderm (Kanai-Azuma et al., 2002). Further confirming its essential role in endoderm formation, Sox17 mutant embryos are deficient of gut endoderm (Kanai-Azuma et al., 2002). In addition to factors belonging to the Sox family, Forkhead genes have been extensively implicated in endoderm formation. Specifically, *Foxa1/HNF3 $\alpha$* , *Foxa2/HNF3 $\beta$* , and *Foxa3/HNF3 $\gamma$*  are expressed during endoderm development (Ang et al., 1993; Monaghan et al., 1993).

As development continues, the single epithelial layer of the endoderm forms into a primitive gut tube which runs along the anterior-posterior axis of the developing embryo, and numerous organs including the pharynx, thyroid, lungs, liver, stomach, pancreas and intestine (Wells and Melton, 1999) are established along an anterior-posterior patterned gut tube endoderm (**Figure 1.3**). The anterior-posterior patterning of the gut endoderm is modulated by numerous signaling pathways, including FGF, Wnt, and retinoic acid. High concentrations of Fgf4 from the overlying mesoderm have a posteriorizing effect on the definitive endoderm (Dessimoz et al., 2006; Wells and Melton, 2000). Similarly, Wnt/ $\beta$ -catenin signaling and retinoic acid from the overlying mesoderm play critical roles in anterior-posterior patterning and serve to posteriorize the gut tube endoderm (Bayha et al., 2009; Kinkel et al., 2009; Li et al., 2008; Martin et al., 2005; McLin et al., 2007; Molotkov et al., 2005; Stafford et al., 2004).

Along the gut tube, the pancreas originates from the posterior foregut endoderm as dorsal and ventral buds which are characterized by their expression of *pancreatic and duodenal homeobox 1 (Pdx1)* and *pancreas specific transcription factor 1a (Ptf1a)* (**Figures 1.3 and 1.4**). The cells in the early pancreatic buds also express *Mnx1/Hlx9* (Harrison et al., 1994; Harrison et al., 1999; Li et al., 1999), *Onecut1/Hnf6* (Landry et al., 1997; Rausa et al.,

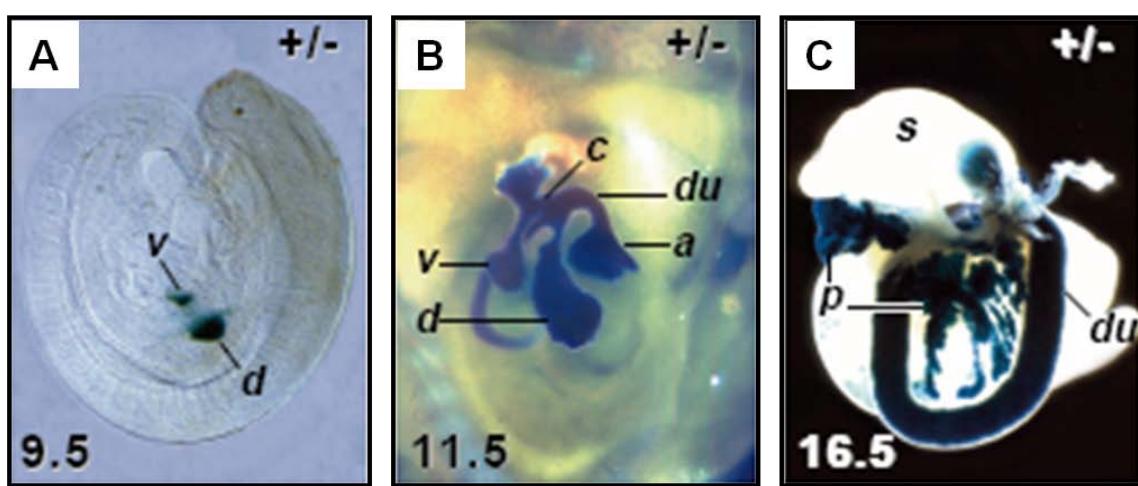


**Figure 1.3 Pancreatic organogenesis: endoderm patterning and branching morphogenesis.** A) Along the gut tube endoderm, the dorsal and ventral pancreatic endoderm are specified. By E9.0, the endoderm evaginates to form the dorsal and ventral pancreatic buds (only dorsal bud depicted), and rapid growth and proliferation cause the epithelium to expand. The pancreatic epithelium consists of multipotent progenitor cells (MPCs, yellow) with periodic endocrine progenitors (grey). B) During early stages of development (8 – 10 somites, E8.5), *Sonic hedgehog (Shh)* expression (blue) is repressed in the dorsal pancreatic epithelium (yellow) by signals from the overlying notochord (green), thereby permitting expression of *Ipf1/Pdx1*. C) By the 20 somite stage (E9.5), the dorsal aorta comes into contact with the dorsal pancreatic epithelium and continues *Shh* repression. Signals from the septum transversum and cardiogenic mesoderm induce *Shh* expression in the endoderm adjacent to the ventral pancreatic endoderm, thus permitting the specification of the ventral endoderm to the pancreatic lineage. D) By E12.5, the dorsal and ventral pancreas fuse and begin to undergo branching morphogenesis and epithelial elongation. Endocrine cell clusters (grey) begin to delaminate from the pancreatic epithelium. E) During the secondary transition (~E13.5), committed cell types begin to form with pancreatic islets (cluster of green, blue, and purple cells) delaminating from the ductal epithelium (yellow) which contains endocrine progenitor cells (light green) while the tips of the branching epithelium (orange) commit to an acinar fate (red). Figure adapted from: Kim and MacDonald (2002).

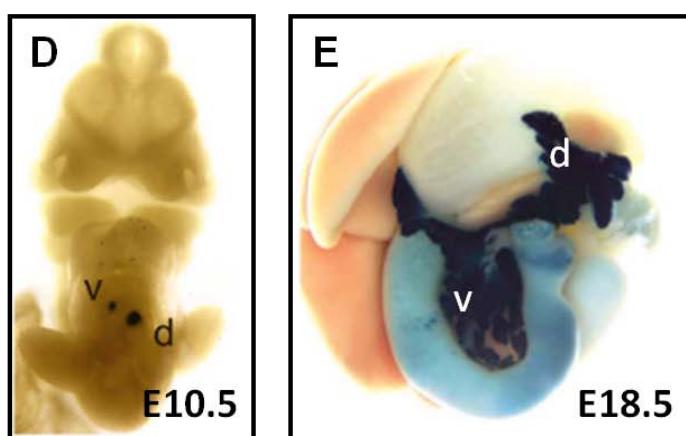
1997), and *Sox9* (Seymour et al., 2007) which all play critical roles in pancreas formation. The pre-pancreatic endoderm receives a variety of instructive and permissive cues from the various mesodermally-derived tissues that it contacts, with the dorsal pancreatic endoderm being in contact with the notochord and dorsal aortae and the ventral endoderm making contact with the cardiac mesoderm, septum transversum mesenchyme, and later, the vitelline veins (**Figure 1.3**) (Deutsch et al., 2001; Kim et al., 1997; Rossi et al., 2001). The notochord secretes a variety of signaling molecules, including the TGF $\beta$  family member, Activin  $\beta$ B, and Fgf2, both of which activate pancreatic gene expression in the dorsal pre-pancreatic epithelium by repressing the expression of *Sonic hedgehog (Shh)* (Hebrok et al., 1998). Additionally, signals from endothelial cells in the dorsal aorta and the vitelline veins promote the specification of the pre-pancreatic endoderm, as evident in *Flk*<sup>-/-</sup> mice, which lack endothelial cells, where *Ptf1a* fails to be expressed in the dorsal pancreatic epithelium (Yoshitomi and Zaret, 2004). Conversely, the ventral pre-pancreatic endoderm diverges from a bipotential progenitor in the ventral foregut that can give rise to both the liver and ventral pancreatic fates (Deutsch et al., 2001). Signals from the cardiogenic mesenchyme, including BMPs produced by the septum transversum and FGFs from the cardiac mesoderm, induce hepatic and pancreatic developmental programs (Jung et al., 1999; Kumar et al., 2003; Rossi et al., 2001). A proper balance of signaling is required, with strong signaling inducing hepatic differentiation and lower levels of signaling from the cardiogenic mesenchyme inducing a pancreatic fate (Deutsch et al., 2001).

Following pancreatic bud formation, the cells further proliferate and branch in response to signals from the adjacent mesenchyme (**Figure 1.3**). While at this point the cells are specified to a pancreatic fate, they are not fully and irreversibly committed to this fate. The majority of the cells within the pancreatic epithelium undergo a primary transition where they enter a “protodifferentiated” state marked by the weak expression of some acinar

*Pdx1*<sup>lacZ/+</sup>



*Ptf1a*<sup>Cre/+; R26R</sup>



**Figure 1.4 *Ptf1a* and *Pdx1* expression during pancreas development.** A – C) Whole mount X-gal histochemistry of *Pdx1*<sup>lacZ/+</sup> embryos. Figure adapted from Offield et al. (1996). A) At E9.5, *Pdx1*<sup>lacZ</sup> expression is evident in the dorsal (d) and ventral (v) endoderm. B) By E11.5, the pancreatic epithelium has expanded and *Pdx1*<sup>lacZ</sup> expression is evident in the antral stomach (a), duodenum (du), dorsal (d) and ventral (v) pancreas, and common bile duct (c). C) By E16.5, *Pdx1*<sup>lacZ</sup> expression is evident in the pancreatic epithelium (p), as well as the stomach (s) and duodenum (du). D – E) Whole mount X-gal histochemistry following *Ptf1a* lineage tracing. Figure adapted from Kawaguchi et al. (2002). D) The progenitor cells contained in the pancreatic epithelium are derived from *Ptf1a*-expressing cells as evident by the lineage label. As compared to *Pdx1*, *Ptf1a* expression is restricted to the dorsal (d) and ventral (v) pancreatic buds within the developing midgut. E) At E18.5, whole mount X-gal histochemistry reveals that the majority of all mature pancreatic cell types were derived from *Ptf1a*-expressing cells.

enzymes and the detection of early differentiated hormone-expressing endocrine cells. Some uncertainty surrounds the exact function and fate of these early endocrine cells with studies both supporting and refuting that these “first wave” endocrine cells can contribute to the mature endocrine pancreas (Gu et al., 2002; Herrera, 2000; Herrera et al., 1994; Lee et al., 1999; Pang et al., 1994; Wilson et al., 2002).

After initial specification and commitment of the pancreatic epithelium, endocrine cell specification begins with the inhibition of Notch signaling in a population of cells, which subsequently express the pro-endocrine gene *Neurogenin 3* (*Neurog3*) (Gradwohl et al., 2000; Jensen et al., 2000; Schwitzgebel et al., 2000). Notch signaling is an evolutionarily conserved pathway and is known principally for its regulation of cell fate decisions; a process termed “lateral specification.” In this pathway, a cell-surface Notch receptor is activated by binding a ligand (e.g. Delta or Jagged) produced by a neighboring cell under the direction of the transcriptional regulator Neurog3. This process leads to the proteolytic release of the intracellular domain of the Notch receptor (NICD) which subsequently binds to Rbpj and thus causes transactivation of target genes, such as Hes1 or related members of the Hairy/Enhancer of Split family which serve as transcriptional repressors that inhibit the expression of pro-endocrine factors. The process of “lateral inhibition” via the Notch signaling pathway is essential for maintaining the undifferentiated state of the pancreatic MPC population, and the inhibition of this signaling mechanism is critical for the widespread commitment of cells to specific cell fates.

The next stage of organogenesis, termed the “secondary transition,” results in the commitment of acinar, ductal and a “second wave” of endocrine cells from the protodifferentiated pancreatic epithelium (**Figure 1.3**). During this transition, *Neurog3*-expressing cells are scattered throughout the epithelial cords of the pancreatic epithelium and identify endocrine progenitor cells that give rise to the hormone-expressing cells of the

mature pancreas. Along with Neurog3, various other transcription factors identify a subset of endocrine progenitor cells, including Pax6, Nkx6-1, Nkx2-2, Insm1 and Neurod1. As the epithelium expands and endocrine cells become specified, the tips of the epithelial branches begin to commit to the acinar lineage and differentiate into pro-acinar cell clusters. Acinar cell commitment is dependent on the trimeric complex PTF1, which is composed of Ptf1a, a ubiquitously expressed class A bHLH protein, and either Rbpj or Rbpjl (Beres et al., 2006; Roux et al., 1989; Sommer et al., 1991). During early stages of pancreatogenesis, Ptf1a interacts with Rbpj and this interaction is essential for early progenitor cell specification and development (Masui et al., 2007). Subsequently, Rbpjl displaces Rbpj within the PTF1 complex and promotes the specification to the acinar lineage (Masui et al., 2010).

### Pancreatic multipotent progenitor cells

During the initial evagination of the foregut endoderm to form the dorsal and ventral pancreatic buds, the pancreatic epithelium consists of multipotent progenitor cells (MPCs) that give rise to all three cell types of the mature pancreas: endocrine, acinar and duct (Gu et al., 2002; Herrera, 2000; Kawaguchi et al., 2002; Zhou et al., 2007). While the nature of these pancreatic MPCs is not fully understood, it is thought that they exist within the pancreatic epithelium from E9.5 until approximately E12.5 (Zhou et al., 2007). A deeper understanding of the specification and differentiation of the pancreatic MPCs is important for multiple reasons. First, the number of pancreatic MPCs present early in development determines the final size of the pancreas (Stanger et al., 2007). Second, the formation of the mature pancreas, including the endocrine islets, depends upon the orderly expansion of the MPC-containing, pre-pancreatic epithelium. Lastly, a better understanding of the signaling pathways that promote progenitor cell specification and establish the genetic regulatory network *in vivo* will

be essential for learning how to mimic these processes during the directed differentiation of pluripotent stem cells towards pancreatic fates *in vitro*.

*Pdx1* and *Ptf1a* are both expressed in pancreatic MPCs; however while they are co-expressed in a subpopulation of MPCs, they also have distinct spatially defined expression domains within the developing midgut (**Figure 1.4**). From E10.5 – E11.5, *Pdx1* is broadly expressed in the epithelium of the caudal stomach, rostral duodenum, common bile duct and extrahepatic biliary ducts, which limits its use for exclusively identifying pancreatic MPCs (Burlison et al., 2008; Offield et al., 1996). Conversely, at this time within the developing midgut, *Ptf1a* is expressed exclusively within the dorsal and ventral pancreatic epithelium (Burlison et al., 2008; Kawaguchi et al., 2002). While both *Pdx1* and *Ptf1a* mutant mice are apancreatic, *Pdx1*-deficient mice also have malformation of the gastro-duodenal junction, defects in the development of the submucosal Brunner's glands, altered enteroendocrine cell numbers in the stomach and duodenum, and abnormalities in the formation of the peribiliary glands and mucin-producing cells of the gall bladder (Burlison et al., 2008; Jonsson et al., 1994; Kawaguchi et al., 2002; Offield et al., 1996).

### **Genetic manipulations in mice and cells**

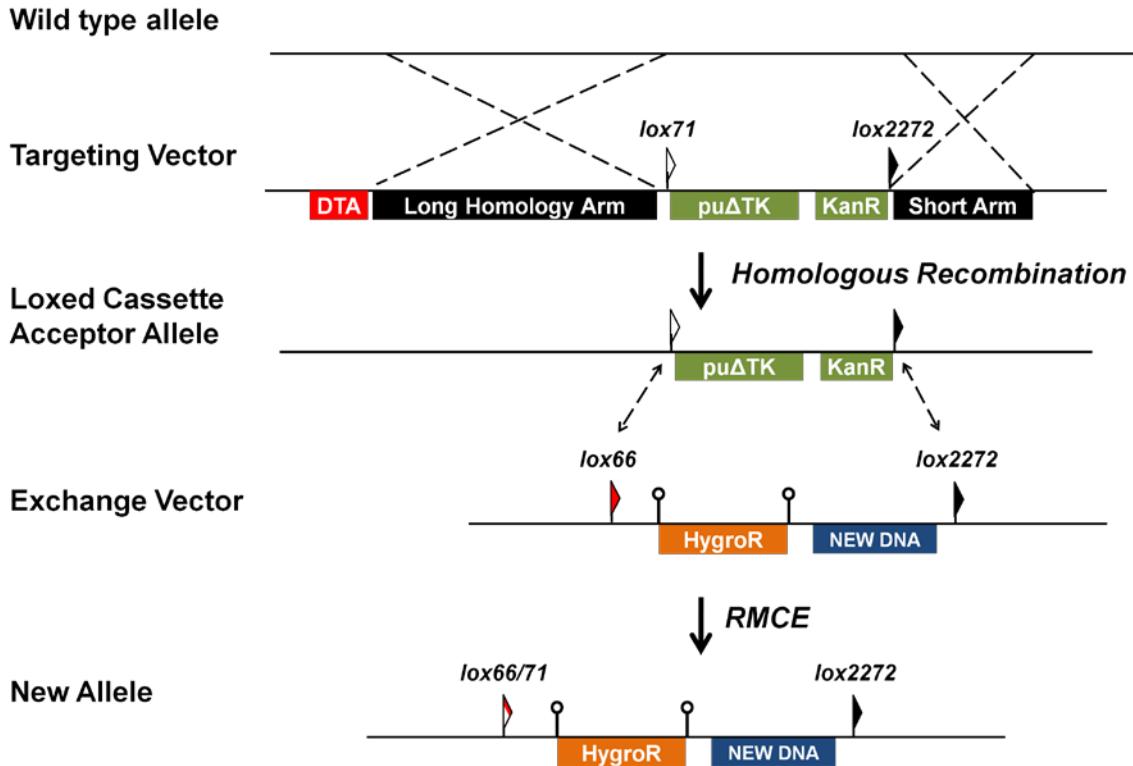
Genetic engineering has proven widely useful for a variety of applications in developmental biology. Specifically, transgenic mice have become valuable models for examining genetic disorders and elucidating embryonic development. Transgenic mice are generated via microinjection of exogenous DNA that incorporates the regulatory elements of a particular gene of interest driving the expression of a mutant version of the coding region into the pronucleus of a fertilized egg (Gordon and Ruddle, 1981; Wagner et al., 1981). However, advances in molecular biology and stem cell biology have permitted the generation of genetically engineered mice through gene targeting in mouse embryonic stem (mES) cells

which hold numerous advantages as compared to pronuclear injection for generating transgenic mice (Smithies et al., 1985; Thomas et al., 1986).

The method of gene targeting relies on homologous recombination in ES cells. The technique was first developed for facilitating site-directed mutagenesis in yeast but has subsequently been adapted for mammalian cells. This technique relies on the generation of a DNA construct which incorporates several kilobases (kb) of DNA sequence homologous to the mouse genome thereby permitting recombination between the homologous sequences (**Figure 1.5**). The DNA construct also contains desired gene modifications, as well as sequences conferring drug sensitivity to permit identification of recombination events. In mammalian genomes, homologous recombination at the correct locus occurs at a very low rate, thus labeling gene targeting as a more laborious method.

In order to accelerate the repetitive generation of mutant and/or reporter alleles for a given gene, strategies that utilize recombinase-mediated cassette exchange (RMCE) have been developed and implemented (**Figure 1.5**). This technique relies on the generation of a loxated cassette acceptor (LCA) allele in ES cells, which is achieved through standard homologous recombination and gene targeting. The LCA allele contains loxP or mutant loxP sites that facilitate the exchange of a DNA cassette containing compatible lox sites. Various mutant lox sites have been described and each present their own advantages and disadvantages, and the use of heteromeric lox sites, such as the compatible lox 66 and lox71 sites, allow more efficient and unidirectional RMCE (Araki et al., 2002).

The sequences removed during the generation of the LCA allele can be manipulated in a variety of ways, thus generating knockout alleles, knockin alleles, point mutations in the sequence, fluorescently-tagged alleles, or insertions of other elements, such as HA-tags for immunoprecipitation or Cre recombinase for lineage labeling. The modified DNA sequence is incorporated into an exchange cassette containing lox sites compatible with the lox sites



**Figure 1.5 Scheme for gene targeting and recombinase-mediated cassette exchange (RMCE).** First, a targeting vector is generated to replace a region of interest of a wild type allele with a desired DNA sequence. For the generation of a loxed cassette acceptor (LCA) allele, this sequence typically contains a dual selection cassette consisting of a fusion of *puromycin resistance* and a mutant *thymidine kinase* driven by the mouse *phosphoglycerol kinase* promoter (*pgk-pu $\Delta$ tk*) and an EM7-driven *kanamycin resistance* (EM7-Kan<sup>R</sup>) flanked by lox71 and lox2272 sites. To facilitate homologous recombination between the targeting vector and the wild type allele, long and short homologous arms are incorporated into the targeting vector and flank the desired DNA sequence. Additionally, the targeting vector contains a *pgk*-driven *diphtheria toxin A* gene (*DTA*) outside the long homology arm for negative selection following targeting. In order to generate a new allele of interest that contains a reporter or mutated DNA sequence, an exchange vector is generated that has lox66 and lox2272 sites which are compatible with the lox sites present in the LCA allele. The exchange cassette contains an FRT-flanked *pgk-Hygro<sup>R</sup>* cassette for positive selection of mES cells after RMCE. Co-electroporation of ES cells with both the exchange vector and a Cre-expressing vector permit the recombination of the compatible lox sites and exchange of the cassette into the LCA allele thereby yielding the new allele of interest.

present in the LCA allele. RMCE is performed by co-electroporating the exchange cassette as well as a Cre-expression vector into ES cells containing the LCA allele of interest. Site-specific recombination between the compatible lox sites in the LCA allele and the exchange cassette is driven Cre recombinase and results in the exchange of the modified cassette into the LCA allele. A staggered positive-negative selection strategy permits the identification of properly exchanged clones (Long et al., 2004).

While RMCE relies on the generation of an LCA allele, which requires the more laborious gene targeting method, the availability of a collection of LCA alleles greatly lessens the effort needed to generate new locus-specific alleles as compared to repetitive gene targeting. Indeed, once an LCA allele is generated, RMCE is significantly more efficient than repetitive gene targeting for the generation of modified alleles. Additionally, this technique is highly advantageous as it permits the insertion of different types of mutations into a precise location in the genome, thus resulting in the ability to study allelic variants in a controlled manner.

## **Transcriptomics**

The transcriptome is the complete set of all messenger RNAs (mRNAs), or transcripts, in a cell (Su et al., 2002). Messenger RNA is transcribed from a DNA template and is eventually translated into a protein product. Unlike the genome, which is encoded by DNA, the transcriptome can vary depending on external conditions, such as environmental influences and developmental timing. Since the transcriptome includes all mRNA transcripts in the cell, it reflects the genes being actively expressed at a given time and therefore governs a cell's development. Examining the transcriptome is critical for understanding the process of cellular differentiation, analyzing pathways that regulate development, and interpreting the functional components of the genome. In addition, comparing the gene expression profiles of

diseased and normal cells/tissues will shed light on the pathology and development of certain diseases and potentially identify new targets for therapeutic intervention.

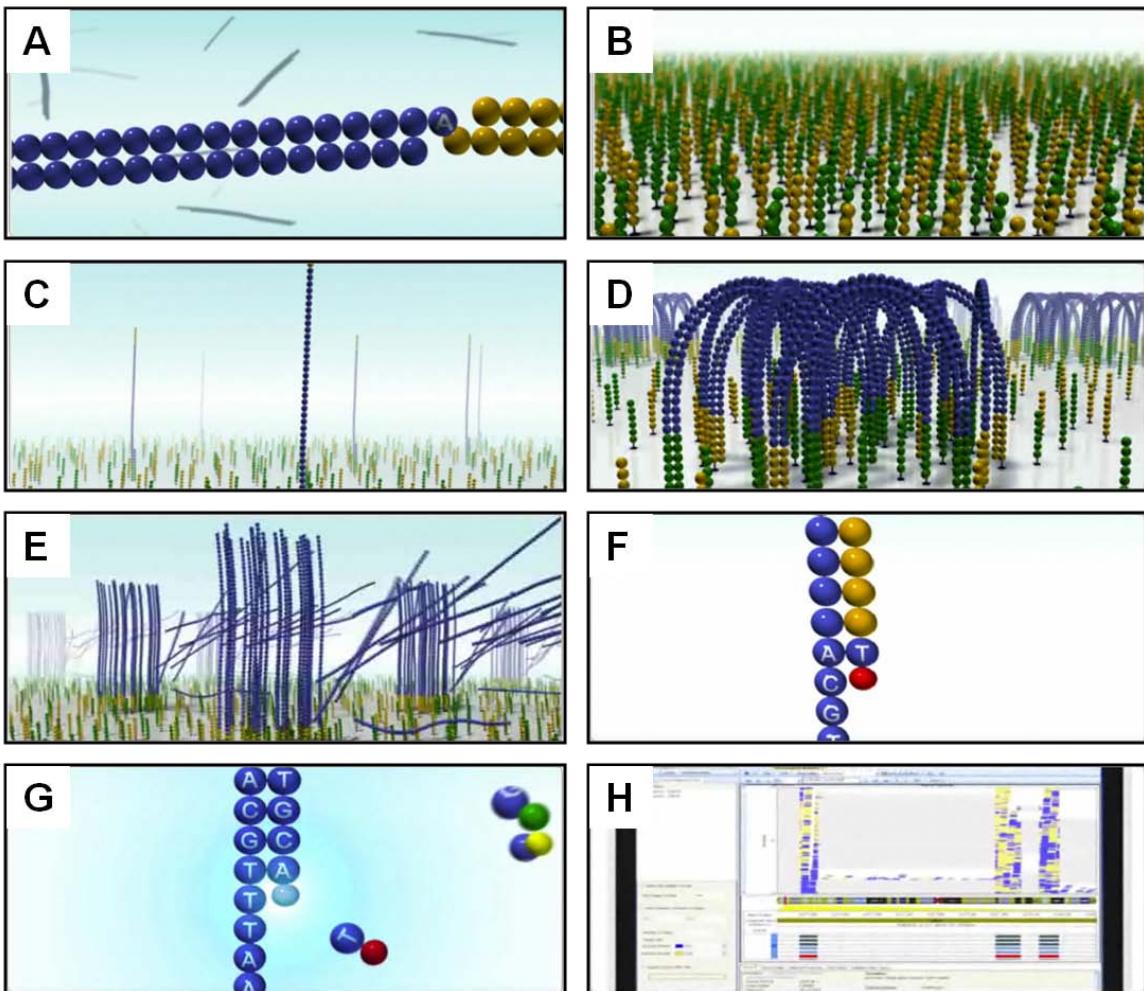
The importance of the transcriptome has made expression profiling an active field. Numerous technologies have been developed to examine the expression level of mRNAs in a given cell, including hybridization- and sequence-based approaches. Hybridization-based approaches, such as microarrays, have been the most commonly used method for transcriptome analysis (Schena et al., 1995). These methods rely on the hybridization of short DNA fragments that are used as probes to fluorescently-tagged target samples, typically cDNA. Hybridization of the probe-target is quantified by detection of the fluorophore-labeled target and used to determine the relative abundance of a particular sequence. However, despite their common use, microarrays have a number of disadvantages that limit their potential, including prior knowledge of genome sequence, limited range of transcriptome coverage, and high background levels due to cross-hybridization (Malone and Oliver, 2011). In contrast, sequence-based methods directly determine the target sample sequence. Developed in 1991, expressed sequence tag (EST) sequencing was a valuable tool for gene discovery (Adams et al., 1991). EST sequencing was critical in driving the human genome project which produced a reference sequence for the human genome (Venter et al., 2001). The method relies upon Sanger sequencing of short cDNA fragments, approximately 500 – 800 nucleotides, which can then be mapped to specific chromosome locations. While a valuable tool for gene discovery, the method is relatively low throughput, dependent on bacterial cloning constraints, not conducive to rare-transcript discovery, and requires costly large scale sequencing (Alba et al., 2004).

In 1995, Velculescu et al. developed a tag-based approach for transcriptome analysis, termed serial analysis of gene expression (SAGE) (Velculescu et al., 1995). Numerous variations of this approach have been developed since its inception, including LongSAGE

(Saha et al., 2002), Robust-LongSAGE (RL-SAGE) (Gowda et al., 2004), SuperSAGE (Matsumura et al., 2005), CAGE (cap analysis of gene expression) (Kodzius et al., 2006; Shiraki et al., 2003), and MPSS (massively parallel signature sequencing) (Brenner et al., 2000). While similar to EST sequencing, SAGE is dependent on generating short sequence “tags,” approximately 9 – 15 nucleotides (nt) in length, which are then concatenated to allow for efficient sequencing. Although more efficient than large-scale EST sequencing, this method is also laborious and dependent on a rather expensive Sanger sequencing. However, in comparison to hybridization approaches, sequence-based methods allow for absolute quantification of gene expression levels.

### **Whole Transcriptome Profiling: RNA-Sequencing technology**

Recent advances in genome sequencing technologies have caused a new method, termed RNA-Seq, to propel forward. Sequencing platforms and methods developed by Illumina (Bennett et al., 2005), Life Science (Margulies et al., 2005), and ABI (Shendure et al., 2005) have numerous advantages over traditional microarrays, including independence from reliance on existing genomic sequence, quantitative gene expression levels, single-base resolution, relatively low background noise, greater dynamic range and reproducibility, low input RNA requirement, and an increased ability to distinguish different isoforms. Similar to SAGE, RNA-Seq is dependent on generating sequence “tags” that can range from 25 – 100 nucleotides (**Figure 1.6**). These tagged sequences are ligated to a flow cell where clonal sequence clusters are generated for numerous rounds of sequencing. The sequence read is subsequently aligned to a genome of interest where the transcriptional profile of a tissue or cell can be examined. RNA-Sequencing is expected to revolutionize transcriptomics (Wang et al., 2009) and has already been applied to numerous species including *Saccharomyces cerevisiae* (Nagalakshmi et al., 2008), *Arabidopsis thaliana* (Lister et al., 2008), human



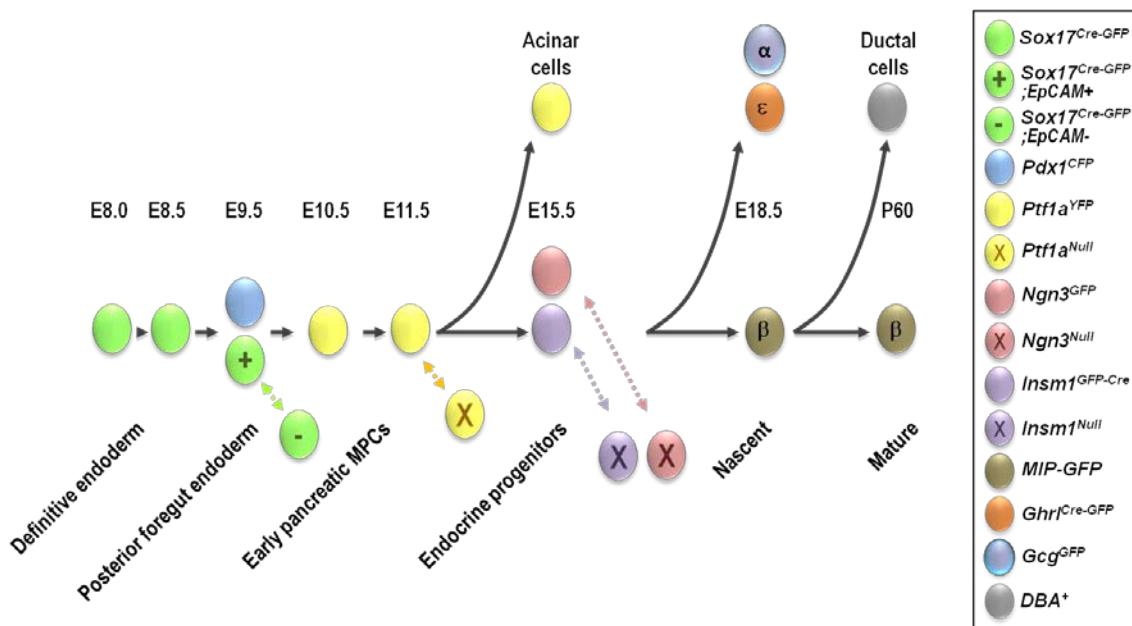
**Figure 1.6 Scheme of Illumina RNA-Sequencing.** A) cDNA (blue) is randomly fragmented and ligated to adaptor oligos (yellow). B) The flow cell has a dense lawn of oligos grafted to the surface. C) The adaptor-ligated cDNA fragments hybridize to the lawn of oligos. D) The bound cDNA fragments are clonally amplified through a series of extensions and isothermal bridge amplifications resulting in millions of unique clusters. E) The reverse stands are cleaved and removed. F) The clusters are sequenced simultaneously where four fluorescently labeled nucleotides compete to bind to the template. G) After each round of synthesis, the clusters are excited by a laser which excites the fluorophore. The emission fluorescence identifies the newly added base. H) Following sequencing, short fragment reads are aligned to the genome of interest and transcript expression values can be quantified. Figure adapted from [http://www.illumina.com/technology/sequencing\\_technology.ilmn](http://www.illumina.com/technology/sequencing_technology.ilmn).

(Sultan et al., 2008), and mouse (Cloonan et al., 2008). While RNA-Seq is still an emerging technology, its advantages over previous methods are clear, and this technology is becoming the predominant method for transcriptional analysis of defined cellular populations.

## Overview and Aims of Dissertation

To enhance cell culture based efforts that aim to generate  $\beta$  cells, a deeper understanding of the native cell types would prove beneficial. While studies of early human organ development would be optimal for such comparisons, these studies are typically limited by the ability to obtain properly staged early human fetal tissues. Although there are developmental differences between human and mouse, gene expression studies in mice have proven useful in guiding the directed differentiation of human ES cells (Bu et al., 2009; Yang et al., 2008). Concerning the directed differentiation of hES cells towards hormone-expressing cells, the most promising methods have attempted to recapitulate the endogenous signaling pathways and transcriptional networks that guide pancreas development (discussed in Chapter II). Thus, it is only logical that a deeper understanding of the various developmental stages would serve to accelerate the efforts pertaining to the directed differentiation of hES cells towards  $\beta$ -like cells.

To facilitate development of improved human ES cell directed pancreatic differentiation protocols, the research described herein represents only a portion of a larger scheme focused on generating and characterizing a series of high quality transcriptional profiles representing key stages in the generation of pancreatic endocrine cells that occur naturally in the mouse (**Figure 1.7**). By using fluorescence-activated cell sorting (FACS) to isolate specific pancreatic progenitor cell populations and by applying whole transcriptome sequencing (RNA-Seq) to these populations, we are obtaining a logical series of transcriptional profiles from specific developmental intermediates in the mouse. Analysis of



**Figure 1.7 Developmental scheme of  $\beta$  cell development and target cell populations for transcriptional profiling.** The graphical representation outlines the developmental pathway underlying  $\beta$  cell differentiation. The definitive endoderm gives rise to a number of organs, including the pancreas. Following formation of the gut tube, the pancreatic endoderm is specified within the posterior foregut endoderm. The early pancreatic epithelium contains multipotent progenitor cells (MPCs) which give rise to all three mature pancreatic cell types: endocrine, acinar and duct. As development continues, the MPCs give rise to endocrine progenitors which will eventually differentiate into one of the five hormone-expressing cell types: insulin-producing  $\beta$  cells, glucagon-producing  $\alpha$  cells, somatostatin-producing  $\delta$  cells, pancreatic polypeptide-producing PP cells, and ghrelin-producing  $\epsilon$  cells. Each developmental intermediate can be identified by the expression of specific transcripts, and we are utilizing fluorescently-tagged alleles to isolate a number of these developmental stages. In addition to analyzing cells that express a transcript of interest, we are isolating and characterizing cells that are genetically deficient (denoted by X). The developmental intermediates and cell populations represented here will allow us to analyze the temporal changes of cells as they differentiate towards a mature  $\beta$  cell fate. Figure from Eunyoung Choi.

these datasets will likely lead to the identification of signaling pathways and gene clusters essential for the formation of functional pancreatic  $\beta$  cells in the mouse, thereby stimulating new hypotheses for identifying pro- $\beta$  cell signals necessary to direct the differentiation of human ES cells into pancreatic  $\beta$  cells.

Our approach has many advantages over prior studies (Chiang and Melton, 2003; Gu et al., 2004; White et al., 2008). First, by utilizing fluorescently-tagged alleles, we are able to isolate genetically defined cell populations. This is a major improvement over previous studies that often relied solely on manual dissection of pancreatic regions and thus represent cell populations possibly contaminated by other non-desired cell types. Second, by utilizing the recently developed RNA-Sequencing approach for whole transcriptome profiling instead of DNA microarrays, we have obtained a less biased and more quantitative view of the gene regulatory changes occurring (Marioni et al., 2008; Mortazavi et al., 2008). Lastly, by collecting information from closely related cell types, we will be able to more accurately assess the spatial and temporal alteration of genetic networks that regulate the step-wise production of  $\beta$  cells, which was not achieved by previous efforts (Gu et al., 2004; White et al., 2008).

To define the temporal, spatial and transcription factor-dependent gene expression of specific developmental intermediates, we have utilized genetically engineered mouse lines to obtain highly purified cell populations for deep sequencing of their transcriptomes. Specifically, we have sought to examine the genetic and temporal requirement of *Ptf1a*, a marker of the pancreatic MPC population and of acinar-specified cells, during pancreas development (discussed in Chapter III). By comparing different transcriptional profiles, we have been able to identify five gene clusters that demonstrate the formation of specific cell lineages. The datasets from these cell populations have identified transcription factors, cell surface receptors, and enzymes that possibly have a role in MPC specification and pancreas

development. Specifically, the analysis of datasets such as these permit the identification of differentially expressed transcripts, such as *Nephrocan* (discussed in Chapter IV), an inhibitor of TGF $\beta$  signaling pathway, that may have a novel role in pancreatic progenitor cell specification or a broader role in pancreas development. Additionally, we have sought to generate fluorescently-tagged alleles to facilitate the combinatorial sorting of *Pdx1*- and *Ptf1a*-expressing cells which will permit analysis of the spatially-distinct expression domains of these two transcripts (discussed in Chapter V).

We anticipate that further analysis of datasets such as these will lead to identification of novel transcriptional effectors which guide the sequential developmental program of  $\beta$  cell differentiation. Parallels/analogies can then be drawn for identifying and isolating human ES cell-derived pancreatic cells and for distinguishing different progenitor cells during the directed differentiation of human ES cells towards a  $\beta$  cell fate.

## CHAPTER II

### DIRECTED DIFFERENTIATION OF HUMAN EMBRYONIC STEM CELLS AND GENERATION OF REPORTER ALLELES

#### Introduction

##### Directed differentiation of human ES cells towards pancreatic fates

In order to generate insulin-producing  $\beta$ -like cells via directed differentiation of human ES cells, it seems necessary to recapitulate, as precisely as possible, normal developmental processes and events. Based on knowledge obtained largely from studies using mice, at minimum it is necessary to sequentially induce the formation of 1) anterior primitive streak cells, 2) definitive endoderm, 3) primitive gut tube, 4) posterior foregut endoderm, 5) pancreatic endoderm and lastly 6) pancreatic  $\beta$  cells (**Figure 1.2**). These defined developmental stages are governed by numerous signaling pathways and molecules that sequentially and spatially promote, or in some cases inhibit, the differentiation of cells to specific fates and serve to establish critical gene regulatory networks that instruct cell development.

There has been remarkable progress by numerous investigators concerning the directed differentiation of hES cells towards hormone-producing fates (Assady et al., 2001; Baharvand et al., 2006; Brolen et al., 2005; Cho et al., 2008; D'Amour et al., 2006; Jiang et al., 2007a; Jiang et al., 2007b; Phillips et al., 2007; Segev et al., 2004; Shim et al., 2007). Most notably, Baetge and colleagues reported a five-stage protocol to direct hES cells through a process that mimics normal pancreatic development, taking into account numerous years of research in vertebrate endoderm development and pancreatic organogenesis

(D'Amour et al., 2006). While the induction produced hormone-expressing endocrine cells, only seven percent of the final endocrine-specified cells expressed *Insulin*. Additionally, there are some significant deficiencies in the  $\beta$ -like cells including improper maturation, as evidenced by a number of polyhormonal cells, as well as the inability to secrete insulin in response to glucose. To circumvent these shortcomings, a later publication reported the competency of hES cell-derived pancreatic progenitors, which are formed during the fourth stage of differentiation, to develop *in vivo* into mature endocrine cells (Kroon et al., 2008). The gene expression pattern and functionality of the *in vivo*-matured insulin-producing cells were similar to mature, functional  $\beta$  cells. However, the *in vivo* maturation of the transplanted heterogeneous population, which quite possibly includes some undifferentiated pluripotent cells and/or cells differentiated to other lineages, resulted in teratoma formation, exposing the critical need for more efficient differentiation protocols and adequate purification of the pancreatic progenitors which are implanted.

While these reports present compelling evidence supporting the therapeutic potential of ES cell research in treatments for diabetes, the ability to generate glucose-responsive insulin-producing cells from hES cells solely from *in vitro* differentiation has yet to be achieved. The limited success stems not only from inefficient differentiation but also from a lack of knowledge concerning the gene regulatory networks established as the cells transition through specific developmental intermediates and the three-dimensional affects leading to cell-to-cell signaling *in vivo*. Further advances could be made if we could purify hES cell-derived developmental intermediates which then could be compared to corresponding authentic intermediates isolated *in vivo* or potentially could be transplanted into recipients without adverse teratoma formation. To overcome these limitations and improve hES cell induction protocols, the utilization of reporter alleles in human ES cells may be vital.

## **Reporter alleles in human ES cells**

The generation of hES cell lines that express reporter cDNAs, such as fluorescent proteins, under transcriptional control of genes that are expressed during specific stages of pancreas development would enable the differentiation of hES cells to be quantitatively monitored and permit the isolation of cell populations that mark specific developmental intermediates. Reporter alleles such as these would allow investigators to analyze the veracity of *in vitro* hES cell differentiation towards pancreatic fates and gain a better understanding of the signals that direct differentiation as well as the transcriptional networks established in these intermediates. To generate reporter alleles in hES cells, investigators have primarily utilized gene targeting (Fischer et al., 2010; Irion et al., 2007; Ruby and Zheng, 2009). Although this method is somewhat laborious, gene targeting eliminates positional effects typically seen in random transgenic insertions, and it enables accurate reporter expression as transcription of the reporter is controlled by endogenous regulatory sequences of the chosen locus.

An alternative approach that can be utilized to generate reporter alleles in ES cells is recombinase-mediated cassette exchange (RMCE). Investigators have reported the generation of a human *ROSA26* allele containing a tandem-dimer red fluorescent protein (tdRFP) flanked by lox sites in hES cells, thus generating an allele that can serve as a cassette acceptor for RMCE (Irion et al., 2007). These cells permit an exchange cassette, containing DNA sequences of interest flanked by compatible lox sites, to be exchanged into the *ROSA26* allele. While the generation of a reporter allele by the exchange of a DNA cassette into the *ROSA26<sup>LCA</sup>* allele is less difficult than gene targeting, it is not without certain limitations. In order to generate a reporter driven by the transcriptional regulatory elements of a particular locus rather than the regulatory elements of *ROSA26*, it is necessary to incorporate those regulatory sequences into the exchange cassette. This presents a major drawback given that

the promoter and enhancer sequences incorporated may not fully recapitulate endogenous expression. Thus, to ensure that critical regulatory regions are incorporated into the exchange cassette for generation of a reporter allele via this method, it is important to have a deep understanding of the regulatory elements driving its expression.

The ability to insert reporters, such as fluorescent proteins, into specific loci revolutionized cell and molecular biology. By using a fluorescent reporter under the transcriptional control of a gene of interest, the developmental expression patterns for numerous genes were elucidated. This technology allowed for examination of temporal and spatial expression patterns in mouse tissues, the identification of new cell types based on specific expression patterns, further characterization of single cells, and transcriptional profiling of isolated cell populations. Similarly, gene targeting in hES cells will be critical for analyzing gene function *in vitro* and examining the function of specific genes. Not only will homologous recombination permit the isolation of fluorescently-tagged cells from hES cell-derived heterogeneous populations, it will also allow for the generation of hES cell lines that contain specific mutations or polymorphisms that will further elucidate the pathogenesis of particular diseases or the physiology of tissues of interest. Regardless of its particular application, homologous recombination and the generation of reporter alleles in hES cells will promote a better understanding of specific genes and accelerate the therapeutic potential of human ES cells.

## Synopsis

Here, we sought to determine the efficiency of the previously published five-stage pancreatic differentiation protocol on the federally-available WA-09 hES cells. We examined the gene expression pattern of the WA-09 hES cells as they transitioned through distinct pancreatic developmental intermediates. Given that the purification of hES cell-derived

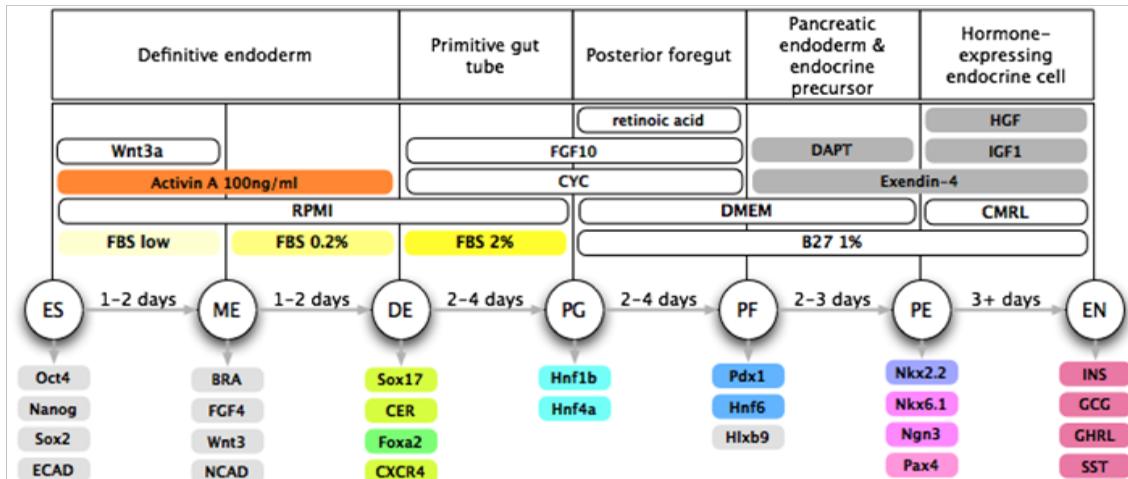
developmental intermediates will prove critical for making further advances, we sought to generate fluorescently-tagged reporter hES cell lines through both RMCE and gene targeting. These reporter cell lines will enable the differentiation protocol to be quantitatively monitored and also permit the isolation of distinct developmental intermediates which then can be compared to corresponding authentic intermediates isolated *in vivo* (discussed further in Chapter III).

## Materials and Methods

### Human ES cell culture and differentiation to pancreatic fates

WA-09 human ES cells (Thomson et al., 1998) were cultured on gelatin-coated tissue culture dishes on a layer of irradiated mouse embryonic fibroblasts (MEFs). Routine culture media consisted of DMEM-F12, 20% Knockout Serum Replacement, 1X non-essential amino acids, 2 mM L-glutamine, 10 ng/ml basic-FGF, 0.1 mM  $\beta$ -mercaptoethanol, 25 units/ml penicillin and 25  $\mu$ g/ml streptomycin (P/S) (all from Gibco/Invitrogen). Cells were passaged every five to seven days with 1 mg/ml collagenase (Invitrogen).

For differentiation, cells were cultured on a layer of MEFs in 24-well plates until they reached approximately 70% confluence. Prior to differentiation, cells were washed briefly with 1X PBS with  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$  (Gibco) and then differentiated for 15 days similar to the five-stage protocol described previously (**Figure 2.1**) (D'Amour et al., 2006). Cells were cultured in RPMI, DMEM, or CMRL media supplemented with varying concentrations of fetal bovine serum (FBS) (Hyclone) or 1% B27 supplement (Invitrogen). Cells were stimulated with combinations of 100 ng/ml Activin A (R&D Systems), 25 ng/ml Wnt3a (R&D Systems), 0.25  $\mu$ M KAAD-Cyclopamine (Toronto Research Chemicals), 50 ng/ml



**Figure 2.1 Schematic diagram of various cell intermediates, transcripts expressed during development towards the endocrine lineage, and outline of differentiation procedure including media, growth factors and times of exposure.** The differentiation protocol is divided into five stages that mark various developmental intermediates that embryonic stem (ES) cells transition through as they develop into endocrine cells, including definitive endoderm (DE), primitive gut tube (PG), posterior foregut (PF), pancreatic endoderm and endocrine precursor (PE), and hormone-expressing endocrine cell (EN). In stage 1, pluripotent hES cells are transitioned through the mesendoderm (ME) to a definitive endoderm intermediate using high concentrations of Activin A and Wnt3a. In stage 2, Activin A is removed and the cells are transitioned to a state resembling the primitive gut tube by exposure to FGF10, or FGF7, and the hedgehog-signaling inhibitor cyclopamine (CYC). In stage 3, the gut tube-like cells are additionally induced with retinoic acid which promotes the expression of posterior foregut markers. In stage 4, the cells are induced to pancreatic endoderm fates by exposure to a  $\gamma$ -secretase, Notch-pathway inhibitor (DAPT) and Exendin-4. During the final stage of differentiation, expression of pancreatic hormones as well as other markers of mature endocrine populations is evident when additionally induced with hepatocyte growth factor (HGF) and insulin-like growth factor (IGF1). Each developmental intermediate is characterized by the expression of specific transcripts; gene symbols are listed below each developmental stage. Figure adapted from D'Amour et al. (2006) by Jonathan Schug.

FGF7 (R&D Systems), 2 µM retinoic acid (Sigma), 1 µM DAPT (Sigma), 50 ng/ml Exendin-4 (Sigma), 50 ng/ml IGF-1 (Sigma), and 50 ng/ml HGF (Peprotech).

### **TaqMan Low Density Array Design**

To analyze changes in gene expression as differentiation proceeded, we utilized customized TaqMan low density arrays (TLDA). These arrays allow for a high-throughput approach to quantitative RT-PCR by profiling up to 384 transcripts simultaneously. We customized a TLDA for the analysis of 48 transcripts, analyzed in technical duplicates, which serve as developmental markers of specific stages of endoderm and pancreas development (**Table 2.1**). Of the 48 transcripts analyzed, 18 have been characterized as part of the endoderm signature panel (ESP, noted in red script) (Sherwood et al., 2007) and 28 transcripts characterize various developmental stages as cells differentiate towards the endocrine lineage.

### **RNA isolation and quantitative reverse transcription PCR (qRT-PCR)**

Total RNA was isolated from cells following each day of differentiation using Stratagene's Absolutely RNA Miniprep Kit according to manufacturer's protocol. RNA quality was assessed by observing the relative intensity of the 18S and 28S ribosomal RNA bands by electrophoresis in 0.5X TBE agarose gels. Total RNA (500 ng) was reverse-transcribed into cDNA using the High Capacity cDNA Archive Kit (Applied Biosystems) according to manufacturer's protocol. Assuming a 1:1 conversion of RNA to cDNA, a 100 µl volume consisting of 100 ng cDNA combined with 1X TaqMan Universal Master Mix was transferred into a loading port on the TLDA. The array was prepped according to manufacturer's guidelines and analyzed on ABI's 7900HT Real-Time PCR System. Relative

**Table 2.1 Summary of transcripts profiled on human ES cell TLDA**

	GENE SYMBOL	GENE NAME	ASSAY ID
1	ANXA4	Annexin A4	Hs00154040_m1
2	BNIPL	BCL2/adenovirus E1B 19kDa interacting protein 1	Hs00414503_m1
3	CACNA1B	Calcium channel, voltage-dependent, N type, alpha 1B subunit	Hs00609480_m1
4	CDCP1	CUB domain containing protein 1	Hs00224587_m1
5	CDX1	Caudal type homeobox 1	Hs00156451_m1
6	CER1	Cerberus 1, cysteine knot superfamily, homolog (Xenopus laevis)	Hs00193796_m1
7	CLDN8	Claudin 8	Hs00273282_s1
8	CLIC6	Chloride intracellular channel 6	Hs00401339_m1
9	CXCR4	Chemokine (C-X-C motif) receptor 4	Hs00607978_s1
10	DSG2	Desmoglein 2	Hs00170071_m1
11	EMB	Emargin homolog (mouse)	Hs00419017_m1
12	EPCAM	Epithelial cell adhesion molecule	Hs00158980_m1
13	ESRP1	Epithelial splicing regulatory protein 1	Hs00214472_m1
14	FOXA1	Forkhead box A1	Hs00270129_m1
15	FOXA2	Forkhead box A2	Hs00232764_m1
16	FOXO1	Forkhead box O1	Hs01054576_m1
17	GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	Hs99999905_m1
18	GCG	Glucagon	Hs00174967_m1
19	GCK	Glucokinase (hexokinase 4)	Hs00175951_m1
20	GHRL	Ghrelin/obestatin prepropeptide	Hs00175082_m1
21	GSC	Goosecoid homeobox	Hs00418279_m1
22	GUSB	Glucuronidase, beta	Hs99999908_m1
23	HNF1B	HNF1 homeobox B	Hs00172123_m1
24	HNF4A	Hepatocyte nuclear factor 4, alpha	Hs00230853_m1
25	INS	Insulin	Hs00355773_m1
26	MAFB	V-maf musculoaponeurotic fibrosarcoma oncogene homolog B (avian)	Hs00534343_s1
27	MNX1	Motor neuron and pancreas homeobox 1	Hs00232128_m1
28	NEUROD1	Neurogenic differentiation 1	Hs00159598_m1
29	NEUROG3	Neurogenin 3	Hs00360700_g1
30	NKX2-2	NK2 homeobox 2	Hs00159616_m1
31	NKX6-1	NK6 homeobox 1	Hs00232355_m1
32	NPNT	Nephronectin	Hs00405900_m1
33	ONECUT1	One cut homeobox 1	Hs00413554_m1
34	PAX4	Paired box 4	Hs00173014_m1
35	PAX6	Paired box 6	Hs00240871_m1
36	PCSK1	Proprotein convertase subtilisin/kexin type 1	Hs00175619_m1
37	PDX1	Pancreatic and duodenal homeobox 1	Hs00236830_m1
38	POU5F1	POU class 5 homeobox 1	Hs01895061_u1

**Table 2.1 continued**

<b>39</b>	PTF1A	Pancreas specific transcription factor, 1a	Hs00603586_g1
<b>40</b>	RAB15	RAB15, member RAS oncogene family	Hs00419246_m1
<b>41</b>	RIPK4	Receptor-interacting serine-threonine kinase 4	Hs00221005_m1
<b>42</b>	SH3GL2	SH3-domain GRB2-like 2	Hs00182352_m1
<b>43</b>	SOX17	SRY (sex determining region Y)-box 17	Hs00751752_s1
<b>44</b>	SPINK1	Serine peptidase inhibitor, Kazal type 1	Hs00162154_m1
<b>45</b>	SST	Somatostatin	Hs00174949_m1
<b>46</b>	ST14	Suppression of tumorigenicity 14 (colon carcinoma)	Hs00222707_m1
<b>47</b>	T	T, brachyury homolog (mouse)	Hs00610080_m1
<b>48</b>	TMPRSS2	Transmembrane protease, serine 2	Hs00237175_m1

changes in mRNA expression were determined using the comparative Ct method ( $\Delta\Delta Ct$ ). Gene expression profiles were displayed as a heat map and scaled as a function of the number of standard deviations relative to the mean of each row, where red notes the maximum fold increase in expression and green represents a relative decrease in gene expression (**Figure 2.2**).

### **Generation of a *Pdx1*-driven GFP/Puro<sup>R</sup>-expressing reporter for RMCE**

To generate a *Pdx1*-driven reporter in human ES cells, we obtained a basal exchange vector, termed pX2.PuroTK, from the laboratory of Dr. Gordon Keller (Ontario, Canada) (Irion et al., 2007) (**Figure 2.3**). The pX2.PuroTK plasmid contains one loxP site and one lox2272 site flanking a *puromycin resistance/thymidine kinase* (*PuroTK*) fusion gene. Unique ClaI and HindIII sites are present at the loxP end, as well as a unique NcoI site at the lox2272 end. The sequences for the lox sites reported in the pX2.PuroTK plasmid are:

loxP sequence:	ATAACTTCGTATA ATGTATGC TATACGAAGTTAT
lox2272 sequence:	ATAACTTCGTATA <u>AG</u> GTATCC TATACGAAGTTAT

As reported in Lee et al. (Lee and Saito, 1998), the sequences for lox2272 and lox2372 sites are:

lox2272 sequence:	ATAACTTCGTATA <u>A</u> AGTATCC TATACGAAGTTAT
lox2372 sequence:	ATAACTTCGTATA <u>AG</u> GTATCC TATACGAAGTTAT

As evident by the nucleotides emphasized by underline, the lox2272 site reported in Irion et al. is a lox2372 site according to the original publication by Lee et al. Therefore, attention needs to be taken concerning future manipulations with this plasmid given that the designations do not match the standard nomenclature detailed in the original publication.

A *Pdx1*-containing BAC (clone 2270K21) isolated from the CTD BAC library was used for the construction of the pX2<sup>Pdx1-GFP/PuroR</sup> exchange cassette. The BAC DNA was electroporated (1800 V, 25  $\mu$ F, 200  $\Omega$ ) into electrocompetent EL350 cells which contain the



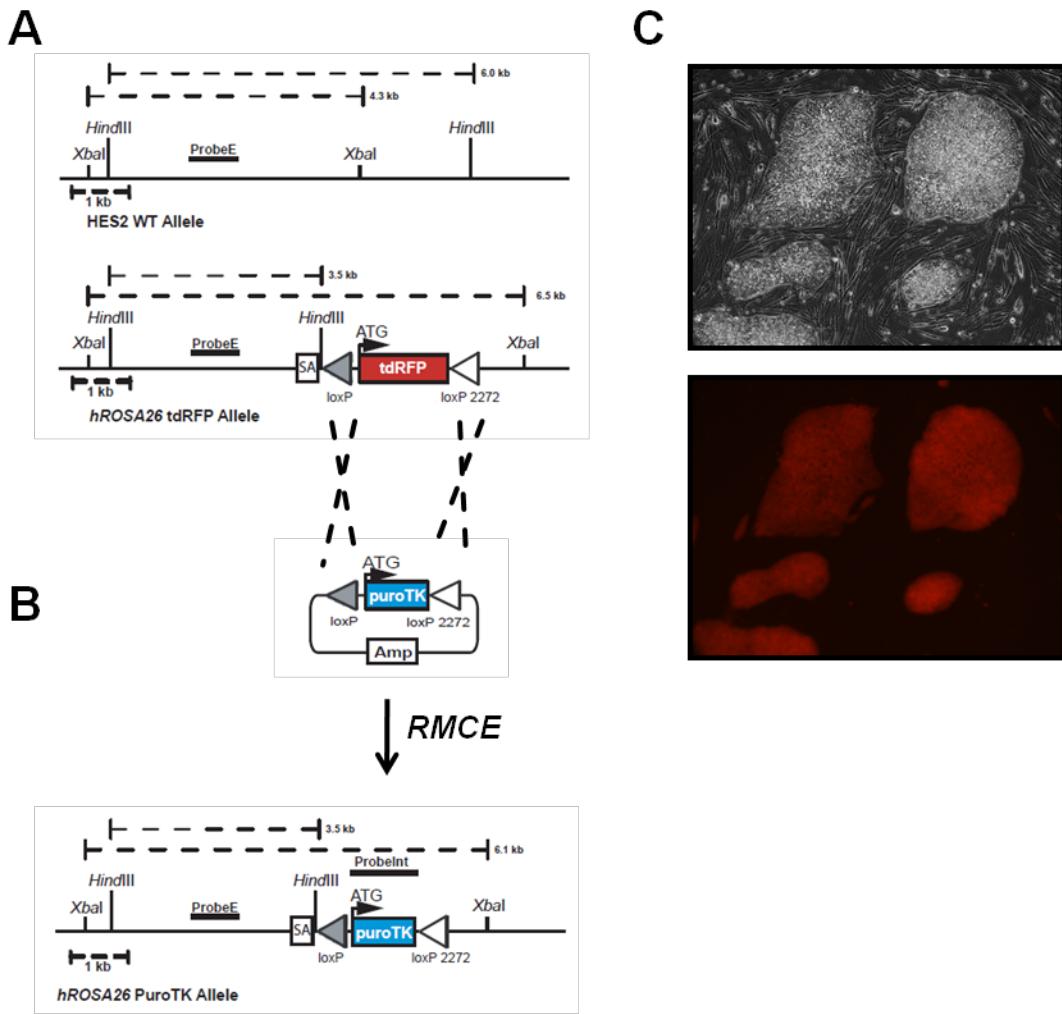
**Figure 2.2 Directed differentiation of human ES cells towards pancreatic fates.** WA-09 human ES cells were differentiated following a five-stage differentiation protocol that induced the cells through various intermediates towards the pancreatic endocrine lineage. Using a customized TaqMan Low Density Array (TLDA), the gene expression profiles of the cells were assayed following each day of differentiation for 15 days. The qRT-PCR results for each gene are displayed as a heat map to visualize the results following data analysis. Red notes the maximum fold increase in gene expression and green represents a relative decrease in gene expression. The intensity of each row is normalized separately.

recombination proteins *exo*, *bet*, and *gam* controlled by the temperature-sensitive repressor *cI857* (Lee et al., 2001). Following electroporation, chloroamphenicol-resistant clones were identified by restriction analysis with EcoRI and EcoRV before proceeding to the first recombineering reaction. Four homologous regions (HR1, HR2, HR3, and HR4) approximately 500 base pairs in length were PCR amplified from the BAC DNA using the following primers sets:

HR1 upper primer:	5' CTAGCCCATGGCATGCAAAGAAATAAGTGTG,
HR1 lower primer:	5' CTAGCCCATGGGGGAGGCTGAGGTGGGTGGAT;
HR2 upper primer:	5' CTAGCATCGATACAGGATAGGAGTAAAGAGGAA,
HR2 lower primer:	5' CTAGCATCGATAGATGGCGCTGAGGATT;
HR3 upper primer:	5' CTAGCATCGATCTCAACATGTCTCCTGTAAACT,
HR3 lower primer:	5' CTAGCGGCCGGCCTGCGGCCGGATTGGGCACCGGGA;
HR4 upper primer:	5' CTAGCATGCATCCGCGGTGGCGCACCTTCACCA,
HR4 lower primer:	5' CTAGCGTTAAACTCGGCTTCCTCGATGTGCACTA.

The primers for amplification introduced appropriate restriction enzyme sites on the 5' and 3'-ends of the homologous regions to facilitate cloning into the appropriate insertion and retrieval vectors.

A DNA fragment containing an enhanced GFP and puromycin resistance reporter was cloned into the PL451 plasmid which contained a *pgk*/EM7 promoter-driven *neomycin resistance* (*Neo*<sup>R</sup>) flanked by tandem FRT sites. Homologous regions 3 and 4 (HR3 and HR4) were cloned into this plasmid, creating the insertion plasmid. EL350 cells containing the BAC were electroporated (1800 V, 25 µF, 200 Ω) with insertion plasmid which was linearized with EcoRV and Pme1, and kanamycin-resistant clones were screened for positive insertions. The second recombineering step entailed the retrieval of DNA from the modified BAC into a basal retrieval vector. Homologous regions 1 and 2 (HR1 and HR2) were cloned into the pX2.PuroTK plasmid utilizing the unique NcoI and ClaI sites, thereby generating the retrieval plasmid. EL350 cells containing the modified BAC were electroporated (1800 V, 25 µF, 200 Ω) with retrieval plasmid which was linearized with HindIII and NdeI, and



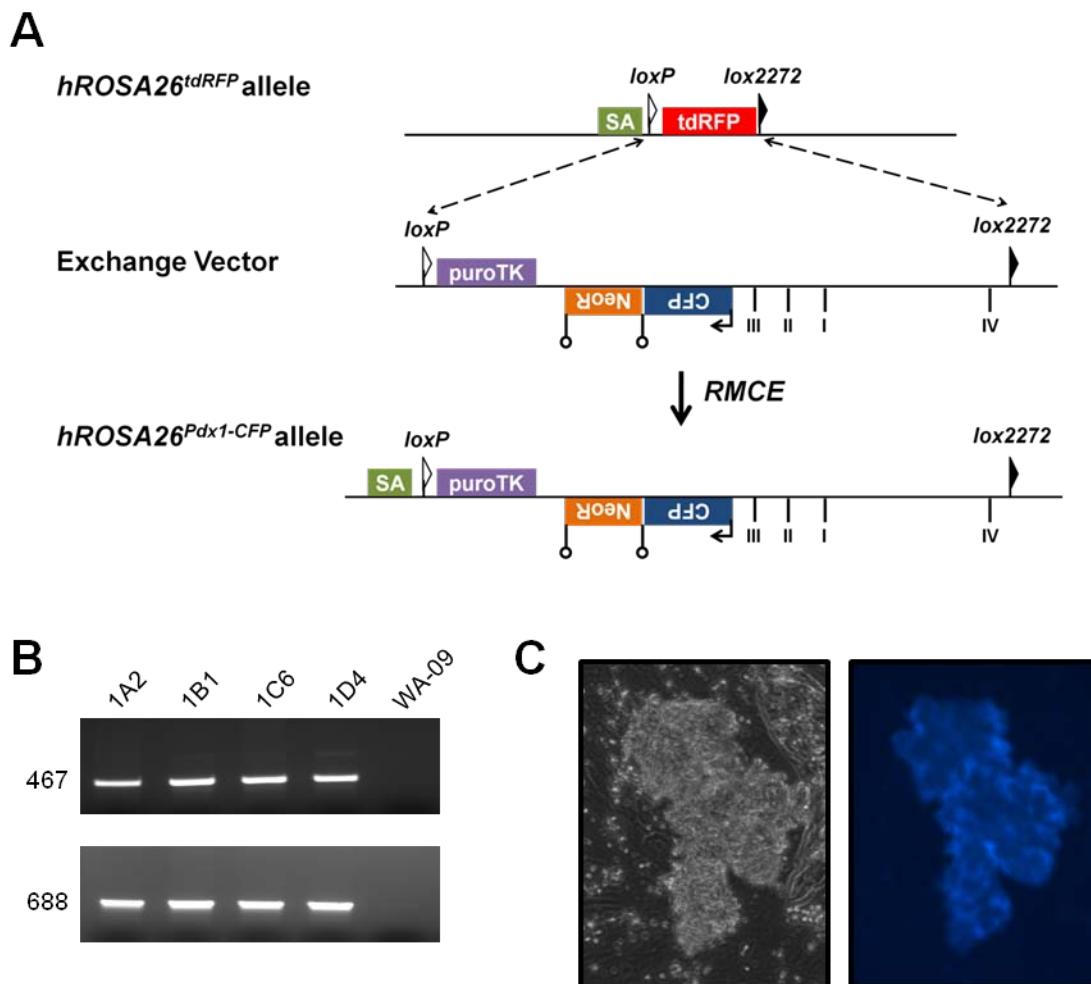
**Figure 2.3 Gene targeting of the *ROSA26* locus in human ES cells.** A) Schematic of human *ROSA26* locus prior to and following targeting. The wild type *hROSA26* allele is depicted in the upper panel. Following homologous recombination, the *hROSA26* allele contains a viral splice acceptor (SA) followed by a loxP and lox2272 flanked *tdRFP* gene. The dashed lines indicate either predicted WT (6.0 kb) and targeted (3.5 kb) Southern band sizes following HindIII digestion or predicted WT (4.3 kb) and targeted (6.5 kb) Southern band sizes following XbaI digestion. B) Recombinase-mediated cassette exchange in the *hROSA26<sup>tdRFP</sup>* allele. The loxP and lox2272 sites allow for RMCE of an exchange cassette into the *hROSA26<sup>tdRFP</sup>* allele. A cassette containing a promoterless fusion of *puromycin resistance* and *thymidine kinase* (puroTK) was exchanged into the *hROSA26<sup>tdRFP</sup>* allele. The splice acceptor site allows for *puromycin resistance* to be driven by the *hROSA26* promoter. Gene targeting and RMCE figures adapted from Irion et al. (2007). C) Brightfield (upper panel) and fluorescence (lower panel) images of *hROSA26<sup>tdRFP</sup>* colonies.

ampicillin-resistant clones were screened for positive retrievals. The placement of homologous regions 1 and 2 resulted in the removal of the *PuroTK* sequence following recombination.

The pX2<sup>Pdx1-GFP/PuroR</sup> plasmid contains approximately 9.4 kb of the *Pdx1* regulatory elements (Areas I-IV) upstream of the transcriptional start site. Following the *Pdx1* 5' UTR are sequences containing a FLAG-tag, enhanced GFP, a 3X SV40 nuclear localization signal, an internal ribosome entry site (IRES2), puromycin resistance and a rabbit β-globin polyadenylation signal. These sequences are followed by a *pgk*/EM7 promoter-driven neomycin resistance sequence flanked by FRT sites. The *pgk* mammalian promoter allows for positive selection of correctly exchanged clones following RMCE, while the EM7 bacterial promoter allows for positive selection during BAC recombineering. Following the reporter and antibiotic selection sequences are approximately 1.2 kb of *Pdx1* exon 1 and intron sequences.

### **Generation of a *Pdx1*-driven CFP-expressing reporter for RMCE**

To generate the CFP-expressing insertion plasmid, linker-primer PCR was used to amplify the nuclear-localized (3X SV40 NLS) Cerluean sequence, as well as the rabbit β-globin polyadenylation sequence, which contains the splicing region and polyA sequences, from the Pdx1<sup>CFP</sup> exchange cassette (discussed in Chapter V) (**Figure 2.4**). The construct was cloned into the aforementioned insertion plasmid, which contained the *pgk*/EM7 promoter-driven neomycin resistance flanked by tandem FRT sites as well as homologous regions 3 and 4 (HR3 and HR4), following excision of the GFP/Puro<sup>R</sup> cassette. EL350 cells containing the BAC were electroporated (1800 V, 25 μF, 200 Ω) with SgrA1- and EcoRV-linearized insertion plasmid. Kanamycin-resistant clones were screened for positive insertions.



**Figure 2.4 Recombinase-mediated cassette exchange (RMCE) in human ES cells.** A) The derivation of a human ES cell line (hES2) expressing a *hROSA26<sup>tdRFP</sup>* loxed cassette acceptor allele has been previously described by Irion et al. (2007). The *Pdx1*-CFP exchange vector contains approximately 9.4 kb of the *Pdx1* regulatory elements containing Areas I-IV and the 5'UTR. Cerulean and an FRT-flanked neomycin resistance ( $\text{Neo}^R$ ) gene are followed by approximately 1.2 kb of *Pdx1* exon 1 and intron sequence. Following RMCE, the final allele allowed for positive selection by the puroTK fusion protein. B) Representative PCR screen revealing exchange events on both the 5' and 3' ends. The four clones shown were among the 24 correctly exchanged clones. H9 DNA represents control DNA. C) Brightfield and fluorescent image of hES cell-derived colony expressing *hROSA26<sup>Pdx1-CFP</sup>* allele following directed pancreatic differentiation. Image courtesy of Eunyoung Choi.

The second recombineering step entailed the retrieval of DNA from the modified BAC into the retrieval vector (pX2.PuroTK). Five unique restriction sites were cloned into the NcoI site of the basal retrieval vector, and homologous regions 1 and 2 (HR1 and HR2) were cloned into the modified pX2.PuroTK plasmid utilizing these unique restriction sites. EL350 cells containing the modified BAC were electroporated (1800 V, 25 µF, 200 Ω) with linearized retrieval plasmid, and ampicillin-resistant clones were screened for positive retrievals. The placement of homologous regions 1 and 2 resulted in the retention of the *PuroTK* sequences following recombination.

The pX2.PuroTK<sup>Pdx1-CFP</sup> plasmid contains a *puromycin resistance/thymidine kinase* (*PuroTK*) fusion gene to allow for positive selection following RMCE. Additionally, it contains approximately 9.4 kb of the *Pdx1* regulatory elements (Areas I-IV) upstream of the transcriptional start site. Following the *Pdx1* BAC 5' UTR is a nuclear-localized Cerulean along with the rabbit β-globin polyadenylation sequence. These sequences are followed by a *pgk/EM7*-promoter driven *neomycin resistance* sequence flanked by FRT sites allowing positive selection following both RMCE and BAC recombineering. Following the reporter and antibiotic selection sequences are approximately 1.2 kb of *Pdx1* exon 1 and intron sequence. The entire cassette is flanked by one loxP site and one lox2272/2372 site (**Figure 2.4**).

### **Generation of targeting vectors for *Pdx1* gene targeting**

Utilizing both the GFP/Puro<sup>R</sup>-expressing and CFP-expressing constructs, two variations of a *Pdx1* gene targeting vector were designed. Homologous regions 1 and 2 were PCR amplified from the BAC (clone 2270K21) using the following primers sets which contain leader sequences, necessary restriction enzyme sites and primer sequences:

HR1 upper primer: 5' CTAGCACTAGTGTAAACCTTAAGTACCCTGTGGAG  
CAGGTGATTA;  
HR1 lower primer: 5' CTAGCAAGCTTATCCTCTGCTTCCGCATCCTCGGA;  
HR2 upper primer: 5' CTAGCGTCGACGACACAGGCCGCCGGAACTT,  
HR2 lower primer: 5' CTAGCATCGATCCCTCACTCCCCGCGCTCGTTA.

The amplified products were cloned into the basal retrieval vector (pBS.DTA) which contains a MC1-driven *diphtheria toxin A* gene (*DTA*). To generate the GFP/Puro<sup>R</sup>-expressing and CFP-expressing gene targeting vectors, the aforementioned EL350 cells containing the modified BAC in which either the GFP/Puro<sup>R</sup>-expressing cassette or the CFP-expressing cassette has been inserted were electroporated (1800 V, 25 µF, 200 Ω) with the ClaI-linearized retrieval plasmid, and ampicillin-resistant clones were screened for positive retrievals.

### Recombinase-mediated cassette exchange in human ES cells

For DNA electroporation, hES2 cells (passage 24) that contained the tdRFP-expressing *ROSA26<sup>LCA</sup>* allele (**Figure 2.3**) were collected by trypsinization and triturated to single cells. 5 x 10<sup>6</sup> cells were electroporated with 40 µg of pX2.PuroTK<sup>Pdx1-CFP</sup> and 40 µg pBS185, a Cre expression plasmid, using a BioRad gene pulser (250 V, 500 µF). Cells were cultured on DR4 MEF cells, and selection with 0.5 µg/ml of puromycin began 72 hours post-electroporation and continued for eleven days. Seventy-two colonies survived selection and were identified as tdRFP-negative, and 24 clones were analyzed by PCR using primers that spanned either the lox66/71 site on the 5' end or the lox2272 site on the 3' end. On the 5' end, the combination of 5' TTTGTGGGTGGGAGGCGCTT and 5' CGTGGCTTGTACTCGGTCA primers resulted in a band size of 467 bp after RMCE and on the 3' end, use of 5' TGCCTGCGCCTGTAATCCTG and 5' CCCCAGGTGAATGACTAAGCTCCA resulted in a 688 bp following RMCE. The resulting PCR products were visualized by electrophoresis in a 1% agarose gel and 24 out of

24 clones were confirmed as correctly exchanged. Clones 1A2, 1B1, 1C6 and 1D4 were expanded and cryopreserved for future experiments (**Figure 2.4**).

### Gene targeting in human ES cells

Two gene targeting experiments were performed using the GFP/Puro<sup>R</sup>-expressing targeting vector. For the first targeting experiment, WA-09 human ES cells (passage 36) were briefly cultured with media containing 10 mM of a ROCK inhibitor (Y-27632, Calbiochem) then collected by trypsinization.  $5 \times 10^6$  cells were electroporated (250 V, 500  $\mu$ F) with 40  $\mu$ g of pBS.DTA<sup>Pdx1-GFP/PuroR</sup> and cultured on DR4 MEFs in ROCK inhibitor-containing hES cell medium. Selection with neomycin sulfate (G418) began 48 – 72 hours post-electroporation and was performed in two stages: 25  $\mu$ g/ml of G418 for four days followed by 50  $\mu$ g/ml for seven days. Nine surviving clones were screened, and none displayed positive targeting. For the second targeting experiment, WA-09 cells (passage 32) were briefly cultured with medium containing 10 mM of a ROCK inhibitor then collected by trypsinization. Four electroporations were performed in which 40  $\mu$ g of pBS.DTA<sup>Pdx1-GFP/PuroR</sup> was electroporated into  $5 \times 10^6$  WA-09 cells and cultured on DR4 MEFs in ROCK inhibitor-containing hES cell medium. Selection with neomycin sulfate (G418) began 48 – 72 hours post-electroporation and was performed in two stages. 130 clones survived selection, and 86 clones were screened. None displayed positive targeting.

Due to poor expression of the GFP/Puro<sup>R</sup> construct in mice (discussed in Chapter V), the human *Pdx1* gene targeting vector was redesigned to incorporate a Cerulean-expressing construct. For electroporation, hES2 cells (passage 32) were briefly cultured with media containing 10 mM of a ROCK inhibitor then collected by trypsinization.  $5 \times 10^6$  cells were electroporated with 40  $\mu$ g of pBS.DTA<sup>Pdx1-CFP</sup> and cultured on DR4 MEFs in ROCK inhibitor-containing hES cell media. Selection with neomycin sulfate (G418) began 48 – 72

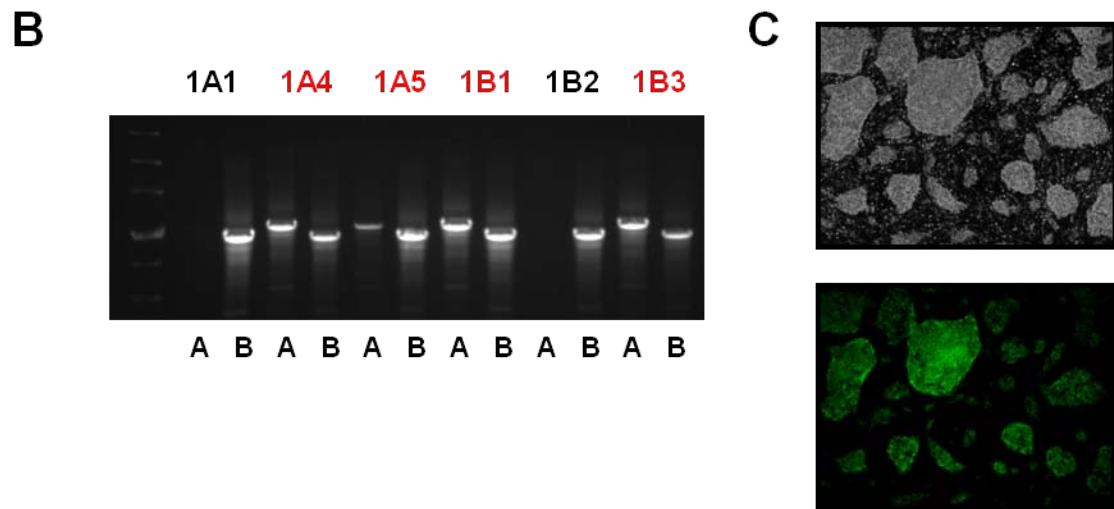
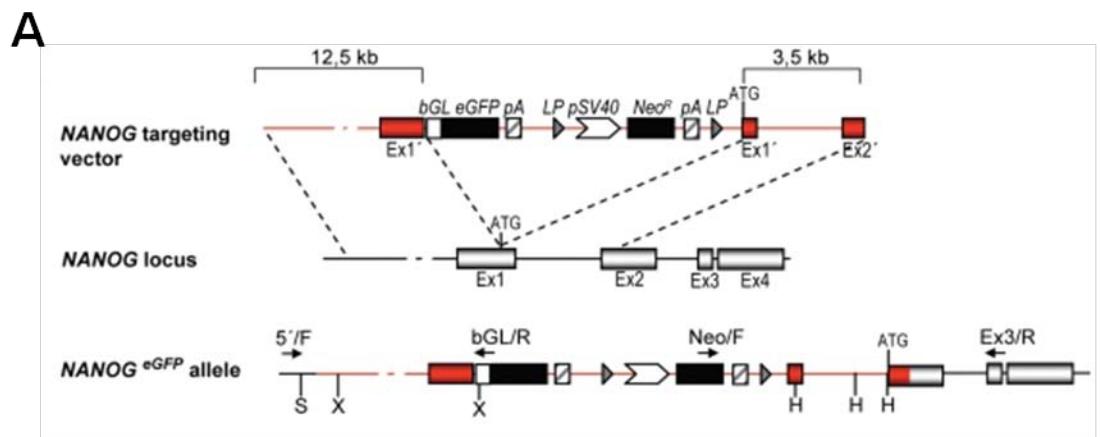
hours post-electroporation and was performed in two stages. 94 clones survived selection, and 90 clones were screened. None displayed positive targeting.

Due to the inefficiency experienced with gene targeting for *Pdx1* and to validate the electroporation parameters, we obtained a targeting vector for *Nanog*, a homeodomain transcription factor that is critical for maintaining stem cell pluripotency (Fischer et al., 2010). The targeting vector consists of eGFP and neomycin resistance flanked by 12.5 kb and 3.5 kb homology arms, and homologous recombination has been shown successful in huES-1 and huES-3 cells (Fischer et al., 2010) (**Figure 2.5**). For electroporation, both WA-09 (passage 31) and hES2 (passage 38) cells were briefly cultured with medium containing 10 mM of a ROCK inhibitor then collected by trypsinization.  $5 \times 10^6$  cells were electroporated with 40 µg of the targeting vector and cultured on DR4 MEFs in ROCK inhibitor-containing hES cell medium. Selection with neomycin sulfate (G418) began 48 – 72 hours post-electroporation and was performed in two stages: 25 µg/ml G418 for four days followed by 50 µg/ml for seven days. From the hES2 electroporation, 81 clones survived selection and 18 GFP-positive clones were screened. Seven displayed positive recombination by PCR screening (clones 1A4, 1A5, 1B1, 1B3, 1A2, 1A3 and 1A6) and displayed GFP fluorescence (**Figure 2.5**). From the WA-09 electroporation, one clone survived selection and was identified as targeted by PCR (clone 2A6).

## Results

### Transition of human ES cells through definitive endoderm intermediate

Signaling through the TGF $\beta$  pathway is critical for the formation of the primitive streak, mesoderm and definitive endoderm in vertebrates (Brennan et al., 2001; Conlon et al., 1994; Lowe et al., 2001; Vincent et al., 2003). One of the members of the TGF $\beta$  family,



**Figure 2.5 Gene targeting of *Nanog* in human ES cells.** A) The targeting vector for *Nanog*, consisting of eGFP and a neomycin resistance gene flanked by loxP sites was inserted into the BAC at the *Nanog* start codon. Long (12.5 kb) and short (3.5 kb) homology arms facilitated homologous recombination resulting in the final allele which contained eGFP following the 5' UTR along with a SV40-driven Neo<sup>R</sup> flanked by loxP sites. Abbreviations: bGL, rabbit *beta-globin* intron 2; eGFP, enhanced green fluorescent protein gene; LP, loxP; Neo<sup>R</sup>, neomycin resistance gene; pA, polyadenylation site; pSV40, SV40 promoter; Ex, exon; Ex1', Ex2', truncated exon 1, 2. Figure from adapted from Fischer et al. (2010). B) Representative PCR screen revealing targeted (lane A, 5.1 kb) and wild type (lane B, 4.6 kb) alleles. The four clones noted in red were among the eight correctly targeted clones. C) Brightfield (upper panel) and fluorescence (lower panel) images of *Nanog*<sup>eGFP</sup> colonies (clone 1B3).

Nodal, is essential for the induction of the definitive endoderm. Thus, given the critical role of TGF $\beta$  signaling in the specification of the definitive endoderm, the first stage of the directed pancreatic differentiation protocol stimulates hES cells with high levels of Activin A, a readily available stimulator of the TGF $\beta$  pathway that mimics Nodal activity *in vitro* (**Figure 2.1**). In addition to stimulating the cells with Activin A, Wnt3a improves the transition through the mesendoderm state. During the initial differentiation of the hES cells to definitive endoderm, it is critical to decrease serum supplementation. The presence of IGF or insulin in serum leads to elevated PI3K signaling which inhibits the differentiation of hES cells to DE (McLean et al., 2007).

The hES cell-derived definitive endoderm expresses transcripts indicative of its definitive endoderm like state, such as *Sox17*, *Gsc* and *Cxcr4*, as well as the primitive streak marker *Mixl1* (Kanai-Azuma et al., 2002; McGrath et al., 1999; Yasunaga et al., 2005) (**Figure 2.1 and 2.2**). Furthermore, the anterior nature of the definitive endoderm was evident by the expression of *Cer1* and *Foxa2* (**Figure 2.1 and 2.2**) (Biben et al., 1998; Sasaki and Hogan, 1993). In addition to these endoderm markers, the TLDA design incorporated 19 of the 22 endoderm signature panel genes which have been shown to be preferentially expressed in E8.25 mouse definitive and visceral endoderm (Sherwood et al., 2007). These included *Emb*, *Cacna1b*, *Cdcp1*, *Cldn8*, *Clic6*, *Npnt*, *Rab15*, *Ripk4*, *Spink1*, *Tmprss2*, *Foxa1*, *Tacstd1*, *Anxa4*, *Dsg2*, *Rbm35a*, *Sh3gl2*, *St14* and *Bnip1* (**Table 2.1 and Figure 2.2**). These genes were incorporated to further analyze the transition of the hES cells through the mesendoderm to the definitive endoderm.

## **Transition of human ES cells through the primitive gut tube and posterior foregut intermediates**

During early pancreas development, it is well known that inhibition of hedgehog signaling is critical for the specification of gut tube endoderm (Kim and Melton, 1998; Lau et al., 2006). Additionally, signaling through members of the FGF family permit pancreatic endoderm proliferation and play a role in endoderm patterning (Bhushan et al., 2001; Hart et al., 2003; Jacquemin et al., 2006; Ye et al., 2005). The addition of FGF10, or FGF7, and cyclopamine, an inhibitor of Hedgehog signaling, induced formation of cells resembling the primitive gut tube endoderm and posterior foregut (**Figure 2.1**). The transition into a primitive gut tube-like state was evident by the expression of both *Tcf2* (*Hnf1b*) and *Hnf4a* during the second stage of differentiation (**Figure 2.2**). During embryogenesis, the transcription factor *Tcf2* (*Hnf1b*) is expressed throughout the primitive gut tube of the mouse at E8.0; subsequently, within 12 hours *Hnf4a* is similarly expressed (Barbacci et al., 1999; Coffinier et al., 1999; Duncan et al., 1994). This similar temporal expression pattern was seen during the directed differentiation of hES cells, as *Hnf1b* expression was evident by the fifth day of differentiation and *Hnf4a* expression peaked approximately one day later. In the next stage of differentiation, the gut tube endoderm-like cells are stimulated with retinoic acid, along with FGFs and cyclopamine, which plays a major role in formation of the pancreatic anlage (Stafford et al., 2004) (**Figure 2.1**). The exposure to retinoic acid induces the cells to transition into a state resembling the posterior foregut endoderm; a number of pancreatic markers such as *Ipfl* (*Pdx1*), *Mafb*, *Onecut1* (*Hnf6*), and *Hlx9* are expressed (**Figure 2.2**).

## **Transition of human ES cells through an endocrine precursor intermediate and into hormone-positive cells**

Following specification of the pancreatic endoderm from the posterior foregut endoderm, a subset of pancreatic endoderm cells are specified to the endocrine lineage. The inhibition of Notch signaling is a critical regulator of this stage and promotes the expression of the pre-endocrine marker *Neurog3* (Wilson et al., 2003). Therefore, in the fourth stage of differentiation the  $\gamma$ -secretase inhibitor DAPT is included to inhibit signaling via the Notch pathway (**Figure 2.1**). Additionally, exendin-4, a GLP-1 agonist well known in its ability to stimulate insulin secretion and promote  $\beta$  cell replication and neogenesis, was added to promote the specification of the posterior foregut endoderm cells to the endocrine lineage (Brubaker and Drucker, 2004). The transition of the hES cells into cell resembling pancreatic endoderm was evident by the expression of not only *Neurog3*, but also *Nkx2-2* (**Figure 2.2**).

In the final stage of differentiation, the endocrine precursor cells are differentiated towards hormone-expressing fates. In addition to exendin-4, the cells are induced with IGF1 and HGF to promote specification to hormone-expressing fates (**Figure 2.1**). At approximately 15 days of differentiation, the expression of various hormones, such as *Insulin* (*Ins*), *Glucagon* (*Gcg*), *Ghrelin* (*Ghrl*) and *Somatostatin* (*Sst*), became evident (**Figure 2.2**).

When comparing *Insulin* expression from human islets to the *Insulin* levels seen in human ES cell-derived cells, we discovered that the efficiency of directed differentiation hES cells to *Insulin*-expressing cells is quite low. When compared to undifferentiated human ES cells, isolated human islets exhibited a  $1.6 \times 10^7$  fold increase in *Insulin* expression and a  $1.3 \times 10^9$  increase in *Glucagon* expression. Conversely, the human ES cell-derived endocrine cells display only a 308-fold and 55,392-fold increase in *Insulin* and *Glucagon*, respectively. These numbers represent only 0.0019% and 0.0042% of the *Insulin* and *Glucagon* expression levels, respectively, that are seen in endogenous human islets. The meager expression levels

induced in hES cell-derived cells further points to the need for more efficient differentiation protocols.

### **Generation of a reporter allele for *Pdx1* in human ES cells**

In order to effectively monitor the directed differentiation of hES cells to pancreatic fates, it would be advantageous to have fluorescently-tagged reporter alleles that would identify specific cell populations as cells transition through various developmental intermediates. To achieve this, we utilized hES cells that contained a loxed cassette acceptor (LCA) allele in the *ROSA26* locus (Irion et al., 2007) (**Figure 2.3**). This allele contains a red fluorescent protein, tdRFP, flanked by loxP and lox2272/2372 sites which facilitate the exchange of DNA cassettes into the allele by Cre-recombinase. We generated an exchange vector for *Pdx1* that used a 9.4 kb fragment of the upstream regulatory sequences of *Pdx1*, containing Areas I – IV and the 5' UTR, to drive the expression of a nuclear-localized cyan fluorescent protein, Cerulean (**Figure 2.4**). A portion of the rabbit  $\beta$ -globin gene, containing both intronic and polyadenylation sequences, was placed downstream of the CFP coding sequences. Additionally, an FRT-flanked *neomycin resistance* (*Neo*<sup>R</sup>) gene was incorporated for positive selection during BAC recombineering and was followed by approximately 1.2 kb of *Pdx1* sequences for exon 1 and intron. The exchange cassette was designed such that when exchanged into the *ROSA26*<sup>tdRFP</sup> allele the expression of the reporter would be transcribed in the opposite direction, thus minimizing the potential for the *Pdx1*<sup>CFP</sup> transgene to be transcribed from the *ROSA26* promoter. RMCE into the *ROSA26*<sup>tdRFP</sup> allele was achieved by the co-electroporation of the exchange vector and a Cre-expression plasmid. Properly exchanged clones were identified through the use of a positive-negative selection strategy where the puroTK selection cassette permitted identification of positively exchanged clones, and clones not properly exchanged were excluded based on tdRFP fluorescence.

The *Pdx1*<sup>CFP</sup> transgene permitted identification of hES cells that had differentiated towards posterior foregut and pancreatic endoderm-like fates. Following the directed differentiation *ROSA26*<sup>*Pdx1*-CFP</sup> hES cells according to a method similar to that described above, cyan fluorescence could be observed in the hES cell-derived pancreatic cells (**Figure 2.4**). However, *Pdx1*<sup>CFP</sup>-expressing cells could not be isolated via fluorescence-activated cell sorting (FACS) due to limitations of excitation wavelengths available (discussed in Chapter V).

### Gene targeting in human ES cells

While the *ROSA26*<sup>*Pdx1*-CFP</sup> allele permits the identification of hES cell-derived *Pdx1*-expressing pancreatic progenitors, the derivation of targeted hES cell lines would be more advantageous than the generation of transgene reporters. The generation of a knockin allele is more advantageous over alternatives for generating fluorescent reporter lines, mainly because it more closely mimics endogenous gene expression without the confounding affects of random gene disruption or random integration. However, gene targeting has proven extremely challenging in hES cells and less robust than in mES cells. Since the initial report in 2003, a small number of publications have demonstrated successful gene targeting in hES cells (Costa et al., 2007; Davis et al., 2008; Fischer et al., 2010; Irion et al., 2007; Urbach et al., 2004; Zwaka and Thomson, 2003). Many limitations and obstacles prevent this method from fully recognizing its potential; however, improved methods and culture conditions are being developed and implemented. For instance, to increase the survival of dissociated hES cells, the addition of an inhibitor of p160-Rho-associated coiled-coil kinase (ROCK), Y-27632, to the culture medium and has proven highly effective in reducing apoptosis and promoting hES cell colony survival (Watanabe et al., 2007).

To generate fluorescent reporter alleles in hES cells that would report the formation of pancreatic progenitors, we have generated a targeting vector for *Pdx1*. However, none of the targeting experiments yielded clones that underwent the desired homologous recombination event which may reflect chromosomal position effects (Yanez and Porter, 2002). Due to the inefficiency experienced with gene targeting for *Pdx1* and to validate the electroporation parameters, we obtained a targeting vector for *Nanog*, a homeodomain transcription factor that is critical for maintaining stem cell pluripotency. The *Nanog*<sup>eGFP</sup> gene targeting vector contains both an *eGFP* reporter and selection cassette that, when correctly inserted into the *Nanog* locus by homologous recombination, will lie immediately 5' of the *Nanog* start codon (Fischer et al., 2010) (**Figure 2.5**). To test the hES cell targeting protocol, we electroporated this vector into WA-09 and hES2 human ES cells. The clones undergoing the desired homologous recombination event exhibited green fluorescence and were identified by PCR screening (**Figure 2.5**). We observed a targeting efficiency of 8.6%, which is slightly less than the reported 11.4% efficiency observed when targeting in hUES3 cells but significantly better than the 0.6% efficiency observed when targeting in hUES1 cells (Fischer et al., 2010). While significant differences between mouse and human ES cells have limited the application of gene targeting in human ES cells, the targeting of *Nanog* demonstrated that the conditions of the established protocol were efficient for homologous recombination.

## Discussion

While progress has been made in understanding how to generate specific endocrine cell precursors, such as the definitive and foregut endoderm and pancreatic endocrine progenitors, the ability to differentiate hES cells to glucose-responsive insulin-producing cell fates, solely through *in vitro* differentiation, has yet to be achieved. Although *insulin-*

expressing cells have been reported, the cells primarily resemble immature endocrine cells and respond poorly to glucose stimulation (D'Amour et al., 2006). While subsequent studies have shown that glucose-responsive insulin-producing cells can be generated following *in vivo* maturation of pancreatic progenitors, the cells contribute to teratoma formation, thus pointing out the critical need for either purification of transplantable progenitors or more efficient *in vitro* differentiation protocols (Kroon et al., 2008). To more efficiently direct the differentiation of hES cells towards  $\beta$ -like cell fates, a better understanding of the signaling mechanisms and the gene regulatory networks induced during *in vivo* pancreas development is needed so that it may be translated to hES cell directed pancreatic differentiation methods.

By applying fluorescence-activated cell sorting (FACS) and whole transcriptome sequencing, transcriptional profiles for hES cell-derived cell populations can be obtained and used to examine specific developmental stages of hES cell differentiation. These transcriptional profiles can also be used in combination with profiles obtained from *in vivo* isolated cell populations from the mouse (discussed in Chapter III) to determine how closely hES cell-derived intermediates mimic the transcriptional profiles of authentic cell populations. However, with comparisons such as this there are a number of concerns. First, it is unknown how closely the transcriptional profiles of mouse and human cells resemble one another. If the transcriptional profiles from *in vivo* isolated mouse and human cells are markedly different, then the comparison of *in vivo* isolated cells from the mouse and *in vitro* hES cell-derived cell populations will inevitably suffer from similar differences. Second, during *in vitro* differentiation, the expression profile for a given transcript can follow numerous patterns (i.e. expression can gradually increase followed by gradual decrease, expression can decrease over time, expression can increase over time and plateau, etc). Given this, the optimal time to isolate hES cell-derived intermediates and compare their transcriptional profile to *in vivo* isolated intermediates in the mouse is uncertain. The

developmental timing of hES cell differentiation protocols is still largely unknown, and most investigators simply transition to the next “stage” of differentiation following peak expression of a specific transcript, or in some cases multiple transcripts, that marks a key developmental stage (i.e. the peak expression of *Pdx1* would indicate the end of differentiation to posterior foregut-like fate and the point to transition into new factors that promote endocrine progenitor differentiation). Thus, it is unknown if the optimal time to isolate specific cell populations is at the peak of a transcript’s expression or at another point during its expression pattern. However, regardless of the unknowns, the transcriptional comparison of hES cell-derived intermediates to *bona fide* cell populations will be critical for assessing the authenticity of the cellular intermediates that are generated during the directed differentiation of hES cells towards pancreatic fates.

While the study by Kroon et al. revealed the competence of hES cell-derived pancreatic endoderm to generate functional glucose-responsive insulin-producing cells following *in vivo* maturation, the study highlights a critical need for the purification of transplanted cells (Kroon et al., 2008). Although the transplanted cells primarily consisted of cells that resemble the pancreatic endoderm, the inefficiency of differentiation resulted in a heterogeneous population that gave rise to teratomas when transplanted into recipient mice. It is thought that purification of the hES cell-derived pancreatic endoderm-like cells will eliminate subsequent teratoma formation observed. Based on the gene expression profiles reported by Kroon et al., at the end of the fourth stage of differentiation, the point when the cells are isolated for transplantation, *Pdx1* reaches its peak expression while *Ptf1a* and *Nkx6-1* are at the onset of expression (Kroon et al., 2008). Therefore, if *Pdx1* is a key marker to identify the readiness of cells for transplant, a fluorescently-tagged *Pdx1* allele would be useful for purifying the pancreatic endoderm-like cells following directed differentiation and addressing two critical issues. First, following transplantation of purified *Pdx1*-positive cells,

we would be able to assess whether or not the purification of transplantable cells eliminates the formation of teratomas in recipient mice. Additionally, we would be able to determine if Pdx1-positive cells, which represent the pancreatic endoderm-like cells, are solely responsible for the functional insulin-producing cells formed following *in vivo* maturation in recipient mice or if the formation is a result of other cells present in the heterogeneous implanted population, such as the few endocrine-positive cells observed in the transplantable cells.

## CHAPTER III

### TRANSCRIPTIONAL PROFILING OF *PTF1A<sup>YFP</sup>*-EXPRESSING PANCREATIC CELLS

#### Introduction

The pancreas arises from dorsal and ventral evaginations of the foregut endoderm that first become evident around embryonic day (E) 9.5. These early multipotent progenitor cells (MPCs) can be distinguished from the surrounding mesenchyme and other foregut descendants by the expression of *pancreas specific transcription factor 1a* (*Ptf1a*, p48) (Burlison et al., 2008; Kawaguchi et al., 2002; Krapp et al., 1998; Roux et al., 1989; Zhou et al., 2007), a key component of both the early and late forms of the heterotrimeric PTF1 complex. In addition to serving as a valuable lineage-specific marker, *Ptf1a* also plays a key role in pancreas development. Mice lacking *Ptf1a* fail to develop a mature pancreas due to the limited expansion of the pancreatic epithelium, thereby suggesting that *Ptf1a* plays an essential role in the proliferation and subsequent lineage specification of pancreatic MPCs to become endocrine, ductal, and acinar cells characterizing the mature pancreas (Kawaguchi et al., 2002; Krapp et al., 1998).

Little is known about how the genetic network within pancreatic MPCs is established, how it changes during normal endocrine, acinar and ductal lineage specification events, and how it is altered in the absence critical regulatory factors, such as *Ptf1a* (Burlison et al., 2008; Kawaguchi et al., 2002; Krapp et al., 1998). An understanding of pancreatic MPC biology and the mechanisms that determine cell proliferation and fate specification, and particularly the role of *Ptf1a* in these processes, is important for multiple reasons. First, the number of pancreatic MPCs present early in development determines the final size of the pancreas

(Stanger et al., 2007). Second, the formation of pancreatic islets, which consist largely of insulin-secreting  $\beta$  cells, depends upon the orderly expansion of the MPC-containing, pre-pancreatic epithelium. Indeed, while some endocrine cells are formed in the absence of *Ptf1a*, there is an insufficient mass of these cells generated to maintain glucose homeostasis in the adult (Fukuda et al., 2008; Kawaguchi et al., 2002). Third, an understanding of step-wise changes in gene expression that occur during pancreatic organogenesis is essential for learning how to mimic these processes during the directed differentiation of pluripotent stem cells to make new pancreatic  $\beta$  cells for therapeutic use.

Several prior studies have explored the gene expression profile of pancreatic MPCs (Chiang and Melton, 2003; Gu et al., 2004; Svensson et al., 2007). These studies utilized early generation DNA microarrays and relied either on manual dissection methods or less than optimal genetic markers to obtain the target cell populations. Over recent years, technological advancements have occurred that enable the gene expression profiles of pancreatic MPCs to be more accurately determined. First, mice that express yellow fluorescent protein (YFP) under control of the *Ptf1a* gene locus were developed, thereby enabling fluorescence-activated cell sorting (FACS) to unambiguously isolate pancreatic progenitor cells (Burlison et al., 2008). Second, next-generation sequencing technologies have been developed that yield gene expression data based on the direct counting of transcripts, rather than hybridization, thereby providing results that have a far broader dynamic range than microarrays and allowing better detection of low abundance mRNAs (Marguerat and Bahler, 2010; Wang et al., 2009).

Here, I describe my work that provides a new and highly detailed view of the gene expression profile of pancreatic MPCs. In these studies, we examined the effects of eliminating *Ptf1a* on gene expression during early pancreas development and have profiled temporal changes in gene expression as pancreatic MPCs adopt an acinar cell fate. These

datasets provide new insights into the dynamic alterations within pancreatic MPCs that are essential for several key aspects of pancreas development and provide valuable new molecular markers for gaining a deeper understanding of pancreatic MPCs and their biology.

## Materials and Methods

### Tissue isolation, embryonic genotyping, and flow cytometry

Embryos were isolated from wild type CD-1 or *Ptf1a*<sup>YFP/+</sup> females crossed with *Ptf1a*<sup>YFP/+</sup> mice, where the presence of vaginal plug at noon was considered as E0.5. Fluorescence imaging was performed using a Leica MZ 16 FA stereoscope with a QImaging RETIGA 4000R camera. Heterozygous embryos were identified using a fluorescent microscope, and embryos obtained from heterozygous interbreeding were genotyped using the following two sets of custom-designed TaqMan primers and probes: 1) *Ptf1a*<sup>WT</sup> allele: 5'-CGAATTGCCACGGATCACT, 5'-CCCGGAAGGACGAATGG and 6FAM-ACAAAGCGTCACCCGA-MGBNFQ, and 2) *Ptf1a*<sup>YFP</sup> allele: 5'-GGGCACAAGCTGGAGTACAAC, 5'-TCTGCTTGTGGCCATGA, and VIC-ACAGCCACAACGTCT-MGBNFQ. Embryonic tissues were lysed with 10 mM Tris pH 8.3, 100 mM KCl, 2.5 mM MgCl<sub>2</sub>, 4.5% Tween, and 0.4 mg of Proteinase K for 15 min at 90°C. qPCR reactions used 1% of the genomic DNA per reaction with 800 nm forward and reverse primers and 200 nm probe in TaqMan Fast Universal PCR Mix (ABI). Analysis was performed using the ABI 7900HT.

Midgut regions containing the dorsal and ventral pancreatic rudiments were dissected from *Ptf1a*<sup>+/+</sup>, *Ptf1a*<sup>YFP/+</sup> or *Ptf1a*<sup>YFP/YFP</sup> embryos in cold phosphate buffered saline (PBS). *Ptf1a*<sup>YFP</sup>-expressing neural tubes were dissected for a FACS compensation control. Tissues were dissociated with Accumax (Sigma) at 37°C for up to 1.5 hours. The dissociation

reaction was quenched, and cells were triturated with FACS medium (L15 medium containing 1 mg/ml BSA, 10 mM HEPES pH 7.4, 1% penicillin/streptomycin, and 37.5 ng/ml DNase I). Following filtration through a 35 µm cell strainer (BD Falcon) and centrifugation, cells were resuspended in FACS medium or FACS medium with 7-Amino-Actinomycin D (7AAD; Invitrogen) to facilitate exclusion of nonviable cells.

Flow cytometry experiments were performed in the VMC Flow Cytometry Shared Resource on a BD Aria I or II. YFP was excited with a 488 nm laser and emission detected with a 502LP and 530/30 bandpass filter. Gates were established using wild type midguts and *Ptf1a<sup>YFP</sup>*-expressing neural tubes. The YFP gate was set conservatively to minimize sorting cellular autofluorescence in the 500 – 600 nm spectral region. Experimental protocols were approved by the Vanderbilt Institutional Animal Care and Use Committee.

### **RNA amplification, library construction, RNA-Sequencing and analysis**

Pools of 13, 9 or 5 pancreata were used to isolate *Ptf1a<sup>YFP/+</sup>* cells at E11.5, *Ptf1a<sup>YFP/YFP</sup>* cells at E11.5 or *Ptf1a<sup>YFP/+</sup>* cells at E15.5, respectively. Total RNA was isolated using TRIzol LS, DNase-treated, column-purified (Zymo) and assessed using an Agilent Bioanalyzer. RNA (0.5 – 10 ng) was amplified using the Ovation RNA-Seq system (NuGEN) which utilized Ribo-SPIA technology. The first step produced a cDNA/mRNA hybrid molecule. The mRNA within this complex is fragmented during the next step and allows for second strand synthesis and generation of a DNA/RNA heteroduplex double-stranded cDNA. Following SPIA amplification and post-SPIA modifications, the double-stranded cDNA was subjected to end repair using T4 DNA polymerase and Klenow polymerase. Klenow 3' and 5' exopolymerase activity was used to incorporate a single ‘A’ base, followed by ligation with a mix of adaptor oligonucleotides using T4 DNA ligase. Size selection was performed on the adaptor-ligated library on a 2% agarose gel, excising the library smear of

approximately 200 bp. Ten to 15 cycles of ligation mediated PCR (LM-PCR) were used to amplify the ligated material in preparation for cluster generation. Following ligation mediated PCR (LM-PCR), the samples were purified and quantitated using the Agilent Bioanalyzer. The resulting cDNA libraries were used at a final concentration of 2 – 4 pM to achieve a cluster density of approximately 160,000 clusters per tile and were analyzed as 36 nucleotide (nt) or greater reads as single-end tags on a Solexa/Illumina Genome Analyzer II. The Illumina pipeline (v1.4.0) was used for image analysis and base calling. Reads were mapped to mm9 genome with Bowtie (v0.11.3) (Langmead et al., 2009). Expression was quantified as reads per kilobase of exon model per million mapped reads (RPKM) (Mortazavi et al., 2008) using a custom Perl script (**Table 3.1**). Gene Ontology analysis was performed using PANTHER (v6.1) (Thomas et al., 2003). Differentially expressed transcripts were identified as those displaying a 5-fold change in RPKM and  $\geq 25$  reads in one sample.

### **Quantitative reverse transcription PCR (qRT-PCR) validation**

Customized 384- and 96-target TaqMan Low Density Arrays (TLDAs) were purchased from Applied Biosystems (ABI) containing the gene specific primers and probes listed in **Table 3.2**. The first phase of qRT-PCR was used as a preliminary platform to examine gene expression. Amplified RNA (WT-Ovation Pico, NuGEN) was analyzed as a single replicate from pooled embryos. 100 ng cDNA was loaded per port, and PCR amplification was performed using an ABI 7900HT. Transcripts for the 384-target TLDA were chosen based on the following criteria: 1) displayed at least a 10-fold difference in expression by RPKM value either between the *Ptf1a<sup>YFP/+</sup>* cells at E11.5 and E15.5 or the *Ptf1a<sup>YFP/YFP</sup>* and *Ptf1a<sup>YFP/+</sup>* cells at E11.5, 2) belonged to the following molecular functions: transcription factors, receptors, signaling molecules, proteases, protein phosphatases, or peptide hormones, and 3) had at least 25 reads in one sample.

**Table 3.1 Summary of RNA-Seq results**

Sample	Sequencing Run	Clusters	Reads	Mapped	Unique Hits	Non-unique Hits	% Mapped
<b>E11.5 Ptf1a<sup>YFP/+</sup></b>	1	145,281	13,094,931	11,527,374	7,639,627	3,887,747	88.03%
	2	166,258	15,086,822	13,728,633	8,832,758	4,895,875	91.00%
	3	180,267	18,537,850	17,134,442	11,469,961	5,664,481	92.43%
			46,719,603	42,390,449	27,942,346	14,448,103	90.49%
<hr/>							
<b>E11.5 Ptf1a<sup>YFP/YFP</sup></b>	1	107,399	9,668,230	8,053,904	5,434,532	2,619,372	83.30%
	2	149,220	13,427,730	11,427,043	7,570,072	3,856,971	85.10%
	3	156,355	15,828,875	14,109,808	9,380,331	4,729,477	89.14%
			38,924,835	33,590,755	22,384,935	11,205,820	85.85%
<hr/>							
<b>E15.5 Ptf1a<sup>YFP/+</sup></b>	1	121,738	11,033,718	9,584,948	5,804,116	3,780,832	86.87%
	2	168,329	15,223,528	13,931,260	8,032,988	5,898,272	91.51%
	3	180,626	18,309,790	17,029,893	10,189,780	6,840,113	93.01%
			44,567,036	40,546,101	24,026,884	16,519,217	90.46%
<hr/>							
	Total		130,211,474	116,527,305	74,354,165	42,173,140	88.93%

The second phase assessed expression across biological replicates ( $n \geq 3$ ), and only RNAs from the  $Ptf1a^{YFP/+}$  cells at E10.5 and  $Ptf1a^{YFP/YFP}$  cells at E11.5 were amplified (WT-Ovation Pico, NuGEN). Genes were selected for the second phase of qRT-PCR using the following criteria: 1) absolute Ct value of  $< 35$  in at least one sample and 2) the qRT-PCR profiles were consistent with RNA-Seq data. Additional filtering based on hierarchical clustering, Gene Ontology analysis and NCBI gene reports, led to the selection of 94 genes. RNA was reverse transcribed using the High Capacity cDNA Archive kit (ABI).

Genes with a Ct  $< 40$  were normalized to *Gapdh* to obtain  $2^{-\Delta Ct}$  values. Values were set to 0 (no expression) for genes with Ct = 40. Clustering was performed with input as the mean expression value across biological replicates for each condition. The GenePattern Hierarchical Clustering module (version 2.0) was employed with Pearson correlation as similarity measure and average-linkage as agglomeration method. Results were visualized with the GenePattern Hierarchical Clustering Viewer module using the relative display mode where colors are scaled as a function of the number of standard deviations relative to the mean of each row (darkest blue = -3 and brightest red = +3). To avoid distributional assumptions, statistical significance was determined using a permutation test ((Ewens and Grant, 2004) section 3.8.1) implemented with a custom Java script. Differential expression analysis comparing  $Ptf1a^{YFP/YFP}$  cells and  $Ptf1a^{YFP/+}$  cells at E11.5 and  $Ptf1a^{YFP/+}$  cells at E15.5 and E11.5 was performed with input as the expression values for the available biological replicates.

Meta-data and processed data for both the RNA-Seq and qRT-PCR, as well as qRT-PCR raw data are available at <http://genomics.betacell.org> and [www.cbil.upenn.edu/RAD](http://www.cbil.upenn.edu/RAD). The RNA-Seq component has been deposited at ArrayExpress (accession number E-MTAB-449) and the Sequence Read Archive (accession number ERP000419).

**Table 3.2 Summary of transcripts profiled on 384- and 96-format TLDA.** Transcripts noted in red are profiled on both the 384- and 96-format TLDA while transcripts in black are profiled only on the 384-format TLDA. Gene symbols in bold font note endogenous controls.

	GENE SYMBOL	GENE NAME	ASSAY ID
1	<b>18S</b>	Eukaryotic 18S rRNA	Hs99999901_s1
2	2210010C04Rik	RIKEN cDNA 2210010C04 gene;2210010C04Rik	Mm00834916_m1
3	2310046K01Rik	RIKEN cDNA 2310046K01 gene;2310046K01Rik	Mm00510189_m1
4	<b>9630041N07Rik</b>	zinc finger protein 879	Mm00557697_m1
5	A430110N23Rik	RIKEN cDNA A430110N23 gene;A430110N23Rik	Mm00557123_m1
6	Ace2	angiotensin I converting enzyme (peptidyl-dipeptidase A) 2;Ace2	Mm01159003_m1
7	<b>ACTB</b>	actin, beta	Mm01205647_g1
8	Acy3	aspartoacylase (aminoacylase) 3;Acy3	Mm00503584_m1
9	Adam33	a disintegrin and metalloproteinase domain 33;Adam33	Mm00459691_m1
10	<b>Adams2</b>	a disintegrin-like and metalloproteinase (reprolysin type) with thrombospondin type 1 motif, 2;Adamts2	Mm00805170_m1
11	Adipoq	adiponectin, C1Q and collagen domain containing;Adipoq	Mm00456425_m1
12	Adora1	adenosine A1 receptor;Adora1	Mm01308023_m1
13	<b>Adora2a</b>	adenosine A2a receptor;Adora2a	Mm00802075_m1
14	Adra2a	adrenergic receptor, alpha 2a;Adra2a	Mm00845383_s1
15	Ager	advanced glycosylation end product-specific receptor;Ager	Mm00545815_m1
16	Agtr1b	angiotensin II receptor, type 1b;Agtr1b	Mm01701115_m1
17	AI854703	expressed sequence AI854703;AI854703	Mm00624982_m1
18	Amac1	acyl-malonyl condensing enzyme 1;Amac1	Mm00840143_g1
19	Amy1	amylase 1, salivary	Mm00651524_m1
20	<b>Amy2</b>	amylase 2a5, pancreatic	Mm02342487_g1
21	Angptl3	angiopoietin-like 3;Angptl3	Mm00803820_m1
22	Angptl4	angiopoietin-like 4;Angptl4	Mm00480431_m1
23	Arhdig	Rho GDP dissociation inhibitor (GDI) gamma;Arhdig	Mm00801450_m1
24	Arid5a	AT rich interactive domain 5A (Mrf1 like);Arid5a	Mm00524454_m1
25	<b>Arx</b>	SUMO1 activating enzyme subunit 2;Sae2	Mm00545903_m1
26	Asgr2	asialoglycoprotein receptor 2;Asgr2	Mm00431863_m1
27	Barhl1	BarH-like 1 (Drosophila);Barhl1	Mm00479842_m1
28	Bgn	biglycan;Bgn	Mm00455918_m1
29	<b>Bhlhe40</b>	basic helix-loop-helix family, member e40	Mm00478593_m1
30	Bmp2	bone morphogenetic protein 2;Bmp2	Mm01340178_m1
31	Bmpr1b	bone morphogenetic protein receptor, type 1B;Bmpr1b	Mm00432117_m1
32	Bnc1	amiloride-sensitive cation channel 1, neuronal (degenerin);Accn1	Mm01324337_m1
33	<b>Calcr</b>	calcitonin receptor;Calcr	Mm00432271_m1
34	Casr	calcium-sensing receptor;Casr	Mm00443375_m1
35	Cbln3	cerebellin 3 precursor protein;Cbln3	Mm00490772_g1
36	Cck	cholecystokinin;Cck	Mm00446170_m1

**Table 3.2 continued**

37	Cckar	cholecystokinin A receptor;Cckar	Mm00438060_m1
38	Ccl5	chemokine (C-C motif) ligand 5;Ccl5	Mm01302428_m1
39	Ccl6	chemokine (C-C motif) ligand 6;Ccl6	Mm01302419_m1
40	Ccr4	chemokine (C-C motif) receptor 4;Ccr4	Mm00438271_m1
41	Cd180	CD180 antigen;Cd180	Mm00434804_m1
42	Cd2	CD2 antigen;Cd2	Mm00488928_m1
43	<b>Cd300a</b>	CD300A antigen;Cd300a	Mm00468054_m1
44	Cd36	scavenger receptor class B, member 1;Scarb1	Mm01135198_m1
45	Cd37	CD37 antigen;Cd37	Mm00514240_m1
46	Cd3e	CD3 antigen, epsilon polypeptide;Cd3e	Mm00599683_m1
47	Cd5	CD5 antigen;Cd5	Mm00432417_m1
48	<b>Cd93</b>	CD93 antigen;Cd93	Mm00440239_g1
49	Cdx2	caudal type homeo box 2;Cdx2	Mm00432449_m1
50	<b>Cela1/Ela1</b>	elastase 1, pancreatic;Ela1	Mm00712898_m1
51	Cela3b/Ela3b	elastase 3, pancreatic;Ela3	Mm00840378_m1
52	<b>Chgb</b>	chromogranin B;Chgb	Mm00483287_m1
53	Chrm1	cholinergic receptor, muscarinic 1, CNS;Chrm1	Mm00432509_s1
54	Chrna3	cholinergic receptor, nicotinic, alpha polypeptide 3;Chrna3	Mm00520145_m1
55	Chrna9	cholinergic receptor, nicotinic, alpha polypeptide 9;Chrna9	Mm01221611_m1
56	Chrnbl	cholinergic receptor, nicotinic, beta polypeptide 1 (muscle);Chrnbl	Mm00680412_m1
57	Clec2d	C-type lectin domain family 2, member d;Clec2d	Mm00474134_m1
58	Cnr2	protocadherin alpha 6;Pcdha6	Mm00438286_m1
59	<b>Cpa1</b>	carboxypeptidase A1;Cpa1	Mm00465942_m1
60	<b>Cpb1</b>	carboxypeptidase B1 (tissue);Cpb1	Mm01289391_m1
61	Creb3	cAMP responsive element binding protein 3;Creb3	Mm00457268_m1
62	Creb3l4	cAMP responsive element binding protein 3-like 4;Creb3l4	Mm00518698_m1
63	Crhr2	corticotropin releasing hormone receptor 2;Crhr2	Mm00438303_m1
64	Crlf2	cytokine receptor-like factor 3;Crlf3	Mm00497362_m1
65	Ctgf	connective tissue growth factor;Ctgf	Mm01192931_g1
66	<b>Ctrb1</b>	chymotrypsinogen B1;Ctrb1	Mm00481616_m1
67	Cx3cl1	chemokine (C-X3-C motif) ligand 1;Cx3cl1	Mm00436454_m1
68	Cx3cr1	chemokine (C-X3-C) receptor 1	Mm00438354_m1
69	Cxcl1	chemokine (C-X-C motif) ligand 1;Cxcl1	Mm00433859_m1
70	Cxcl16	chemokine (C-X-C motif) ligand 16;Cxcl16	Mm00469712_m1
71	Cxcr5	Burkitt lymphoma receptor 1;Blr1	Mm00432086_m1
72	Cxcr6	chemokine (C-X-C motif) receptor 6;Cxcr6	Mm00472858_m1
73	<b>Cysltr2</b>	cysteinyl leukotriene receptor 2;Cysltr2	Mm02620584_s1
74	Dll3	delta-like 3 (Drosophila);Dll3	Mm00432854_m1
75	Dtx3	deltex 3 homolog (Drosophila);Dtx3	Mm00472859_m1
76	E2f2	E2F transcription factor 2;E2f2	Mm00624964_m1

**Table 3.2 continued**

77	Edn3	endothelin 3;Edn3	Mm00432986_m1
78	Ednra	endothelin receptor type A;Ednra	Mm01243722_m1
79	Efna2	ephrin A2;Efna2	Mm00433011_m1
80	Efna3	ephrin A3;Efna3	Mm01212723_g1
81	<b>Efna4</b>	ephrin A4;Efna4	Mm00433013_m1
82	Efnb3	ephrin B3	Mm00433016_m1
83	<b>Egfr</b>	epidermal growth factor receptor;Egfr	Mm01187858_m1
84	Ephb3	Eph receptor B3	Mm00802553_m1
85	Ephb6	Eph receptor B6;Ephb6	Mm00432456_m1
86	<b>Etv4</b>	ets variant gene 4 (E1A enhancer binding protein, E1AF);Etv4	Mm00476696_m1
87	F12	coagulation factor XII (Hageman factor);F12	Mm00491349_m1
88	F2rl3	coagulation factor II (thrombin) receptor-like 3;F2rl3	Mm00433161_g1
89	Fas	fatty acid synthase;Fasn	Mm00433237_m1
90	Fbln2	fibulin 2	Mm00484266_m1
91	Fcer2a	Fc receptor, IgE, low affinity II, alpha polypeptide;Fcer2a	Mm00442792_m1
92	Fcgrt	Fc receptor, IgG, alpha chain transporter;Fcgrt	Mm00438887_m1
93	Fcrl1	Fc receptor-like A;Fcrl	Mm00462105_m1
94	<b>Fev</b>	FEV (ETS oncogene family);Fev	Mm00462220_m1
95	Ffar2	free fatty acid receptor 2;Ffar2	Mm01175249_g1
96	<b>Fga</b>	fibrinogen, alpha polypeptide;Fga	Mm00802584_m1
97	Fgb	fibrinogen, B beta polypeptide;Fgb	Mm00805336_m1
98	Fgf16	fibroblast growth factor 16;Fgf16	Mm00651404_m1
99	Fgf17	fibroblast growth factor 17;Fgf17	Mm00433282_m1
100	Fgf23	fibroblast growth factor 23;Fgf23	Mm00445621_m1
101	Fgg	fibrinogen, gamma polypeptide;Fgg	Mm00513575_m1
102	Fhl5	Muscle development	Mm00480451_m1
103	<b>Fosl1</b>	fos-like antigen 1;Fosl1	Mm00487429_m1
104	Gabra4	gamma-aminobutyric acid (GABA-A) receptor, subunit alpha 4;Gabra4	Mm00802631_m1
105	Gabra5	gamma-aminobutyric acid (GABA-A) receptor, subunit alpha 5;Gabra5	Mm00621092_m1
106	Gabrb3	gamma-aminobutyric acid (GABA-A) receptor, subunit beta 3;Gabrb3	Mm00433473_m1
107	<b>GAPDH</b>	glyceraldehyde-3-phosphate dehydrogenase	Mm99999915_g1
108	<b>Gast</b>	gastrin;Gast	Mm00772211_g1
109	Gata1	GATA binding protein 1;Gata1	Mm01352636_m1
110	Gata3	GATA binding protein 3;Gata3	Mm00484683_m1
111	Gbx1	gastrulation brain homeobox 1;Gbx1	Mm00803826_m1
112	<b>Gcg</b>	glucagon;Gcg	Mm00801714_m1
113	<b>Gegr</b>	glucagon receptor;Gcgr	Mm00433546_m1
114	Gdap1	ganglioside-induced differentiation-associated-protein 1;Gdap1	Mm00494579_m1
115	Gdap111	ganglioside-induced differentiation-associated protein 1-like 1;Gdap111	Mm00523187_m1
116	Gdf7	growth differentiation factor 7;Gdf7	Mm00807130_m1

**Table 3.2 continued**

117	Gdf9	growth differentiation factor 9;Gdf9	Mm00433565_m1
118	<b>Gdnf</b>	glial cell line derived neurotrophic factor;Gdnf	Mm00599849_m1
119	<b>Gfi1</b>	growth factor independent 1;Gfi1	Mm00515855_m1
120	Ghrh	growth hormone releasing hormone;Ghrh	Mm00439100_m1
121	Gip	gastric inhibitory polypeptide	Mm00433601_m1
122	<b>Gipr</b>	gastric inhibitory polypeptide receptor;Gipr	Mm01316351_g1
123	Glra1	glycine receptor, alpha 1 subunit;Glra1	Mm00445063_m1
124	Gnrrhr	gonadotropin releasing hormone receptor;Gnrrhr	Mm00439143_m1
125	Gp5	glycoprotein 5 (platelet);Gp5	Mm00515021_s1
126	Gpr1	G protein-coupled receptor 1;Gpr1	Mm00461557_m1
127	Gpr142	G protein-coupled receptor 142;Gpr142	Mm00725194_m1
128	Gpr157	G protein-coupled receptor 157;Gpr157	Mm00616683_m1
129	Gpr3	G-protein coupled receptor 3;Gpr3	Mm00433719_s1
130	Gpr37l1	G protein-coupled receptor 37-like 1;Gpr37l1	Mm00661872_m1
131	Gpr84	G protein-coupled receptor 84;Gpr84	Mm00518921_m1
132	<b>Gpr98</b>	G protein-coupled receptor 98;Gpr98	Mm00475232_m1
133	<b>Gria2</b>	Gria2 glutamate receptor, ionotropic, AMPA2 (alpha 2)	Mm00442822_m1
134	Gria3	glutamate receptor, ionotropic, AMPA3 (alpha 3);Gria3	Mm00497506_m1
135	Grin2c	glutamate receptor, ionotropic, NMDA2C (epsilon 3);Grin2c	Mm00439180_m1
136	Grm2	G protein-coupled receptor, family C, group 1, member B;Grm2	Mm01235831_m1
137	Guca2a	guanylate cyclase activator 2a (guanylin);Guca2a	Mm00433863_m1
138	<b>GUSB</b>	glucuronidase, beta	Mm00446953_m1
139	Hcrtr1	hypocretin (orexin) receptor 1;Hcrtr1	Mm01185776_m1
140	Hes3	hairy and enhancer of split 3 (Drosophila);Hes3	Mm00468603_m1
141	<b>Hey1</b>	hairy/enhancer-of-split related with YRPW motif 1;Hey1	Mm00468865_m1
142	Heyl	hairy/enhancer-of-split related with YRPW motif-like;Heyl	Mm00516555_m1
143	<b>Hoxa1</b>	homeo box A1;Hoxa1	Mm00439359_m1
144	<b>Hoxc4</b>	homeo box C4;Hoxc4	Mm00442838_m1
145	<b>HPRT</b>	hypoxanthine guanine phosphoribosyl transferase 1	Mm01324427_m1
146	Hr	hairless	Mm00498963_m1
147	Hrh4	histamine H4 receptor;Hrh4	Mm00467633_m1
148	Hsf4	heat shock transcription factor 4;Hsf4	Mm00442428_m1
149	Icam2	intercellular adhesion molecule 2;Icam2	Mm00494862_m1
150	Ifi204	interferon activated gene 204;Ifi204	Mm00492602_m1
151	Igf1	insulin-like growth factor 1;Igf1	Mm00439560_m1
152	Il17b	interleukin 17B;Il17b	Mm00444686_m1
153	Il17f	interleukin 17F;Il17f	Mm00521423_m1
154	Il1r2	Cytokine and chemokine mediated signaling pathway	Mm00439622_m1
155	Il22ra1	interleukin 22 receptor, alpha 1;Il22ra1	Mm00663697_m1
156	Il27ra	interleukin 27 receptor, alpha;Il27ra	Mm00497259_m1

**Table 3.2 continued**

157	<b>Il2rg</b>	interleukin 2 receptor, gamma chain;Il2rg	Mm00442885_m1
158	Il3ra	interleukin 3 receptor, alpha chain;Il3ra	Mm00434273_m1
159	Insl6	insulin-like 6;Insl6	Mm01964671_s1
160	<b>Insm1</b>	insulinoma-associated 1	Mm02581025_s1
161	<b>Insrr</b>	insulin receptor-related receptor;Insrr	Mm00442243_m1
162	Irf5	interferon regulatory factor 5;Irf5	Mm00496477_m1
163	Irx1	Iroquois related homeobox 1 (Drosophila);Irx1	Mm01352526_m1
164	Irx2	Iroquois related homeobox 2 (Drosophila);Irx2	Mm01340315_m1
165	<b>Isl1</b>	ISL1 transcription factor, LIM/homeodomain;Isl1	Mm00627860_m1
166	<b>Isx</b>	intestine specific homeobox;Isx	Mm01243745_m1
167	<b>Itgb8</b>	integrin beta 8;Itgb8	Mm00623991_m1
168	Kir3dl2	killer cell immunoglobulin-like receptor, three domains, long cytoplasmic tail, 2;Kir3dl2	Mm00844610_s1
169	Kiss1r	KISS1 receptor;Kiss1r	Mm00475046_m1
170	<b>Klf15</b>	Kruppel-like factor 15;Klf15	Mm00517792_m1
171	Klf2	Kruppel-like factor 2 (lung);Klf2	Mm00500486_g1
172	Klhdc8a	kelch domain containing 8A;Klhdc8a	Mm00522717_m1
173	Klk1	kallikrein 1;Klk1	Mm00834006_g1
174	Klk1b24	kallikrein 1-related peptidase b24;Klk1b24	Mm00658591_g1
175	Klk1b27	kallikrein 1-related peptidase b27;Klk1b27	Mm00834759_gH
176	Klk1b3	kallikrein 1-related peptidase b3;Klk1b3	Mm01203825_gH
177	Klk1b5	kallikrein 1-related peptidase b5;Klk1b5	Mm00833453_g1
178	Klk1b8	kallikrein 1-related peptidase b8;Klk1b8	Mm00776302_g1
179	Lair1	leukocyte-associated Ig-like receptor 1;Lair1	Mm00618113_m1
180	Ldb3	LIM domain binding 3	Mm00522021_m1
181	Lect2	leukocyte cell-derived chemotaxin 2;Lect2	Mm00521920_m1
182	Lgals9	lectin, galactose binding, soluble 9;Lgals9	Mm00495295_m1
183	Lgi3	leucine-rich repeat LGI family, member 3;Lgi3	Mm00507490_m1
184	<b>Lhx1</b>	LIM homeobox protein 1;Lhx1	Mm00521776_m1
185	Lhx3	LIM homeobox protein 3;Lhx3	Mm01333633_m1
186	Lhx4	LIM homeobox protein 4;Lhx4	Mm00521928_m1
187	Lhx5	LIM homeobox protein 5;Lhx5	Mm00521778_m1
188	Lilrb3	leukocyte immunoglobulin-like receptor, subfamily B (with TM and ITIM domains), member 3;Lilrb3	Mm01700366_m1
189	<b>Lin28</b>	lin-28 homolog (C. elegans);Lin28	Mm00524077_m1
190	<b>Loxl2</b>	lysyl oxidase-like 2;Loxl2	Mm00804740_m1
191	Loxl4	lysyl oxidase-like 4;Loxl4	Mm00446385_m1
192	Lrfn3	leucine rich repeat and fibronectin type III domain containing 3	Mm00615455_m1
193	<b>Lrp1</b>	low density lipoprotein receptor-related protein 1;Lrp1	Mm00464608_m1
194	Lrrc3b	leucine rich repeat containing 3B;Lrrc3b	Mm00524764_m1
195	Lrrtm1	leucine rich repeat transmembrane neuronal 1;Lrrtm1	Mm00551337_g1

**Table 3.2 continued**

196	Lrrtm3	leucine rich repeat transmembrane neuronal 3;Lrrtm3	Mm00618457_m1
197	Lsp1	lymphocyte specific 1;Lsp1	Mm00497788_m1
198	Ltk	leukocyte tyrosine kinase;Ltk	Mm00434790_m1
199	<b>Mafb</b>	v-maf musculoaponeurotic fibrosarcoma oncogene family, protein B (avian);Mafb	Mm00627481_s1
200	Mds1	myelodysplasia syndrome 1 homolog (human);Mds1	Mm00491303_m1
201	Mef2c	myocyte enhancer factor 2C;Mef2c	Mm01340839_m1
202	<b>Meox1</b>	mesenchyme homeobox 1;Meox1	Mm00440285_m1
203	Mip	major intrinsic protein of eye lens fiber;Mip	Mm00434949_m1
204	Mixl1	Mix1 homeobox-like 1 ( <i>Xenopus laevis</i> );Mixl1	Mm00489085_m1
205	Mlf1	myeloid leukemia factor 1;Mlf1	Mm00440290_m1
206	Mmp2	matrix metallopeptidase 2;Mmp2	Mm00439506_m1
207	Mpl	myeloproliferative leukemia virus oncogene;Mpl	Mm00440310_m1
208	Msx1	homeo box, msh-like 1;Msx1	Mm00440330_m1
209	<b>Mt1</b>	metallothionein 1;Mt1	Mm00496660_g1
210	Mt2	metallothionein 2;Mt2	Mm00809556_s1
211	Myoc	myocilin;Myoc	Mm00447900_m1
212	<b>Myt1</b>	myelin transcription factor 1;Myt1	Mm00456190_m1
213	Necab2	EF hand calcium binding protein 2;Efcbp2	Mm00475387_m1
214	<b>Nepn</b>	nephrocan;Nepn	Mm00481816_m1
215	<b>Neurod1</b>	neurogenic differentiation 1	Mm01280117_m1
216	Neurog2	neurogenin 2;Neurog2	Mm00437603_g1
217	<b>Neurog3</b>	neurogenin 3;Neurog3	Mm00437606_s1
218	NfkB2	nuclear factor of kappa light polypeptide gene enhancer in B-cells 2, p49/p100;NfkB2	Mm00479807_m1
219	Nhlh1	nescent helix loop helix 1;Nhlh1	Mm00440478_m1
220	<b>Nkx6-3</b>	NK6 transcription factor related, locus 3 ( <i>Drosophila</i> );Nkx6-3	Mm01211932_g1
221	<b>Nlgn3</b>	neuroligin 3;Nlgn3	Mm00556834_m1
222	Nodal	nodal;Nodal	Mm00443040_m1
223	<b>Notch3</b>	Notch gene homolog 3 ( <i>Drosophila</i> );Notch3	Mm00435270_m1
224	Nrlh3	nuclear receptor subfamily 1, group H, member 3;Nr1h3	Mm00443454_m1
225	<b>Nrlh4</b>	nuclear receptor subfamily 1, group H, member 4;Nr1h4	Mm00436419_m1
226	Nrlh5	nuclear receptor subfamily 1, group H, member 5;Nr1h5	Mm01308716_m1
227	Nr2e1	nuclear receptor subfamily 2, group E, member 1	Mm00455855_m1
228	Nr2f1	nuclear receptor subfamily 2, group F, member 1;Nr2f1	Mm00657937_m1
229	<b>Nr4a1</b>	nuclear receptor subfamily 4, group A, member 1;Nr4a1	Mm01300401_m1
230	Nr4a3	nuclear receptor subfamily 4, group A, member 3;Nr4a3	Mm00450074_m1
231	Nr5a1	nuclear receptor subfamily 5, group A, member 1;Nr5a1	Mm00446826_m1
232	Nrl	neural retina leucine zipper gene;Nrl	Mm00476550_m1
233	Nrp2	neuropilin 2;Nrp2	Mm00803099_m1
234	Nrtn	neurturin;Nrtn	Mm00435387_m1

**Table 3.2 continued**

235	Ntf3	neurotrophin 3;Ntf3	Mm01182924_m1
236	Ntsr2	neurotensin receptor 2;Ntsr2	Mm00435426_m1
237	<b>Nxph1</b>	neurexophilin 1;Nxph1	Mm01165166_m1
238	<b>Nxph2</b>	neurexophilin 2;Nxph2	Mm00801892_m1
239	Nxph4	neurexophilin 4;Nxph4	Mm00806440_m1
240	Olfm2	olfactomedin 2;Olfm2	Mm00620619_m1
241	Olfml3	olfactomedin-like 3;Olfml3	Mm00513567_m1
242	Olfr267	olfactory receptor 267;Olfr267	Mm00528927_s1
243	<b>Onecut2</b>	one cut domain, family member 2;Onecut2	Mm00815708_s1
244	<b>Onecut3</b>	one cut domain, family member 3;Onecut3	Mm00653012_m1
245	Oprl1	opioid receptor-like 1;Oprl1	Mm00440563_m1
246	P2rx1	purinergic receptor P2X, ligand-gated ion channel, 1;P2rx1	Mm00435460_m1
247	P2ry1	purinergic receptor P2Y, G-protein coupled 1;P2ry1	Mm00435471_m1
248	Pax2	paired box gene 2;Pax2	Mm01217939_m1
249	<b>Pax6</b>	paired box gene 6;Pax6	Mm00443081_m1
250	<b>Pbx4</b>	pre-B-cell leukemia transcription factor 4;Pbx4	Mm00453088_m1
251	Pecam1	platelet/endothelial cell adhesion molecule 1;Pecam1	Mm00476702_m1
252	Penk1	preproenkephalin 1;Penk1	Mm01212875_m1
253	Pf4	chemokine (C-X-C motif) ligand 4;Cxcl4	Mm00451315_g1
254	Pgp	ATP-binding cassette, sub-family B (MDR/TAP), member 1A;Abcb1a	Mm00546522_g1
255	<b>Phf1</b>	PHD finger protein 1;Phf1	Mm00493478_m1
256	Pira2	paired-Ig-like receptor A2;Pira2	Mm02768273_g1
257	<b>Pitx2</b>	paired-like homeodomain transcription factor 2;Pitx2	Mm00440826_m1
258	Plg	plasminogen;Plg	Mm00447087_m1
259	Plxnc1	plexin C1;Plxnc1	Mm00450687_m1
260	Plxnd1	plexin D1;Plxnd1	Mm01184367_m1
261	<b>Pnliprp1</b>	pancreatic lipase related protein 1	Mm00479741_m1
262	Podnl1	podocan-like 1;Podnl1	Mm01247693_m1
263	Pou1f1	POU domain, class 1, transcription factor 1 (Pit1, growth hormone factor 1);Pou1f1	Mm00476852_m1
264	Pou3f4	POU domain, class 3, transcription factor 4;Pou3f4	Mm00447171_s1
265	Pou6f2	POU domain, class 6, transcription factor 2;Pou6f2	Mm00558158_m1
266	Ppara	peroxisome proliferator activated receptor alpha	Mm00440939_m1
267	Ppbp	chemokine (C-X-C motif) ligand 7;Cxcl7	Mm00470163_m1
268	Ppp1rla	protein phosphatase 1, regulatory (inhibitor) subunit 1A;Ppp1rla	Mm00451727_m1
269	<b>Ppp1rlb</b>	protein phosphatase 1, regulatory (inhibitor) subunit 1B;Ppp1rlb	Mm00454892_m1
270	Prdm12	PR domain containing 12;Prdm12	Mm01324476_m1
271	Prdm14	PR domain containing 14;Prdm14	Mm01237814_m1
272	Procr	protein C receptor, endothelial;Procr	Mm00440992_m1
273	Prok1	prokineticin 2;Prok2	Mm01204733_m1

**Table 3.2 continued**

274	<b>Prokr1</b>	prokineticin receptor 1;Prokr1	Mm00517546_m1
275	<b>Prph2</b>	peripherin 2;Prph2	Mm00435972_m1
276	<b>Prss16</b>	protease, serine, 16 (thymus);Prss16	Mm00457601_m1
277	<b>Prss2</b>	protease, serine, 2;Prss2	Mm00657001_m1
278	<b>Prss27</b>	protease, serine 27;Prss27	Mm00841353_g1
279	<b>Psmb9</b>	proteosome (prosome, macropain) subunit, beta type 9 (large multifunctional peptidase 2);Psmb9	Mm00479004_m1
280	<b>Ptafr</b>	platelet-activating factor receptor;Ptafr	Mm02621061_m1
281	<b>Ptf1a</b>	pancreas specific transcription factor, 1a	Mm00479622_m1
282	<b>Ptger4</b>	prostaglandin E receptor 4 (subtype EP4);Ptger4	Mm00436053_m1
283	<b>Ptgir</b>	prostaglandin I receptor (IP);Ptgir	Mm00801939_m1
284	<b>Pth1r</b>	parathyroid hormone receptor 1;Pthr1	Mm00441046_m1
285	<b>Pyy</b>	peptide YY;Pyy	Mm00520715_m1
286	<b>Rarg</b>	retinoic acid receptor, gamma;Rarg	Mm00441091_m1
287	<b>Rasgrp4</b>	RAS guanyl releasing protein 4;Rasgrp4	Mm00460898_m1
288	<b>Rbp1</b>	retinol binding protein 1, cellular;Rbp1	Mm00441119_m1
289	<b>Rbpj</b>	recombination signal binding protein for immunoglobulin kappa J region	Mm03053645_s1
290	<b>Rbpjl</b>	recombination signal binding protein for immunoglobulin kappa J region-like	Mm00485631_m1
291	<b>Reep2</b>	receptor accessory protein 2;Reep2	Mm00523031_m1
292	<b>Reep5</b>	receptor accessory protein 5;Reep5	Mm00492230_m1
293	<b>Reg1</b>	regenerating islet-derived 1;Reg1	Mm00485651_m1
294	<b>Relt</b>	tumor necrosis factor receptor superfamily, member 19-like;Tnfrsf19l	Mm00723872_m1
295	<b>Ren1</b>	renin 1 structural;Ren1	Mm02342889_g1
296	<b>Ret</b>	ret proto-oncogene;Ret	Mm00436304_m1
297	<b>Rln1</b>	relaxin 1;Rln1	Mm01208503_m1
298	<b>Rnase1</b>	ribonuclease, RNase A family, 1 (pancreatic)	Mm00726747_s1
299	<b>Robo2</b>	roundabout homolog 2 (Drosophila);Robo2	Mm00620713_m1
300	<b>Ror1</b>	RAR-related orphan receptor alpha;Rora	Mm00443462_m1
301	<b>Rorc</b>	RAR-related orphan receptor gamma;Rorc	Mm01261022_m1
302	<b>Rtn4r</b>	reticulon 4 receptor;Rtn4r	Mm00452228_m1
303	<b>Rtn4rl2</b>	reticulon 4 receptor-like 2;Rtn4rl2	Mm01336368_g1
304	<b>Rxfp4</b>	relaxin family peptide receptor 4;Rxfp4	Mm00731536_s1
305	<b>Rxrg</b>	retinoid X receptor gamma;Rxrg	Mm00436410_m1
306	<b>Sag</b>	ring finger protein 7;Rnf7	Mm00485903_m1
307	<b>Sall1</b>	sal-like 1 (Drosophila);Sall1	Mm00491266_m1
308	<b>Sall4</b>	sal-like 4 (Drosophila);Sall4	Mm01240680_m1
309	<b>Scarf1</b>	scavenger receptor class F, member 1;Scarf1	Mm00464144_m1
310	<b>Scg2</b>	secretogranin II;Scg2	Mm00843883_s1
311	<b>Scrt1</b>	scratch homolog 1, zinc finger protein (Drosophila);Scrt1	Mm00459966_m1
312	<b>Sct</b>	secretin;Sct	Mm00441235_g1

**Table 3.2 continued**

313	Sdc3	syndecan 3;Sdc3	Mm01179832_m1
314	Sdpr	serum deprivation response	Mm00507087_m1
315	Sds	serine dehydratase;Sds	Mm00455126_m1
316	Selp	selectin, platelet;Selp	Mm00441295_m1
317	Selplg	selectin, platelet (p-selectin) ligand;Selpl	Mm01204601_m1
318	Sema5a	sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5A;Sema5a	Mm00436500_m1
319	Sfrp1	secreted frizzled-related sequence protein 1;Sfrp1	Mm00489161_m1
320	Sfrp2	secreted frizzled-related protein 2;Sfrp2	Mm01213947_m1
321	Sfrp5	secreted frizzled-related sequence protein 5;Sfrp5	Mm00490044_m1
322	Shh	sonic hedgehog;Shh	Mm00436528_m1
323	Slit1	slit homolog 1 (Drosophila);Slit1	Mm01198620_m1
324	Sox10	SRY-box containing gene 10;Sox10	Mm01300162_m1
325	Sox11	SRY-box containing gene 11;Sox11	Mm01281943_s1
326	Sox5	SRY-box containing gene 5;Sox5	Mm00488381_m1
327	Sp6	trans-acting transcription factor 6	Mm02527757_s1
328	Spdef	SAM pointed domain containing ets transcription factor;Spdef	Mm00600221_m1
329	Spib	Spi-B transcription factor (Spi-1/PU.1 related);Spib	Mm01719550_s1
330	Spred3	sprouty-related, EVH1 domain containing 3;Spred3	Mm00805393_g1
331	Sstr3	somatostatin receptor 3;Sstr3	Mm00436695_s1
332	Stambpl1	Stam binding protein like 1;Stambpl1	Mm00472562_m1
333	Syp	Regulated exocytosis;Neurotransmitter release	Mm00436850_m1
334	Syt1l	synaptotagmin-like 1;Syt1l	Mm00473300_m1
335	Syt2l	synaptotagmin-like 2;Syt2l	Mm00473315_m1
336	T	brachyury;T	Mm01318252_m1
337	Taf7l	TAF7-like RNA polymerase II, TATA box binding protein (TBP)-associated factor;Taf7l	Mm00459354_m1
338	Tal1	T-cell acute lymphocytic leukemia 1;Tal1	Mm00441665_m1
339	Tbx3	T-box 3;Tbx3	Mm00809779_s1
340	Tbxa2r	thromboxane A2 receptor;Tbxa2r	Mm00436917_m1
341	Tcf15	transcription factor 15;Tcf15	Mm00493442_m1
342	Tcf15	transcription factor-like 5 (basic helix-loop-helix);Tcf15	Mm00626495_m1
343	Tgfb1	transforming growth factor, beta 1;Tgfb1	Mm00441724_m1
344	Theg	testicular haploid expressed gene;Theg	Mm00803557_m1
345	Tlr13	toll-like receptor 13;Tlr13	Mm01233819_m1
346	Tlr7	toll-like receptor 7;Tlr7	Mm00446590_m1
347	Tmprss4	transmembrane protease, serine 4;Tmprss4	Mm00520486_m1
348	Tmprss5	transmembrane protease, serine 5 (spinesin);Tmprss5	Mm00446105_m1
349	Tnfrsf18	tumor necrosis factor receptor superfamily, member 18;Tnfrsf18	Mm00437136_m1
350	Tnfrsf25	tumor necrosis factor receptor superfamily, member 25;Tnfrsf25	Mm01263821_m1
351	Tnfsf10	tumor necrosis factor (ligand) superfamily, member 10;Tnfsf10	Mm01283606_m1

**Table 3.2 continued**

352	Trpm5	transient receptor potential cation channel, subfamily M, member 5;Trpm5	Mm00498453_m1
353	Tshr	thyroid stimulating hormone receptor;Tshr	Mm00442027_m1
354	<b>Tshz3</b>	teashirt zinc finger family member 3;Tshz3	Mm00549466_m1
355	Tspan17	tetraspanin 17	Mm00512218_m1
356	<b>Tspan8</b>	tetraspanin 8;Tspan8	Mm00524563_m1
357	<b>Unc13a</b>	unc-13 homolog A (C. elegans);Unc13a	Mm00550016_s1
358	<b>Unc13c</b>	unc-13 homolog C (C. elegans);Unc13c	Mm00463432_m1
359	<b>Unc5c</b>	unc-5 homolog C (C. elegans);Unc5c	Mm00494093_m1
360	V1rd3	vomeronasal 1 receptor, D3;V1rd3	Mm00835727_s1
361	Vax2	ventral anterior homeobox containing gene 2;Vax2	Mm00496315_m1
362	<b>Vdr</b>	vitamin D receptor;Vdr	Mm00437297_m1
363	Vgll2	vestigial like 2 homolog (Drosophila);Vgll2	Mm00464115_m1
364	Wnt11	wingless-related MMTV integration site 11;Wnt11	Mm00437328_m1
365	<b>Wnt16</b>	wingless-related MMTV integration site 16;Wnt16	Mm00446420_m1
366	Wnt3	wingless-related MMTV integration site 3;Wnt3	Mm00437336_m1
367	<b>Wnt4</b>	wingless-related MMTV integration site 4;Wnt4	Mm01194003_m1
368	Wnt5a	wingless-related MMTV integration site 5A;Wnt5a	Mm00437347_m1
369	Wnt7a	wingless-related MMTV integration site 7A;Wnt7a	Mm00437355_m1
370	Wnt7b	wingless-related MMTV integration site 7B;Wnt7b	Mm00437357_m1
371	Wnt8a	wingless-related MMTV integration site 8A;Wnt8a	Mm00436822_m1
372	Wnt9a	wingless-type MMTV integration site 9A;Wnt9a	Mm00460518_m1
373	<b>Xpnpep2</b>	X-prolyl aminopeptidase (aminopeptidase P) 2, membrane-bound;Xpnpep2	Mm00460007_m1
374	Zbtb7b	zinc finger and BTB domain containing 7B;Zbtb7b	Mm00784709_s1
375	<b>Zfp462</b>	zinc finger protein 462;Zfp462	Mm00556263_m1
376	Zfp513	zinc finger protein 513;Zfp513	Mm00614402_m1
377	<b>Zfp57</b>	zinc finger protein 57;Zfp57	Mm00456405_m1
378	Zfp811	zinc finger protein 811;Zfp811	Mm00806009_m1
379	<b>Zfp92</b>	zinc finger protein 92;Zfp92	Mm00494326_m1
380	Zp2	zona pellucida glycoprotein 2;Zp2	Mm00442173_m1
381	Zp3	zona pellucida glycoprotein 3;Zp3	Mm00442176_m1

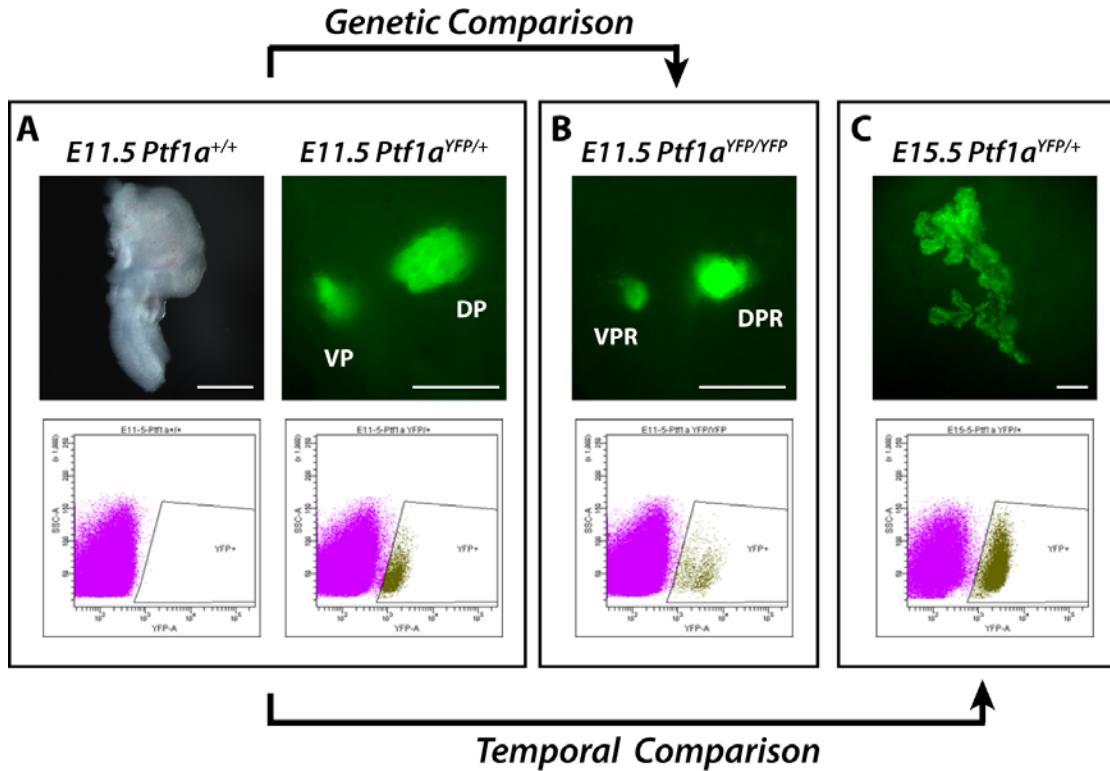
## Results

### Isolation of three distinct *Ptf1a*<sup>YFP</sup>-expressing cell populations

Using FACS, we isolated three different pancreatic cell populations from mouse embryos that express Citrine, a yellow fluorescent protein (YFP), under control of the *Ptf1a* gene locus (Burlison et al., 2008), as illustrated in **Figure 3.1**. By conservatively setting the FACS isolation gates, the cell populations we obtained had less than one percent contamination as assessed by autofluorescence from wild type cells. An average of 1.06 x 10<sup>3</sup>, 4.23 x 10<sup>2</sup>, and 1.42 x 10<sup>4</sup> cells were isolated per *Ptf1a*<sup>YFP/+</sup> embryo at E11.5, *Ptf1a*<sup>YFP/YFP</sup> embryo at E11.5, or *Ptf1a*<sup>YFP/+</sup> embryo at E15.5, respectively. RNA from these cells was then analyzed by RNA-Seq.

### Whole transcriptome profiling

It has been suggested that the sequencing of > 40 million reads is required to detect and quantify RNAs from biologically relevant classes (Mortazavi 2008). For this reason, we performed three lanes of whole transcriptome sequencing (Illumina) on each of the three samples from which we obtained 38.9 to 46.7 million reads of at least 36 nucleotides (nt) for each of the three sample RNAs, as summarized in **Table 3.1**. Approximately 89% of these reads aligned to the mouse genome (mm9) using the short-read aligner, Bowtie (Langmead et al., 2009), thereby resulting in 42.4, 33.6 and 40.5 million mapped reads for the E11.5 *Ptf1a*<sup>YFP/+</sup>, E11.5 *Ptf1a*<sup>YFP/YFP</sup>, and E15.5 *Ptf1a*<sup>YFP/+</sup> cells, respectively (**Table 3.1**). To allow comparison of transcript levels between samples, transcripts were quantified in reads per kilobase of exon model per million mapped reads (RPKM), which normalizes the sample based on transcript length and the total number of mapped reads following sequencing (Mortazavi et al., 2008).

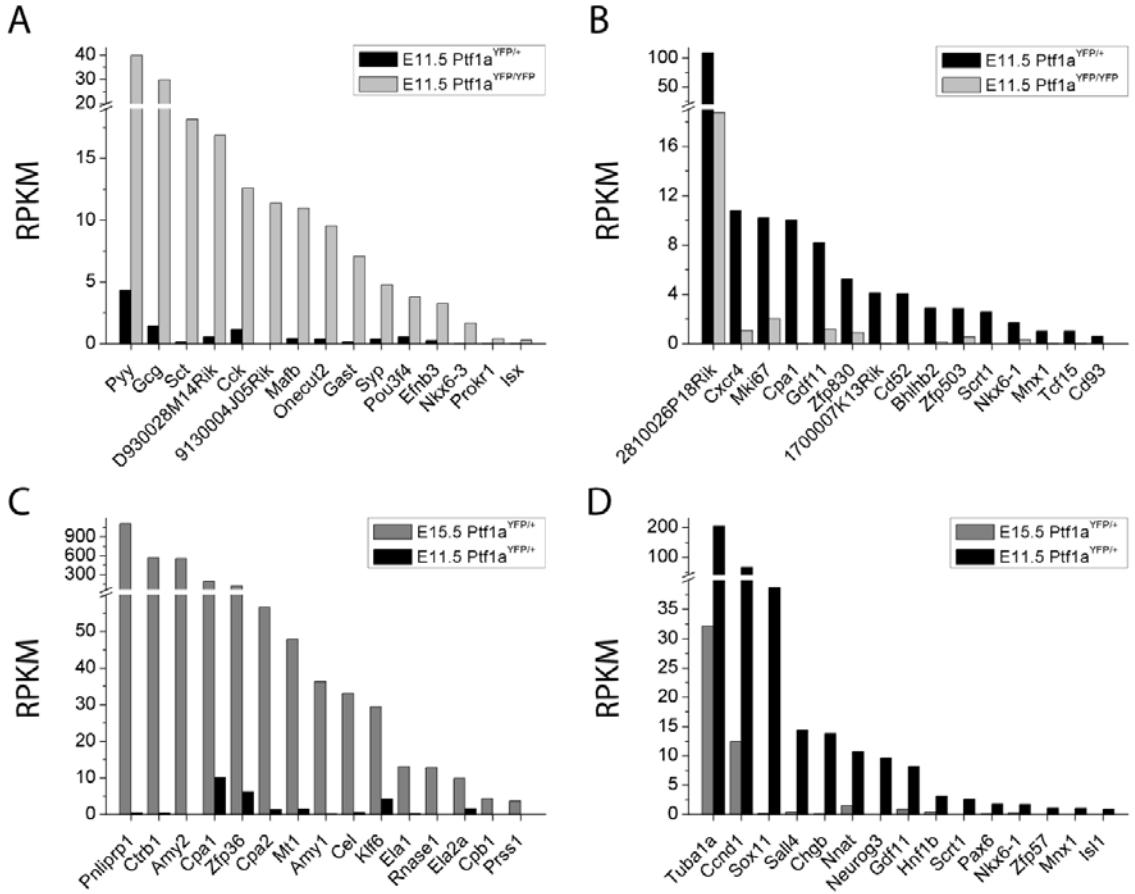


**Figure 3.1 Genetic and temporal comparisons of cells isolated from pancreatic MPCs, *Ptf1a*-deficient progenitor cells, and nascent acinar-specified cells.** A) YFP expression was observed in the dorsal (DP) and ventral (VP) pancreatic buds of E11.5 *Ptf1a*<sup>YFP/+</sup> embryos by whole mount fluorescence microscopy. Wild type embryos were used to establish sorting parameters (left), resulting in less than one percent contamination from autofluorescent cells when *Ptf1a*<sup>YFP</sup>-positive MPCs were isolated via FACS (right). B) In *Ptf1a*<sup>YFP/YFP</sup> mice, the expansion of the pancreatic epithelium was severely inhibited by E11.5, leaving only dorsal (DPR) and ventral (VPR) pancreatic remnants. The genetic comparison of E11.5 *Ptf1a*<sup>YFP/+</sup> and *Ptf1a*<sup>YFP/YFP</sup> expression profiles permitted analysis of genes that are genetically dependent on *Ptf1a*. C) By E15.5, YFP expression was observed broadly throughout the pancreas, indicative of acinar-cell expansion. The temporal comparison of E11.5 and E15.5 *Ptf1a*<sup>YFP/+</sup> cells enabled analysis of the changes seen as *Ptf1a*-expressing cells transition from a multipotent to a differentiated state. Scale bars = 30 $\mu$ m.

### Ptf1a-dependent gene expression at E11.5

1,105 genes were identified in the RNA-Seq analyses as increased in the *Ptf1a*-deficient progenitors compared to *Ptf1a*-expressing MPCs. Among these were pre-endocrine and endocrine markers, such as *Gcg*, *Mafb* and *Pou3f4* (Jorgensen et al., 2007), as well as enteroendocrine and gastrointestinal markers, such as *Pyy*, *Sct*, *Cck*, *Gast*, *Syp*, *Prokr1* and *Isx* (Sancho et al., 2004) (**Figure 3.2**). The presence of pre-endocrine and gastrointestinal markers in the *Ptf1a*-deficient cells is consistent with prior results showing the existence of endocrine cells in the dorsal duct-like structure of E18.5 *Ptf1a*-null embryos and the lineage tracing studies that revealed the presence of morphologically-indistinguishable *Ptf1a*-null cells in the developing duodenum and common bile duct epithelia (Burlison et al., 2008; Kawaguchi et al., 2002). However, the detection of these markers at E11.5 suggests that the differentiation of *Ptf1a*-deficient cells towards other fates begins much earlier than was previously determined.

Conversely, there were 997 genes that were more highly expressed in the *Ptf1a*-expressing cells compared to the *Ptf1a*-null cells at E11.5. Among these were *Cpa1*, *Gdf11*, *Nkx6-1*, and *Mnx1* (*Hlx9*), all of which have previously been identified to be expressed during pancreas development (**Figure 3.2**). Although it has been shown that the Ptf1a-containing PTF1 complex binds to the *Pdx1* promoter (Wiebe et al., 2007), we did not see a significant change in *Pdx1* expression between the E11.5 *Ptf1a*-expressing and *Ptf1a*-deficient cells. Other genes that were found to be more abundant in the *Ptf1a*-expressing MPCs were transcription factors such as *Scrt1*, *Zfp830*, *Tcf15*, and *Bhlhb2*; CD antigens and other cell surface markers such as *Cd52*, *Cd93*, *Cd4*, *Cxcr4*, and *Ager*; and markers of cell proliferation such as *mKi67* and *Cenpf*.



**Figure 3.2 Genetic and temporal comparisons of differentially expressed transcripts as determined by RNA-Seq.** A and B) The genetic comparison of *Ptf1a*-deficient progenitors (light gray bars) and *Ptf1a*-expressing progenitors (black bars) revealed increased levels of enteroendocrine and gastrointestinal transcripts in *Ptf1a*-deficient progenitors (A) and transcripts upregulated by the expression of *Ptf1a* (B). C and D) The temporal comparison of acinar-specified cells (dark gray bars) and *Ptf1a*-expressing progenitors (black bars) revealed increased transcript levels of numerous enzymes in acinar-specified cells (C) and transcripts upregulated in the progenitors (D). Differential expression was defined as being at least a 5-fold change in RPKM.

### Temporal changes as MPCs are acinar-specified

In the temporal comparison, 657 genes were identified as upregulated at E15.5 as compared to E11.5 based on the RNA-Seq analyses. As expected, many different proteases, hydrolases and other enzymes were highly expressed in the E15.5 *Ptf1a*-expressing cells (**Figure 3.2**). These findings closely paralleled a prior study that examined transcript levels of ten digestive enzymes during pancreas development as profiled in the rat (Han et al., 1986). Specifically, *Ctrb1*, *Amy2* and *Cpa1* were expressed at the highest levels, whereas *Cpa2*, *Ela1*, *Rnase1*, *Ela2a*, *Cpb1* and *Prss1* (*Trypsin*) were less abundant.

Conversely, there were 2,136 genes identified as downregulated at E15.5 compared to E11.5. These genes included pancreatic markers such as *Mnx1* (*Hlx9*), *Neurog3*, *Nkx6-1*, *Hnf1b*, *Pax6* and *Isl1*, as well as other factors with previously identified roles in pancreas and endocrine development such as *Gdf11* (Harmon et al., 2004), *Chgb* (Obermuller et al., 2010), and *Nnat* (Chu and Tsai, 2005) (**Figure 3.2**). Interestingly, *Sall4*, which plays a crucial role in controlling lineage commitment of hepatoblasts and regulating stem cell pluripotency (Oikawa et al., 2009); *Zfp57*, which plays a role in imprinting and has been associated with transient neonatal diabetes (Mackay et al., 2008); *Scrt*, which is expressed in the developing and adult brain (Marin and Nieto, 2006); and *Nepn*, which is expressed in the midgut and hindgut endoderm (Hou et al., 2007), were also more abundant in the pancreatic MPCs compared to the acinar-specified cells.

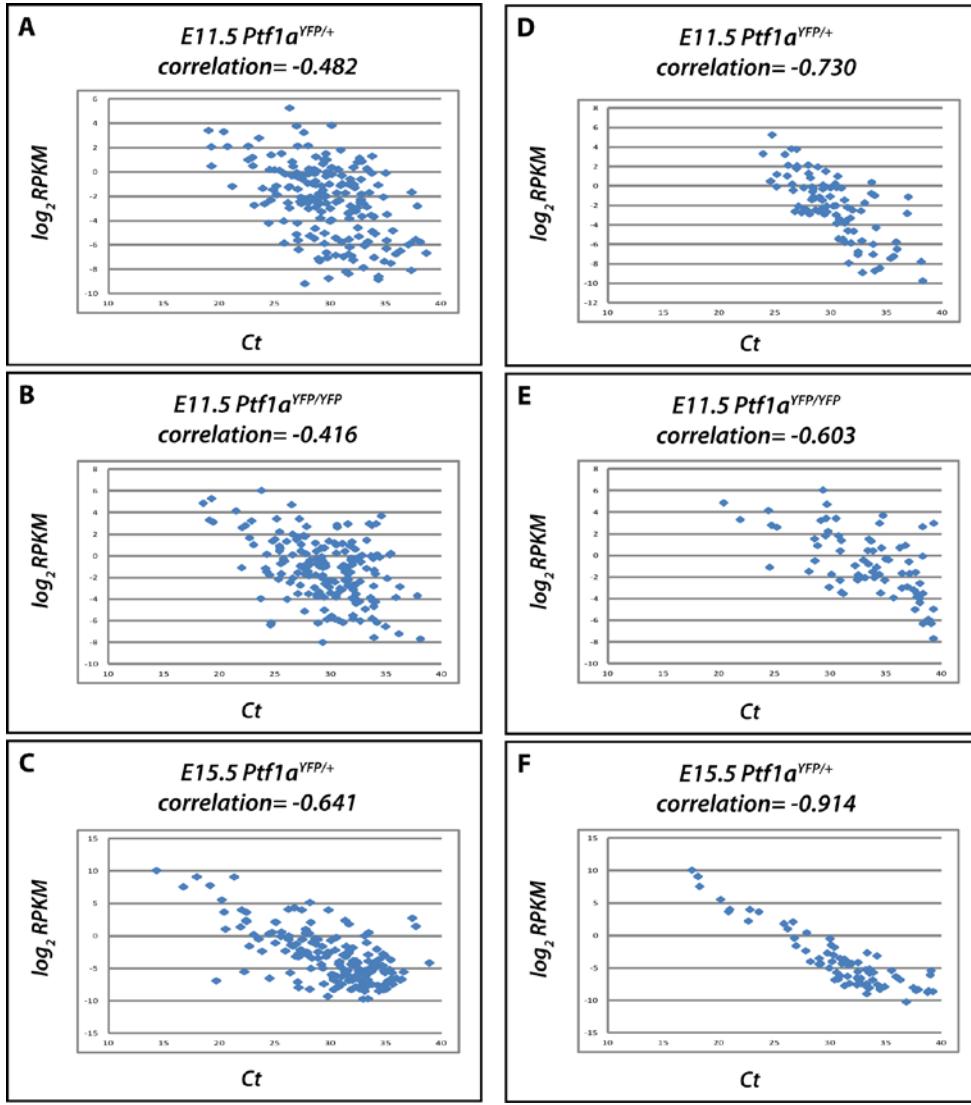
### Validation of specific genes and temporal profiling by qRT-PCR

Since the RNA-Seq datasets were generated from single libraries made from cells obtained after pooling at least five embryos, we sought to validate a subset of the transcripts across biological replicates by quantitative (q) RT-PCR profiling using high-throughput TaqMan arrays (**Table 3.2**). To perform this analysis, we isolated additional RNA samples

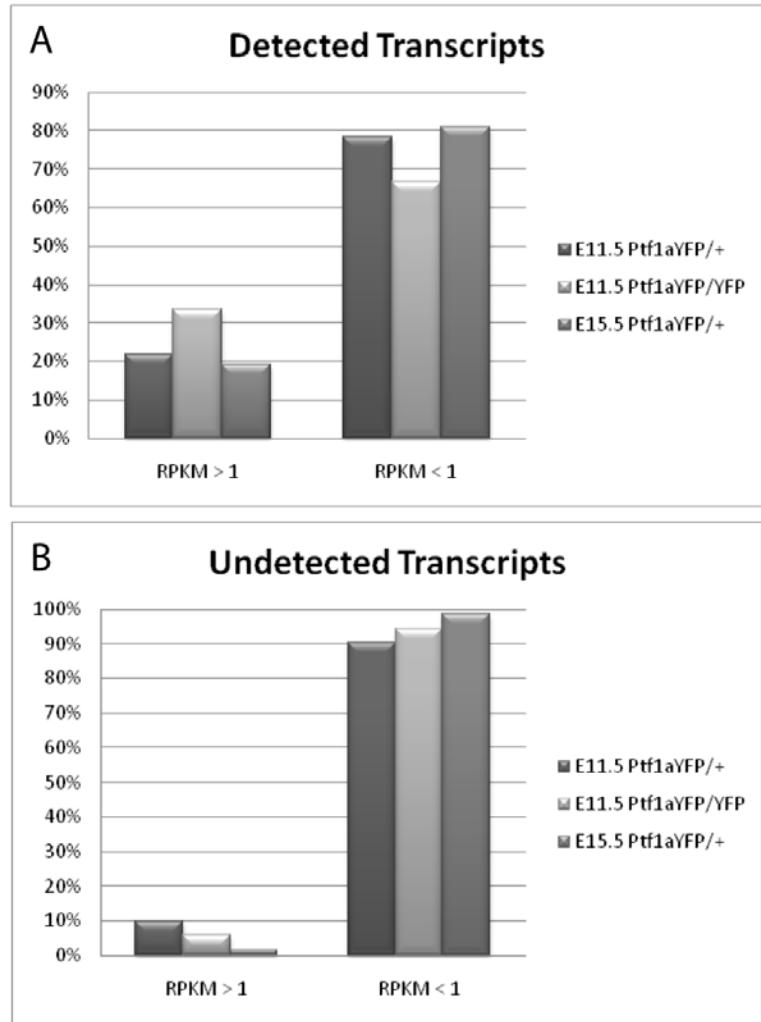
and used a two phase selection strategy to identify 94 genes that could be reliably detected by qRT-PCR under the experimental conditions of our study. In the first phase, we screened 376 transcripts identified from the RNA-Seq dataset as differentially expressed either in a temporal or genetic manner. This analysis utilized seven different amplified RNA samples from FACS-purified *Ptf1a*<sup>YFP/+</sup> cells between E10.5 and E15.5 and *Ptf1a*<sup>YFP/YFP</sup> cells at E11.5. 296 of these genes (79%) exhibited an absolute Ct value < 40 in at least one sample. A negative correlation (-0.416 to -0.641) between the absolute Ct and RPKM values was observed for detectable transcripts (**Figure 3.3A – C**). The moderate nature of the correlation may reflect the number of low-abundance transcripts (< 1 RPKM) that were analyzed. An analysis of the detected and non-detected transcripts revealed that transcripts with an RPKM value of less than one were less likely to be detected by qRT-PCR (**Figure 3.4**). Hierarchical clustering allowed the transcripts to be grouped into ten primary cluster sets (data not shown).

In the second phase, 94 genes were chosen for additional qRT-PCR analysis using at least three biological replicates for each of ten different experimental conditions. This included analyzing *Ptf1a*<sup>YFP/+</sup> cells over a span of nine days (E10.5 – E18.5) and *Ptf1a*-deficient cells (*Ptf1a*<sup>YFP/YFP</sup>) at E11.5. A negative correlation (-0.603 to -0.914) was again observed between the absolute Ct and RPKM values (**Figure 3.3D – F**). Hierarchical clustering resulted in the transcripts being grouped into five principal clusters sets based on their abundance in 1) *Ptf1a*-deficient progenitors (*Ptf1a*<sup>YFP/YFP</sup> at E11.5), 2) *Ptf1a*-deficient progenitors and early MPCs (*Ptf1a*<sup>YFP/+</sup> at E10.5), 3) early MPCs (*Ptf1a*<sup>YFP/+</sup> at E10.5 – E11.5), 4) early and late MPCs (*Ptf1a*<sup>YFP/+</sup> at E11.5 – E13.5), and 5) acinar-specified cells (*Ptf1a*<sup>YFP/+</sup> at E15.5 – E18.5) (**Figure 3.5**).

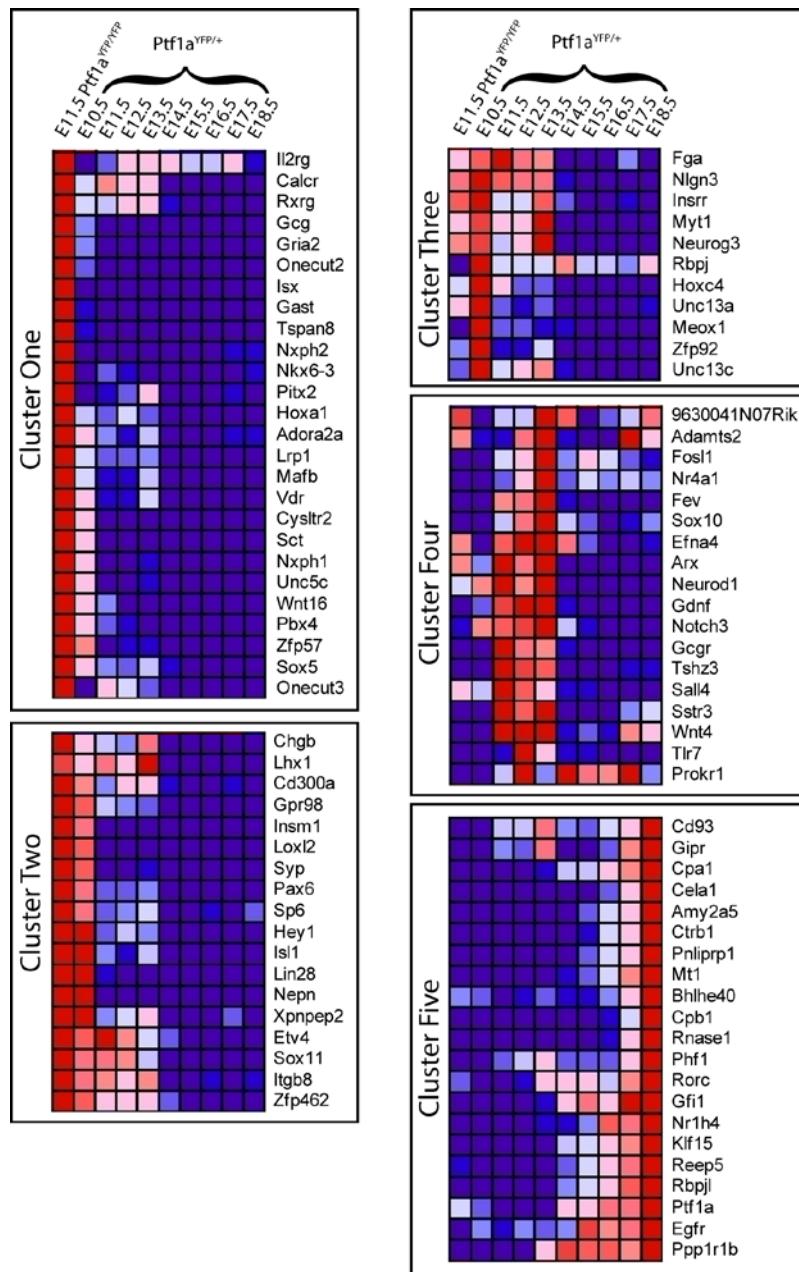
Gene cluster one contains 26 transcripts that are most highly expressed in the *Ptf1a*-deficient progenitors (E11.5 *Ptf1a*<sup>YFP/YFP</sup>). In this cluster, pre-endocrine, enteroendocrine and



**Figure 3.3 Correlation of gene expression between RNA-Seq and qRT-PCR.** The comparison of RPKM values and cycle threshold (Ct) values determined by RNA-Seq and qRT-PCR, respectively, was analyzed. Ct values are in logarithmic scale, thus RPKM values are graphed on a  $\log_2$  scale. Correlations are plotted for  $E11.5 Ptf1a^{YFP/+}$ ,  $E11.5 Ptf1a^{YFP/YFP}$ , and  $E15.5 Ptf1a^{YFP/+}$  samples following both the first phase (A – C) and the second phase (D – F) of qRT-PCR profiling.



**Figure 3.4 Analysis of detectable transcripts by qRT-PCR relative to RPKM value.**  
Transcripts profiled in the first phase of qRT-PCR were analyzed based on RPKM value and cycle threshold (Ct) value. Transcripts receiving a Ct value of less than 40 (A) represented transcripts with an RPKM > 1 (average 24.8%) and RPKM < 1 (average 75.2%). Transcripts that were undetected (Ct = 40) by qRT-PCR (B) primarily displayed an RPKM < 1 (average 94.4%).

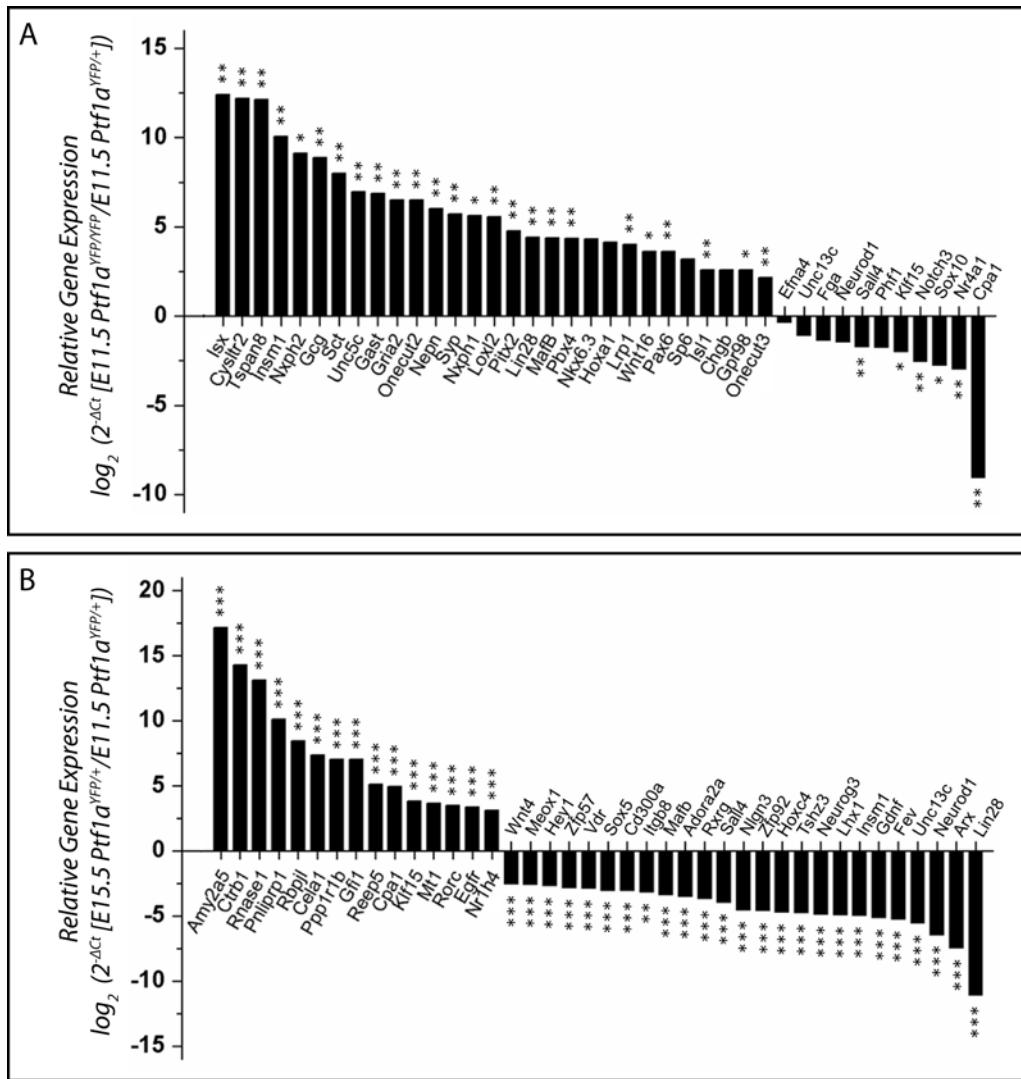


**Figure 3.5 Quantitative RT-PCR profiling of differentially expressed transcripts.** 94 differentially expressed transcripts were further profiled by qRT-PCR using ten different FACS-isolated cell samples: *Ptfla*<sup>YFP/YFP</sup> cells at E11.5 and *Ptfla*<sup>YFP/+</sup> cells from E10.5 – E18.5. Mean pre-processed expression values across biological replicates are scaled as a function of the number of standard deviations relative to the mean of each row, where the darkest blue corresponds to -3 (less abundant) and the brightest red to +3 (more abundant). Hierarchical clustering permitted the identification of five main cluster sets based on abundance in: 1) *Ptfla*-deficient progenitors (*Ptfla*<sup>YFP/YFP</sup> at E11.5), 2) *Ptfla*-deficient progenitors and early MPCs (*Ptfla*<sup>YFP/+</sup> at E10.5), 3) early MPCs (*Ptfla*<sup>YFP/+</sup> at E10.5 – E11.5), 4) early and late MPCs (*Ptfla*<sup>YFP/+</sup> at E11.5 – E13.5) or 5) acinar-specified cells (*Ptfla*<sup>YFP/+</sup> at E15.5 – E18.5).

gastrointestinal markers, such as *Gast*, *Gcg*, *Isx*, *Mafb*, *Nxk6-3*, *Onecut2*, *Onecut3* and *Sct*, were most abundant. In addition, *Cysltr2*, *Gria2*, *Hoxa1*, *Nxph1*, *Nxph2*, *Pbx4*, *Tspan8*, *Unc5c* and *Zfp57* follow a similar expression pattern.

Gene cluster two contains 18 transcripts that are most highly expressed in *Ptf1a*-deficient pancreatic MPCs (E11.5 *Ptf1a*<sup>YFP/YFP</sup>) as well as early pancreatic MPCs (E10.5 *Ptf1a*<sup>YFP/+</sup>). Pre-endocrine, enteroendocrine and gastrointestinal markers, such as *Insm1*, *Syp*, *Pax6* and *Isl1*, were abundant in this cluster. Other genes, such as *Gpr98*, *Lox12*, *Sp6*, *Hey1*, *Lin28* and *Nepn*, were primarily expressed in the *Ptf1a*-deficient progenitors and early pancreatic MPCs, while others, such as *Chgb*, *Lhx1*, *Cd300a*, *Xpnpep2*, *Etv4*, *Sox11*, *Itgb8* and *Zfp462*, exhibited continued expression in later developmental stages. Analysis of the relative gene expression of transcripts in these two cluster sets revealed a 4.5 – 5.4 x 10<sup>3</sup> fold increase as compared to the *Ptf1a*-expressing progenitor cells at E11.5 (**Figure 3.6**).

Gene cluster three contains 11 transcripts that are upregulated in the early pancreatic MPCs (*Ptf1a*<sup>YFP/+</sup> at E10.5 – E11.5). Among the genes identified are pancreatic MPC and pre-endocrine markers such as *Myt1*, *Neurog3* and *Rbpj*. In addition, this cluster also contains *Fga*, *Nlgn3*, *Hoxc4*, *Unc13a*, *Meox1*, *Zfp92* and *Unc13c*. Analysis of the differential expression at E11.5 between the *Ptf1a*-expressing and *Ptf1a*-deficient progenitors revealed the abundance of *Sox10*, *Notch3*, *Neurod1*, *Sall4*, *Fga*, *Unc13c* and *Arx* (**Figure 3.6**) in the *Ptf1a*-expressing MPCs. Similarly, *Cpa1*, *Nr4a1*, *Klf15* and *Phf1* all showed a significant reduction in the E11.5 *Ptf1a*-deficient cells compared with the *Ptf1a*-expressing progenitors. Although it has been previously proposed that *Cpa1* can be used as a marker to identify pancreatic MPCs at the distal tip of the branching epithelium (Zhou et al., 2007), both our analysis as well as a recent study (Masui et al., 2010) suggest that *Cpa1* is a fortuitous marker of the pancreatic MPCs. Since *Nr4a1*, *Klf15* and *Phf1* are all expressed in a pattern similar to *Cpa1* and display diminished expression in *Ptf1a*-deficient cells, this subcluster may have

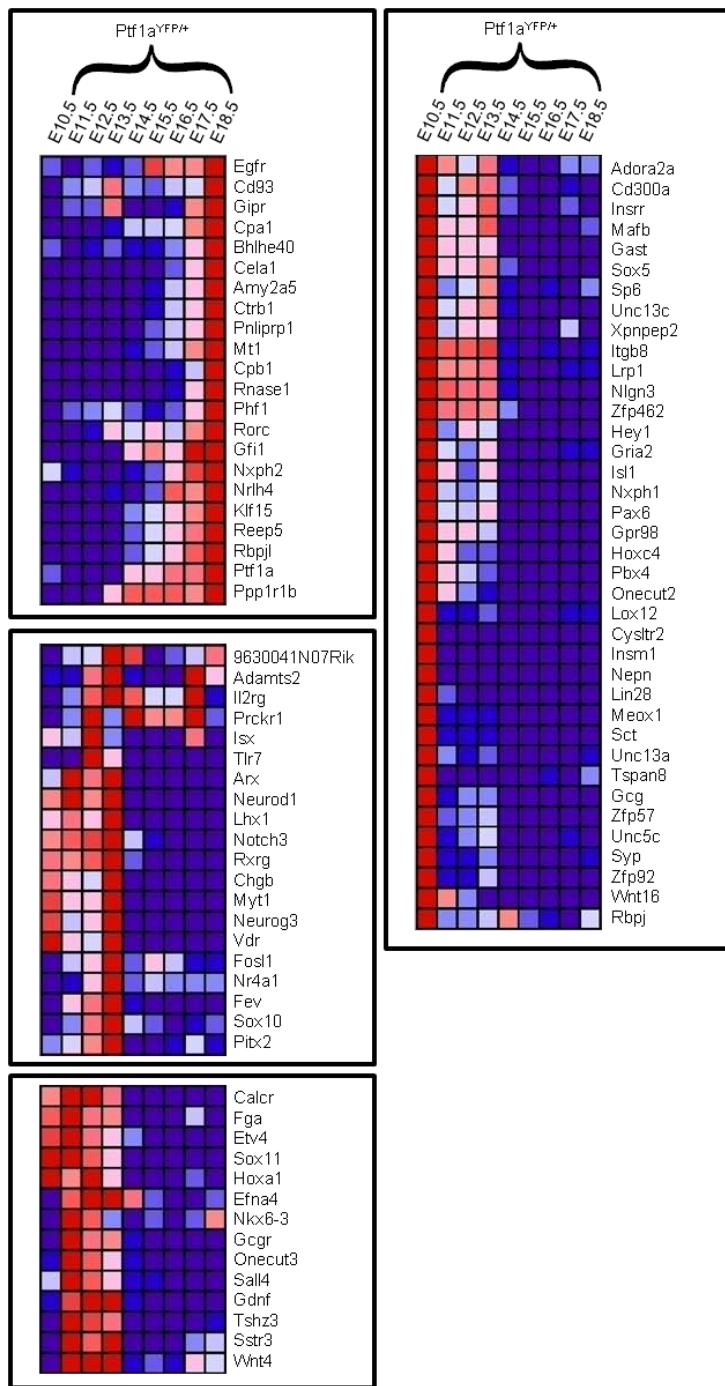


**Figure 3.6 Genetic and temporal comparisons of differentially expressed transcripts as determined by qRT-PCR profiling.** A) The expression levels of transcripts in *Ptf1a*<sup>YFP/YFP</sup> cells at E11.5 were analyzed by qRT-PCR and expressed relative to expression in *Ptf1a*<sup>YFP/+</sup> cells at E11.5. B) The expression levels of transcripts in *Ptf1a*<sup>YFP/+</sup> cells at E15.5 were analyzed by qRT-PCR and expressed relative to expression in *Ptf1a*<sup>YFP/+</sup> cells at E11.5. Relative expression is expressed in  $\log_2$  scale. Relative expression of 5 = ~35-fold change and 10 = ~1000-fold change. \* $p < 4.8 \times 10^{-2}$ , \*\* $p < 1.6 \times 10^{-2}$ , \*\*\* $p < 8.0 \times 10^{-3}$ .

unappreciated roles in progenitor cell development and lineage specification. Additionally, the comparison of the progenitors at E11.5 to the acinar-specified cells at E15.5 revealed the upregulation of a number of transcripts including *Meox1*, *Hey1*, *Sall4*, *Nlgn3*, *Tshz3*, *Unc13c* and *Lin28* (**Figure 3.6**). As expected, pancreatic MPC and pre-endocrine markers such as *Mafb*, *Neurog3*, *Neurod1* and *Arx*, were more abundant in the E11.5 cells as compared to the acinar specified cell. An alternative display of the qRT-PCR profiling, excluding the E11.5 *Ptf1a<sup>YFP/YFP</sup>* sample, allows for the temporal assessment of gene expression among only the *Ptf1a*-expressing samples (**Figure 3.7**).

Gene cluster four consists of 18 genes that are most abundant as *Ptf1a*-expressing cells are becoming restricted to the developing epithelial tips (E12.5 – E13.5) (**Figure 3.5**). Interestingly, pre-endocrine markers such as *Neurod1* and *Arx*, as well as *Neurog3* and *Myt1*, which are grouped in cluster three, were also expressed in these late progenitors. The abundant expression of these transcripts during this developmental stage can be attributed to the perdurance of the YFP which was previously described to occur around E12.5 (Burlison et al., 2008). Several transcripts, including *9630041N07Rik*, *Efna4*, *Sall4*, *Fev*, *Sox10*, *Wnt4*, *Gdnf* and *Tshz3*, were found to have a similar expression pattern to these pre-endocrine markers.

Gene cluster five contains 21 genes that are most highly expressed in pre-acinar and acinar-committed cells. As expected, this cluster contained many digestive enzymes including *Amy2a5*, *Ctrb1*, *Rnase1*, *Pnliprp1*, *Cela1*, *Cpa1*, and *Cpb1*. Additionally, we observed other genes that exhibited a similar expression pattern including *Egfr*, *Gfi1*, *Klf15*, *Mt1*, *Nrlh4*, *Ppp1rlb*, *Rbpjl*, *Reep5*, *Phf1* and *Rorc*. The relative gene expression of these transcripts between E15.5 and E11.5 *Ptf1a<sup>YFP/+</sup>* revealed varying fold changes of differential expression, from an 8.9 – 1.5 x10<sup>5</sup> fold increase (**Figure 3.6**).



**Figure 3.7 Temporal analysis of qRT-PCR profiling of differentially expressed transcripts.** This analysis is similar to the heat map and hierarchical clustering in Figure 3.5; however, the *Ptf1a*-deficient progenitor population (E11.5 *Ptf1a*<sup>YFP/YFP</sup>) is removed, thus displaying the expression pattern for only *Ptf1a*<sup>YFP/+</sup> cells from E10.5 – E18.5. Expression values depicted in the heat map represent the mean  $2^{-\Delta Ct}$  and are scaled as a function of the number of standard deviations relative to the mean of each row, where dark blue corresponds to -3 (less abundant) and red to +3 (more abundant).

## Discussion

Here, we utilized mice that express YFP under control of the *Ptf1a* gene locus to purify and characterize *Ptf1a*-expressing pancreatic cells 1) as they undergo a temporal transition from early MPCs to nascent acinar cells and 2) as cells lacking *Ptf1a* undergo a change in cell fate. In both cases, these studies utilized cell populations that were at least 99% pure. Moreover, we observed high correlations between the RPKM values obtained by RNA-Seq and the absolute Ct values obtained by qRT-PCR (**Figure 3.3**).

Due to the small number of cells isolated during early pancreas development, especially in the *Ptf1a*-null genotype, use of an RNA amplification technology was unavoidable. Although prior studies have shown the fidelity of mRNA expression is highly preserved after DNA amplification (Feldman et al., 2002), this procedure has been reported to affect certain GC-rich transcripts (McDowell et al., 1998). For this reason, we selected 94 genes using a dual phase screening strategy and analyzed these genes using multiple biological replicates. While the number of transcripts profiled in this manner represents only a small fraction of the 2,793 and 2,102 genes detected in the temporal or genetic RNA-Seq screen, it allowed for the identification of five different gene clusters, each of which provides insights into the dynamics of gene expression during specific aspects of pancreas development.

### Gastrointestinal specification of pre-pancreatic endoderm occurs in the absence of *Ptf1a*

Gene clusters one and two, which contain genes that are most highly expressed in *Ptf1a*-deficient cells, provide several insights into mechanisms of pancreatic cell specification. Prior to E9.5, posterior foregut endoderm becomes programmed in a manner to specify pancreas development. Although these events remain poorly understood, they lead to the expression of *Ptf1a* in both the dorsal and ventral evaginations of the pre-pancreatic

foregut endoderm (Krapp et al., 1998). At this developmental stage, lineage tracing experiments have shown that the pancreatic progenitor cells are multipotent, e.g. they go on to form endocrine, acinar and ductal cells that characterize the adult pancreas (Burlison et al., 2008; Kawaguchi et al., 2002). Both the expression of *Ptf1a* and formation of the PTF1 complex are clearly essential for the formation of a mature pancreas. However, exactly which genes are affected by PTF1 and how they are regulated has not been globally defined.

The identification of *Isx*, an intestine-specific transcription factor that parallels the expression pattern of *Cdx2* (Choi et al., 2006), *Tspan8*, which is expressed at high levels in the stomach and small intestine (Champy et al., 2010), and enteroendocrine markers such as *gastrin*, *secretin* and *synaptophysin* in the *Ptf1a*-null cells indicates that pancreatic MPCs, in the absence of *Ptf1a*, are being quickly redirected towards gastrointestinal fates. Indeed, given that the ventral pancreas is closely associated with the duodenum and previous studies showed that the ventral pancreas and common bile duct share a common origin (Sumazaki et al., 2004), cells of the ventral pancreas may be better positioned to adopt gastrointestinal fates. However, whether the dorsal and ventral buds are re-specified to similar or different fates, as well as the mechanism(s) by which this occurs, remains to be determined.

### **Developmental arrest of pre-pancreatic endoderm occurs in the absence of *Ptf1a***

The increase in pre-endocrine and endocrine specific genes in *Ptf1a*-deficient MPCs suggests that formation of early endocrine cells in the developing pancreas can occur in a *Ptf1a*-independent manner. This notion is supported by two prior observations. First, lineage tracing experiments have shown that approximately 50% of endocrine cells of the mature pancreas are derived from progenitors that do not express *Ptf1a* (Kawaguchi et al., 2002). Second, expression profiling of single cells from the dorsal pancreatic epithelium at E10.5 has suggested the existence of six different progenitor cell types based on combinations of

*Pdx1*, *Nkx2-2*, *Nkx6-1*, *Ptf1a*, *Neurog3*, *Pax6*, *Pax4*, *Isl1*, *Neurod1*, *Gcg*, *Ins*, *Sst*, and *Ppy* (Chiang and Melton, 2003). Of the six types identified, only two express *Ptf1a*. Thus, we hypothesize that the *Ptf1a*-deficient progenitors can transition easily into other progenitor cell states, thereby explaining the increase in the expression of *Neurog3*, *Pax6*, *Pax4*, *Isl1*, *Neurod1*, *Gcg* and *Ppy* in the E11.5 *Ptf1a*<sup>YFP/YFP</sup> cells. While the notion of a metastable state of pancreatic MPCs is speculative, previous studies have examined dynamic states of cellular gene expression and how transient gene expression patterns govern cell fate decisions (Orkin and Zon, 2002).

At the same time, the nature of the genes expressed in cluster two suggests that Ptf1a acts to prevent developmental arrest of pancreas-specified progenitor cells. The RNA-Seq profile of the *Ptf1a*-deficient cells argues against induction of apoptosis, since the majority of pro-apoptotic genes show little or no change. However, the profiles suggest that proliferation is rapidly attenuated, as is evident by the decrease in the expression of many markers of cellular proliferation (*mKi67*, *Foxm1*, *Cenpf*, *Ccne1*, *Top2a*, *Ccnb1*, *Cdk1*, *Mcm4*, *Ccnd1* and *Mcm3*). This is consistent with the Ptf1a-containing PTF1 complex being broadly important for the proliferation of pancreatic MPCs. Another line of reasoning to support the notion of developmental arrest is the presence of *Nephrocan* (*Nepn*) (Hou et al., 2007), a definitive endoderm marker, *Lin28*, an important pluripotency factor (Zheng et al., 2009), and *Hey1*, a downstream target of Notch signaling (Ghosh and Leach, 2006), in both the E10.5 *Ptf1a*<sup>YFP/+</sup> and E11.5 *Ptf1a*-deficient cells. Indeed, some of the genes expressed in early *Ptf1a*-expressing MPCs and *Ptf1a*-deficient cells may reflect the ancestry of the cells from a preceding stage, e.g. E9.5 or earlier, prior to or at the earliest onset of *Ptf1a* expression. However, other possibilities are that the genes in cluster two are regulated in a Ptf1a-independent manner or that these genes may be repressed by PTF1. In either case, these

results further confirm that the function of *Ptf1a* during early stages of pancreas development is to commit pre-pancreatic foregut endoderm towards pancreatic lineages.

### ***Ptf1a<sup>YFP</sup>* perdurance allows for the transient isolation of MPC progeny**

Clusters three and four provide insights into events that are occurring as the cells first become specified to the pancreatic lineage and then begin to undergo branching morphogenesis, a process that is also seen during lung, kidney and mammary development (Affolter et al., 2009). At this stage, pancreatic progenitor cells have developed from the foregut, and the pre-pancreatic epithelium undergoes elongation and branching, where the expression of *Ptf1a* becomes restricted to the peripheral tips of the epithelial tree. Recently, it has been shown that pancreatic MPCs become restricted to the developing periphery of the epithelial tree, and bipotential (duct/endocrine) progeny that arise from the MPCs are located in the central region of the epithelial tree (Zhou et al., 2007). This central region which has been termed the trunk domain (Zhou et al., 2007) does not express *Ptf1a*. However, due to the perdurance of YFP (Burlison et al., 2008), some *Ptf1a<sup>YFP/+</sup>* cells from E12.5 to E13.5 may represent cells localized to the central region of the epithelial tree. The YFP perdurance, while generally undesirable for a study such as this, provides further suggestive evidence that pre-endocrine progenitors residing in the trunk epithelium are, in fact, derived from *Ptf1a*-positive ancestors.

Consistent with the idea that the perdurance of YFP results in isolation of endocrine progenitors, clusters three and four revealed increased expression of pre-endocrine markers, such as *Neurog3*, *Myt1*, *Arx* and *Neurod1*. While the hierarchical clustering revealed that some of these transcripts are most abundant by E12.5 or E13.5, an analysis of relative gene expression revealed that many of these transcripts are more abundant in the E11.5 *Ptf1a*-expressing MPCs compared to the E15.5 acinar-specified cells (**Figure 3.6**). Thus, their

expression might begin in early pancreatic MPCs, continue as the progenitors become restricted to the developing peripheral tips, and further increase in the pre-endocrine precursors which are located in the bipotential trunk domain.

### **Ptf1a promotes differentiation to acinar cell fates**

Finally, cluster five consists of transcripts that reflect nascent acinar cells. Consistent with this is the expression of many proteases, hydrolases, phosphatases and other enzymes, as well as the increasing expression of *Rbpjl*, which is consistent with previous studies (Masui et al., 2010). The expression profile of *Rbpjl* can be contrasted with *Rbpj*, which is most abundant at E10.5, where it is part of the early PTF1-J complex. While the majority of the transcripts in cluster five are expressed at high levels in acinar-specified cells, as displayed in the heat maps, some are also expressed at earlier stages of pancreas development (**Figure 3.6**). For example, by expression analysis *Ptf1a* and *Cpa1* are most abundant in the E18.5 acinar-specified cells, yet both are expressed in the early MPC population, as previously shown by *in situ* and immunohistochemical analyses (Burlison et al., 2008; Kawaguchi et al., 2002; Krapp et al., 1998; Zhou et al., 2007). Interestingly, the expression profiles of *Klf15* and *Gfi1* closely parallel those of *Ptf1a* and *Cpa1*, suggesting that the further analysis of some transcripts in cluster five may reveal genes that similarly identify the MPC population and promote its specification towards the acinar lineage.

## **Conclusions**

Further studies of the transcription factors, cell surface receptors and pathway-specific genes identified from these studies will likely be useful for further characterizing the biology of pancreatic MPCs *in vivo* and for more precisely analyzing events within these cells as they first undergo a transition to become bipotential trunk cells and then terminally

differentiate to become acinar cells. While the data we obtained points to a very high degree of complexity in the genetic program of pancreatic MPCs, our datasets promise to provide an important platform for further discovery of mechanisms, pathways and factors that control pancreas development, especially when compared with other pancreatic lineages, particularly the endocrine cell lineage.

## **CHAPTER IV**

### **CHARACTERIZATION OF NEPHROCAN DURING PANCREAS DEVELOPMENT**

#### **Introduction**

##### **Nephrocan, a unique member of the SLRP family**

Nephrocan (Nepn) was first identified through a bioinformatics query examining proteins similar to the small leucine-rich protein (SLRP), Decorin (Mochida et al., 2006). SLRPs are a conserved family of proteins which contain multiple repeats of a leucine-rich motif (LRR) flanked by cysteine residues and can be synthesized as either glycoproteins, which contain N-linked oligosaccharides, or as proteoglycans, which contain chondroitin/dermatan or keratan sulfate chains (Hocking et al., 1998; Iozzo, 1997, 1998). These post-translational modifications of SLRPs are thought to modify the function of SLRPs, which include acting as modulators of cellular pathways and playing key roles in a number of biological processes, including collagen fibrillogenesis, blood vessel viscoelasticity, cell migration, and cell proliferation (Iozzo, 1997, 1999). Currently, SLRPs are classified into five classes based on several parameters, including their evolutionary protein conservation, the presence of a distinct cysteine-rich cluster at the N-terminus, the number of leucine-rich repeats, and their chromosomal organization (Schaefer and Iozzo, 2008).

Specifically, Decorin is grouped as a class I SLRP and is characterized by ten LRR motifs, N- and C-terminal cysteine clusters, N-linked glycosylation, and a chondroitin/dermatan sulfate side chain (Krusius and Ruoslahti, 1986; Schaefer and Iozzo, 2008). Decorin is a small cellular matrix proteoglycan which interacts with the transforming

growth factor  $\beta$  (TGF $\beta$ ) pathway, affects cell proliferation, and modifies the extracellular environment (Hildebrand et al., 1994; Schonherr et al., 1998; Takeuchi et al., 1994).

Similarly, Nepn consists of multiple LRR motifs flanked by cysteine-rich clusters, as well as putative N-glycosylation sites; however, there are several features that suggest Nepn belongs to a new class of SLRPs. These distinct features include the presence of 17 LRR repeats, four cysteine residues in the C-terminus, and a unique polyacidic tail at the C-terminus, as well as the absence of a conserved propeptide domain that may function as a recognition sequence for xylosyltransferase (Mochida et al., 2006).

### **Nephrocan, an inhibitor of TGF $\beta$ signaling**

Similar to Decorin, Nepn interacts with TGF $\beta$  signaling and specifically acts as an inhibitor of TGF $\beta$  activity (Mochida et al., 2006). The TGF $\beta$  family members include not only TGF $\beta$ , but also activins and bone morphogenetic proteins (BMPs) which control a variety of cellular processes, including proliferation, apoptosis, cellular differentiation, and epithelial-mesenchymal transition (EMT), a mechanism essential for numerous developmental processes and characterized by loss of cell adhesion, polarity and cell-cell contacts, as well as increased cell mobility (Massague et al., 2000). While there are various routes through which the TGF $\beta$  family members can act, the Smad pathway is the canonical pathway through which TGF $\beta$  family members signal through (Deryck and Zhang, 1996). In this signaling pathway, TGF $\beta$  binds to a type II receptor dimer, subsequently recruiting and phosphorylating a type I receptor. A receptor-regulated Smad, such as Smad3, is phosphorylated by the type I receptor and binds to a co-mediator Smad protein. This heterodimeric Smad complex translocates into the cell nucleus where it can serve as a modulator of gene transcription. Studies examining the effect of Nepn on TGF $\beta$ /Smad signaling revealed that Nepn inhibited TGF $\beta$  activity and downregulated Smad3.

phosphorylation, indicating that Nepn may function upstream of Smad3 to inhibit the TGF $\beta$ /Smad signaling pathway (Mochida et al., 2006).

### **Role of TGF $\beta$ signaling in endoderm and pancreas development**

The TGF $\beta$  pathway is a major regulator of pancreatic endoderm, progenitor, and endocrine development. During early stages of embryogenesis, the formation of definitive endoderm is a process dependent on the canonical TGF $\beta$ /Smad signaling pathway (Tremblay et al., 2000). The role of TGF $\beta$  signaling in development has become further apparent as recent studies have shown that TGF $\beta$  signaling molecules, such as activin, can induce human and mouse ES cells to differentiate into definitive endoderm-like cells (D'Amour et al., 2005; Kubo et al., 2004; Yasunaga et al., 2005). While TGF $\beta$  signaling promotes early endoderm development, it has been shown that the TGF $\beta$  pathway inhibits specification of early pancreatic progenitors. Utilizing embryonic explants, studies have revealed that continuous TGF $\beta$  signaling in the foregut endoderm limits the number of ventral pancreatic progenitors that become specified (Wandzioch and Zaret, 2009). As cell specification proceeds, TGF $\beta$  plays a critical role in regulating endocrine and exocrine cell fates in the later stages of pancreas development. Exposure of *in vitro*-cultured pancreatic rudiments to TGF $\beta$ 1 inhibited the development of acinar tissue and promoted the development of endocrine cells (Sanvito et al., 1994). Conversely, exposure of *in vitro*-cultured pancreatic rudiments to follistatin, an antagonist of TGF $\beta$  signaling, promoted acinar development and inhibited endocrine differentiation (Miralles et al., 1998).

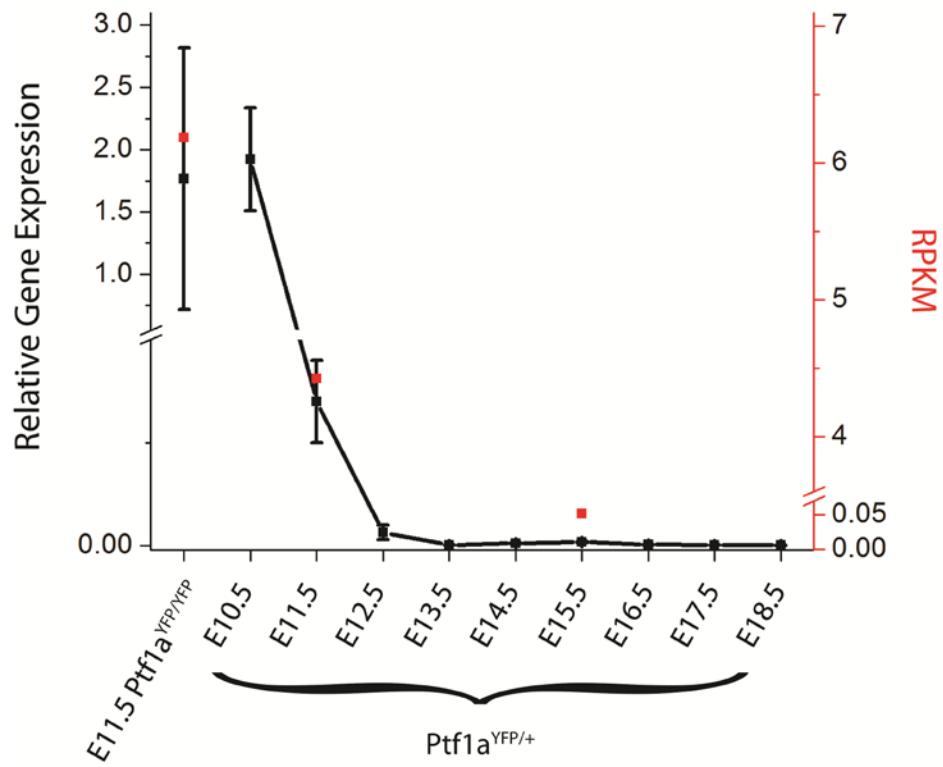
Collectively, these studies provide evidence for both the stimulatory and inhibitory affects of the TGF $\beta$  signaling pathway on pancreas development during three distinct temporal stages. First, during early embryogenesis, signaling by TGF $\beta$  serves to promote endoderm development and specification. Subsequently, pancreatic progenitor cell

specification is promoted only when TGF $\beta$  signaling is effectively inhibited or the cells spatially move outside the TGF $\beta$  signaling domain. Lastly, following specification of pancreatic progenitors, TGF $\beta$  signaling serves to regulate the balance between endocrine and exocrine cell fates.

### Expression of *Nephrocan* during mouse development

*Nepn* was first identified to be predominantly expressed in kidney of adult mice and was such named given that the Greek word νεφρός, ‘nephros,’ means kidney (Mochida et al., 2006). In the same study, *Nepn* was also detected to a lesser extent in heart, lung and skeletal muscle, with virtually no expression in brain, spleen, liver, or testis. Immunolabeling revealed the expression of Nepn extensively in the epithelial cells of distal tubules and collecting ducts in the kidney of adult mice and diminished labeling in the proximal epithelial cells. Further expression profiling at various embryonic stages revealed *Nepn* to be upregulated at E11.0, as compared to E7.0, E15.0 and E17.0, marking a critical time for cell specification and organogenesis. Similarly, we observed *Nepn* to be upregulated at E10.5 in *Ptf1a<sup>YFP</sup>*-expressing pancreatic progenitors (**Figure 3.5, 3.7, and 4.1**). Additionally, *Nepn* is expressed in the pylorus, the region of the caudal stomach that connects to the duodenum (Li et al., 2009). Microarray analysis of the transcriptome of stomach, pyloric and intestinal tissues from E14.5 to E16.5 revealed the enrichment of *Nepn* in the pylorus, and *in situ* analysis confirmed expression in the pyloric epithelium with more robust expression near the antral stomach. Given that Nepn is a secreted modulator of TGF $\beta$  signaling, this study was the first to highlight a secreted signaling protein in the pyloric epithelium.

In addition to expression in the kidney and pylorus, *Nepn*, which is also known as the RIKEN gene 5730521E12Rik, is expressed in the definitive endoderm during early stages (0



**Figure 4.1 Gene expression profile of *Nepn* by qRT-PCR and RNA-Seq.** The normalized expression level of *Nepn* is shown in *Ptf1a*<sup>YFP/YFP</sup> cells at E11.5 and *Ptf1a*<sup>YFP/+</sup> cells from E10.5 to E18.5 by qRT-PCR (black) and in *Ptf1a*<sup>YFP/YFP</sup> cells at E11.5 and *Ptf1a*<sup>YFP/+</sup> cells at E11.5 and E15.5 by RNA-Seq (red). *Nepn* is expressed in the *Ptf1a*<sup>YFP/YFP</sup> cells at E11.5 and *Ptf1a*<sup>YFP/+</sup> cells at E10.5; however, *Nepn* expression is diminished in *Ptf1a*-expressing cells at E11.5 and barely detectable thereafter.

– 6 somites) of endoderm patterning and in the posterior midgut and hindgut during later stages (8 – 12 somites) of endoderm patterning (Hou et al., 2007). Using a serial analysis of gene expression (SAGE) approach, *Nepn* displayed a 7-fold increase in expression in the posterior midgut and hindgut as compared to the foregut and anterior midgut. Further characterization by *in situ* analysis showed *Nepn* expression as early as E7.25 in the posterior region of the definitive endoderm, and its expression expanded throughout the posterior region at E7.75. As development continues, *Nepn* expression extended anteriorly and was expressed bilaterally in the midgut region. Given the lack of definitive endoderm specific transcripts, the identification of *Nepn* as an early definitive endoderm and midgut marker will prove to be valuable for examining the molecular mechanisms regulating endoderm development.

## Materials and Methods

### Generation of a ROSA26<sup>LCA</sup> targeting vector

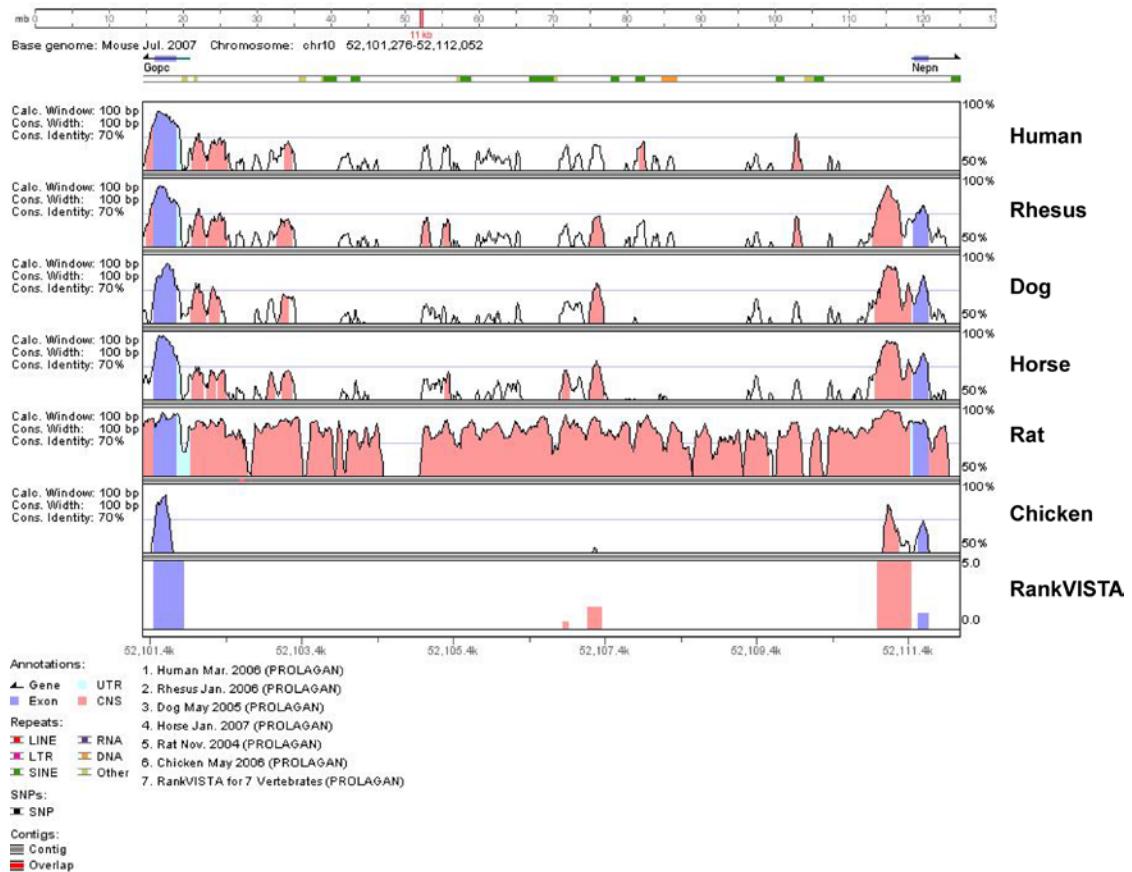
The ROSA26<sup>LCA</sup> targeting vector was made by a two-step BAC recombineering procedure (Copeland et al., 2001). First, two inner homology regions from the *ROSA26* locus were PCR-amplified from a mouse *ROSA26* BAC (clone number: 58-D17, RPCI-22 library) and cloned into a plasmid (termed pLCA.71/2272) containing a dual selection cassette consisting of a fusion of *puromycin resistance* and a mutant *thymidine kinase* driven by the mouse *phosphoglycerol kinase* promoter (*pgk-puΔtk*) and an EM7-driven *neomycin resistance* (EM7-*Neo*<sup>R</sup>) flanked by lox71 and lox2272 sites. The resulting construct was used to insert the lox71, *pgk-puΔtk*, EM7-*Neo*<sup>R</sup>, and lox2272-containing cassette into the *ROSA26* BAC. Second, two outer homology regions from the *ROSA26* locus were cloned into a *diphtheria toxin A* containing plasmid (termed pMCS.DT-A), and the resulting plasmid was

used to retrieve the lox71, *pgk-puΔtk*, EM7-*Neo*<sup>R</sup>, and lox2272 cassette along with flanking upstream and downstream homology arms of 8.232 kb and 3.770 kb, respectively.

The targeting vector generated replaced a 5.165 kb region of the *ROSA26* gene, containing both the regulatory sequences and first exon, with the dual *pgk-puΔtk* and EM7-*Neo*<sup>R</sup> selection cassette. 200 µg of the *ROSA26*<sup>LCA</sup> targeting vector was linearized with NotI for electroporation. Targeting vector construction and preparation were performed by Kathy Shelton.

### **Generation of a *ROSA26*<sup>Nepn-Cherry</sup> exchange cassette**

The Nepn-Cherry exchange vector was made by standard BAC recombineering methods (Copeland et al., 2001; Lee et al., 2001). First, two inner homology regions from the *Nepn* gene locus were PCR-amplified from a BAC containing mouse *Nepn* (clone number: 167-J16, RP23 library) and cloned into a plasmid containing monomeric Cherry (pmCherry-C1; Clontech), a rabbit *β-globin* polyadenylation sequence containing intronic sequences to allow splicing (phspPolyA in pBluKS(-)), and a dual *pgk*/EM7-driven *neomycin resistance* sequence flanked by FRT sites (PL451 plasmid). The resulting construct was used to insert the Cherry-containing cassette into the *Nepn* BAC. Second, two outer homology regions were cloned into a lox66/2272-containing plasmid (termed pLS66/2272), permitting the retrieval of the Cherry construct flanked by 9.368 kb of 5' *Nepn* sequence and 833 bp of 3' *Nepn* sequence. The inclusion of 9.365 kb of *Nepn* 5' regulatory sequences was determined by an analysis of homology of the *Nepn* locus among species (**Figure 4.2**). An alternative exchange cassette containing 12.9 kb of *Nepn* 3' sequence was also constructed. The exchange cassettes did not reincorporate the 5.165 kb region of the *ROSA26* gene which contains both the regulatory sequences and first exon. Exchange vector design and construction were performed by Weiping Yuan.



**Figure 4.2 VISTA plot displaying homology of *Nepn* locus among species.** The conservation of the *Nepn* locus is shown for human, rhesus, dog, horse, rat and chicken relative to the mouse genome. The histograms depict sequence lengths of 100 base pairs (bp) with  $\geq 70\%$  sequence conservation indicated by pink shading, and exon conservation is indicated by purple shading. The last panel reveals the RankVISTA plot which quantitatively predicts conserved regions across the seven species. A 9.4 kb region upstream of *Nepn* exon 1 (small purple box) was included in the *Nepn*-Cherry exchange cassette to incorporate potential upstream regulatory elements.

### **Generation of a *Nepn*<sup>Cherry</sup> targeting vector**

The *Nepn*<sup>Cherry</sup> targeting vector was made by cloning two outer homology regions from the mouse *Nepn* BAC into a *diphtheria toxin A* containing plasmid (termed pMCS.DT-A). The resulting plasmid was used to retrieve a portion of the mouse *Nepn* BAC containing monomeric Cherry, the rabbit  $\beta$ -globin polyadenylation sequence, and the dual *pgk*/EM7-driven *neomycin resistance* flanked by FRT sites. The Cherry cassette was flanked upstream and downstream by homology arms of 8.392 kb and 3.264 kb, respectively. The targeting vector was linearized with *Cla*I for electroporation. As summarized in **Table 4.1**, the screening strategy utilizes digestion with EcoRI and hybridization with a 5' probe (wild type band: 14.36 kb, targeted band: 12.11 kb) and digestion with SphI and hybridization with a 3' probe (wild type band: 16.7 kb, targeted band: 9.13 kb)

### **Mouse ES cell culture and electroporation**

TL1 mouse ES cells (Labosky et al., 1994) were cultured in 0.1% gelatin-coated tissue culture dishes on a layer of irradiated mouse embryonic fibroblasts (MEFs) using complete ES cell media which consisted of DMEM (Gibco, Grand Island, NY) plus 15% FBS (Atlas Biologicals, Fort Collins, CO), 2 mM L-glutamine (Invitrogen, La Jolla, CA), 0.1 mM non-essential amino acids (Invitrogen), 1,000 units/ml leukemia inhibitory factor (Chemicon, Temecula, CA), and 0.05 mg/ml Gentamicin (Invitrogen). MEFs were harvested from E13.5 antibiotic resistant (DR4) C57BL/6J embryos, expanded *in vitro*, and were mitotically inactivated by  $\gamma$  irradiation ( $^{137}\text{Cs}$ ). ES cells were routinely passed every two to three days with 0.25% trypsin, 1mM EDTA • 4Na, pH 8.

For the generation of the *ROSA26*<sup>LCA</sup> allele, a total of four electroporations were performed in which 40  $\mu\text{g}$  of the linearized *ROSA26*<sup>LCA</sup> targeting vector was electroporated into  $5.6 \times 10^6$  TL1 ES cells (passage 14) per electroporation at 240 V and a capacitance

**Table 4.1 Southern blot hybridization probes for *Nephrocan* gene targeting**

5' Hybridization Enzyme: EcoRI Wild type allele: 14.36 kb Targeted allele: 12.11 kb	CTCACTCAATACTAGAAGTTATTGAATATTGACAGCT CTAAAGACATAGCTGCCCGTAATATCTGCCAAGTGTGA ATGTAAACTGAGAAGGTAGATAGGAACATATATATTCT CATTACATTCACCAAGAGCTTGATCGGATTCCCCCTA AGTTCATGCATGTCTGTGTACAGACACGGACGTTGA TCTGGGCTATTGGATGAAAATAATCAGAGCTGCATCTG TCTGAAAATACATCATCTAGCTTGTCAATTATCAAAGCA GAAGAGGAGAGTTGAGGTATTCCCCGACAGAACAGATCT TGGTAATAACACACCTACAAGGCAGGTAGTGTGATTAGCT CCTGTCCCCACCGCATGAGATGCTGGGAGGTGGATGTA CAGGACGTCTGGAGACCCACACATCAGTCTGCATCCTT TTCCTCCCCACCCGCTCCGACTCCTTGCTTCACTCA CGTATAAAACGCACTCTCCATCGCATTGAAAGCTGAGG TGTCCAGTGGTAGGTTAAGGATCA
3' Hybridization Enzyme: SphI Wild type allele: 16.7 kb Targeted allele: 9.13 kb	GGAAATTTCAACTTCTTAAGCCTGGCTTCCTCACTG AAGGGTCAGTAATCATTCTTGAGAAAAAAATTAAGC AGCAAATTTAGAAAATTGTGACTACTTTAATTGACA GACATCAAAACCAATCCCTGCTGACTCTGGCAAAGAG AGTAAGAAAATATTTCCATGCTTAATAGAAATATTAA GATTCTAGCTACTAGATTCTAGCTAGGACCAGGCAGAAT CTCTCTGTTATGTAGGCCACTCTCTGCACCTCATATGTA GAAATGAAAACAAATTGAAACCTGATAATGTGCTTGT CTCATCCCCAAATGAATCACGTATCTGGATCTATGCCA TCTTTAGACAAGCCTTCAGAGTTATCAGGGCTTGTG GCCATTCTAGATAAGTAAAGTTAGTGACACCCAGGT TCTTCTTTGGGATTCTGACTTCTTAACAAGTCAAG AAGTGAGGGTCGGTGTCTATTCCAACCTCTGTATAAAA GGTAA

setting of 500 µFD. After selection using 1.5 µg/ml of puromycin (Sigma) for seven days, DNA was isolated from 270 of the 384 surviving clones for analysis by Southern blot. Southern probes for hybridization on the 5' and 3' ends were PCR-amplified from the mouse *ROSA26* BAC (clone number: 58-D17, RPCI-22 library), and clones were screened on the 5' end by digestion with SphI and hybridization with a 5' probe (wild type band: 16.129 kb, targeted band: 8.116 kb) and on the 3' end by digestion with NsiI and hybridization with a 3' probe (wild type band: 19.296 kb, targeted band: 11.046 kb) (**Table 4.2**). Three correctly targeted clones were identified by Southern analysis (1H8, 2A6 and 5B9) (**Figure 4.3**).

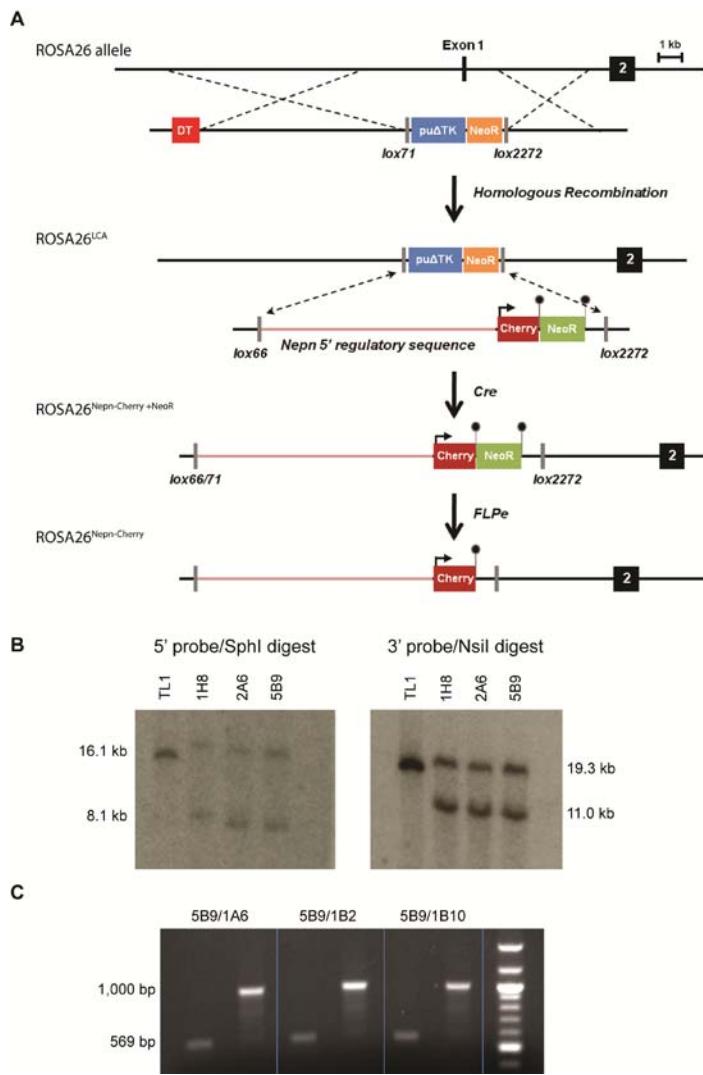
Following blastocyst injections, clones 5B9 and 2A6 both exhibited germline transmission.

Electroporation and screening were performed by Susan Hipkens and Kathy Shelton.

For the generation of the *Nepn*<sup>Cherry</sup> targeted allele, a total of four electroporations were performed in which 40 µg of the linearized *Nepn*<sup>Cherry</sup> targeting vector was used to electroporate 5.0 x 10<sup>6</sup> TL1 ES cells (passage 12) per electroporation at 240 V and a capacitance setting of 500 µFD. After selection using 200 µg/ml of G418-Geneticin (Invitrogen) for nine days, DNA was isolated from 366 surviving clones for analysis by Southern blot. Southern probes for hybridization on the 5' and 3' ends were PCR-amplified from the mouse *Nepn* BAC (clone number: 167-J16, RP23 library), and clones were screened on the 5' end by digestion with EcoRI and hybridization with a 5' probe (wild type band: 14.36 kb, targeted band: 12.11 kb) and on the 3' end by digestion with SphI and hybridization with a 3' probe (wild type band: 16.70 kb, targeted band: 9.13 kb). Thirteen correctly targeted clones were identified by Southern analysis, and eight clones were expanded for karyotyping (2A10, 2F5, 2H9, 4A5, 4B5, 4B8, 4C1, and 4C7). Clones 2H9 and 4A5 exhibited the highest percentage of normality following karyotyping, 75% and 65%, respectively, and were prepared for blastocyst injections. Electroporation and screening were performed by Susan Hipkens and Rama Gangula.

**Table 4.2 Southern blot hybridization probes for *ROSA26* gene targeting**

5' Hybridization Enzyme: SphI Wild type allele: 16.129 kb Targeted allele: 8.116 kb	TCTTCAAGTACTGAGTAAGAACTCGTCCAAGGGATGC TTTCTAACAGAACCAACTCTCAGGGTTACTGGCATGTC TACCCAGTTATAACTTAATACACAACAAGACAATTTC ACTATACTGTATCGTTAAGTGAAAGAAAGGCTACTCCT TTTACCTTTGCAATTCCCTGAATTCCAGGCCACTAGA GAAACGAATCACAATACCAAATGAAGGTATGCCATT TACTAGATGTGTTAGAATATCTGGGCTCCAAGACTTAT CTACCTCATAGGTGCATTGCTCTCTTCTAAGTCTAT CTGAAAGGACACTTAAACCAATTATAAATGACTAC CAGTTAAAAAAAATGACTACCAGTTGAACATTATA CGCATATTATACGCATATTCAAGTCTAAGTCACCACT AGTCAGGCCTCTAATTCAAATTAAACTGAAAAAAA AAAAAAAAAAACTAGTTAGCCAAAAGTTGTAATAAAA TTTATTCTACTTATGATCTAATTATTGTAGATATGTA CTCAATAATT
3' Hybridization Enzyme: NsiI Wild type allele: 19.296 kb Targeted allele: 11.046 kb	AACTGTAGATCTTAAGGGATGCTTCTGCTCTGAGATG ATACAAAGAATTAGACCATAAAACAGTAGGTTGCAC AAGCAATAGAATATGGCCTAAAGTGTCTGACACTTA GAAGCCAAGCAGTGTAGGCTCTTAAGAAATACCATT ACAATCACCTTGCTAGAAATCAAGCATTCTGGAGTGG TCAAGCAGTGTAAACCTGTACTGTAAGTTACTTTCTGC TATTTTCTCCCAAAGCAAGTTCTTATGCTGATATTTC CAGTGTAGGAACACTAAATATTAAAGTTGCTTCA CTCTTTCTTACCAAGGAGGGTCTTCCTCATCTG ATCTGAAGGATGAACAAAGGCTTGAGCAGTGCCTT AGAAGATAAACTGCAGCATGAAGGCCCCGATGTTCA CCCAGACTACATGGACCTTCGCCACACATGTCCCATT CCAGATAAGGCCTGGCACACACAAAAAACATAAGTCA TTAGGCTACCACTGATTCTAAAACAACCTAAAATCT TCCCACCTAAATGCTATGGGTGGTGGGTTGGAAAGTT GACTCAGAAAATCACTTGCTGTTTAGAGAGGATCT GGGTCAGTTCTGATACATTGTGGCTACAACATCTTCTGT CTCCAGTTCTAGGGGGTCCATCCAACATCCTCTTCTGT TGAGGGCACCAATAATGTATTGTGTACAAACAGGG AGGTGAGTGATTAACTCTCGTGTATAGTACCTTGGTA AAACATTCTTGCTGAGTAAGCAGTACAGCTGCC TGTCCCTGGTCTACAGACACGGCTCATTCCGAAGGC AAGCTGGATAGAGATTCCAATTCTCTTCTGGATCCC ATCCTATAA



**Figure 4.3 Generation of a *ROSA26*<sup>Nepn-Cherry</sup> reporter allele by RMCE.** A) Schematic representation of the *ROSA26* locus, *ROSA26*<sup>LCA</sup> allele, and the *ROSA26*<sup>Nepn-Cherry</sup> reporter allele. The *ROSA26*<sup>LCA</sup> allele contains a *pgk*-driven *puromycin resistance*- $\Delta$ *thymidine kinase* fusion gene (*pu* $\Delta$ *TK*) and an *EM7*-driven *neomycin resistance* (*Neo*<sup>R</sup>) flanked by *lox71* and *lox2272* sites (gray bars) which replace a 5.165 kb fragment containing the *ROSA26* promoter and first exon. The *lox66* and *lox2272*-flanked *Nepn*-Cherry exchange cassette consists of a 9.365 kb fragment from the *Nepn* locus, *Cherry* coding sequence and an *FRT*-flanked (black circles) dual *pgk*/*EM7*-driven *Neo*<sup>R</sup> cassette. RMCE was performed using a positive-negative selection strategy mediated by *Cre*. Mice containing the *ROSA26*<sup>Nepn-Cherry+NeoR</sup> allele were mated to *FLPe*-expressing mice resulting in the final *ROSA26*<sup>Nepn-Cherry</sup> reporter allele. B) Southern analysis of three ES cell clones following gene targeting. Clones 1H8, 2A6 and 5B9 were identified as correctly targeted. TL1 wild type DNA was used for representation of the non-targeted allele. C) PCR screening of *ROSA26*<sup>Nepn-Cherry(+NeoR)</sup> exchanged ESC clones using primer sets to detect wild type and exchanged alleles. Clones 5B9/1A6, 5B9/1B2, and 5B9/1B10 were identified as properly exchanged on both the 5' (lane 1, 569 bp) and 3' end (lane 2, 1000 bp).

### **Recombinase-mediated cassette exchange**

RMCE was performed as previously described (Long et al., 2004). In brief, a total of two electroporations were performed in which equal amounts (40 µg) of the Nepn-Cherry exchange vector and pBS185, a Cre-expression vector, were used to electroporate  $5.0 \times 10^6$  mouse ES cells containing the *ROSA26<sup>LCA</sup>* allele, clone 5B9 (passage 24) (Chen et al., 2011). After a staggered positive-negative selection strategy, clones surviving exposure to both G418-Geneticin (88 survived) and gancyclovir (14/88 survived) were screened by PCR on both the 5' and 3' ends using the following primers sets: 5' analysis: 5'-AGACTTATCTACCTCATAGGTG and 5'-GCTATTGCGCATGCACAC; and 3' analysis: 5'-GCAGAATCCAGCACCTTC and 5'-TCACAAGCAATAAACCTGTAGT. Properly exchanged clones yielded 569 bp and 1,000 bp bands on the 5' and 3' ends, respectively. Three clones (1A6, 1B2 and 1B10) were identified as correctly exchanged (**Figure 4.3**) and displayed 25.9%, 75.0%, and 50.0% normal karyotypes, respectively. An exchange efficiency of 3.4% was observed for the plasmid containing the 13.8 kb insert. Conversely, the alternative exchange cassette, containing an additional 12.9 kb of *Nepn* 3' sequence, thus resulting in a total insert size of 26.4 kb, yielded zero positively exchanged clones following two independent electroporations. Electroporation and screening were performed by Susan Hipkens.

### **Blastocyst injections, mouse husbandry and genotyping**

For the *ROSA26<sup>Nepn-Cherry</sup>* allele, chimeric mice were generated by microinjection of exchanged ES cells (clones 5B9/1B2 and 5B9/1B10) into C57BL/6J blastocysts. Resulting chimeras were subsequently mated to C57BL/6J mice to produce mice harboring the *ROSA26<sup>Nepn-Cherry+NeoR</sup>* allele. The FRT-flanked *Neo<sup>R</sup>* sequence was removed by breeding with FLPe-expressing mice (*Tg(ACFLPe)9205Dym*). The final allele, *ROSA26<sup>Nepn-Cherry</sup>* was

maintained on an outbred background by breeding to CD-1 mice. Mouse genotyping was performed using the same primers utilized for RMCE screening on genomic DNA isolated from ear punch tissue using a Puregene genomic DNA isolation kit (Gentra). Following removal of the FRT-flanked *Neo<sup>R</sup>* sequence, the FLPe transgene was removed by via breeding and screening for the absence of FLPe by PCR. Mice were maintained on a 12 hour light-dark cycle, and all experimental protocols were approved by the Vanderbilt Institutional Animal Care and Use Committee. Blastocyst injections were performed by the Vanderbilt Transgenic Mouse/ES Cell Shared Resource.

### **Immunolabeling**

For immunodetection, dissected tissues were fixed in 4% paraformaldehyde in 1X PBS for at least one hour at 4°C, saturated with 30% sucrose at 4°C, embedded in Tissue-Tek Optimal Cutting Temperature (OCT) compound (Sakura), and frozen on dry ice. Sections (7-10 µm) were thawed at room temperature for 5-10 minutes, permeabilized with 0.2% TritonX-100 in 1X PBS for 20 minutes, and washed in 1X PBS. Non-specific binding was blocked by incubating slides with 5% normal donkey serum (NDS; Jackson ImmunoResearch) in 1X PBS. Primary antibodies were used at the following dilutions in 1X PBS with 1% BSA (Sigma): goat anti-Pdx1 (C.V.E. Wright), 1:5000; guinea pig anti-Insulin (Linco), 1:1000; guinea pig anti-Glucagon (Linco), 1:1000; rabbit anti-RFP (Rockland), 1:800. Tissue sections were incubated with primary antibodies overnight at 4°C in a humidified chamber and then washed with 1X PBS. Secondary antibodies were diluted in 1X PBS with 1% BSA as follows: Alexa Fluor 488 donkey anti-goat IgG (Invitrogen), 1:1000; Alexa Fluor 555 donkey anti-rabbit IgG (Invitrogen), 1:1000; donkey anti-guinea pig IgG conjugated to Cy5 (Jackson Immuno Research), 1:500. Sections were incubated with secondary antibodies for 1-2 hours at room temperature in a humidified chamber and then

washed with 1X PBS. Coverslips were mounted with Prolong Gold with DAPI (Invitrogen). Imaging was performed using an Axioplan2 microscope (Zeiss) with a QImaging RETIGA EXi camera.

### Fluorescence-activated cell sorting

Embryos were isolated from wild type CD-1 or *Ptf1a<sup>YFP/+</sup>* females crossed with *ROSA26<sup>Nepn-Cherry</sup>* mice, where the presence of vaginal plug the following day was considered E0.5. Embryo genotypes were identified using a fluorescent microscope, and fluorescence imaging was performed using a Leica MZ 16 FA stereoscope with a QImaging RETIGA 4000R camera. Midgut regions containing the posterior stomach, dorsal and ventral pancreatic buds, and anterior intestine were dissected from wild type, *Ptf1a<sup>YFP/+</sup>*, *ROSA26<sup>Nepn-Cherry</sup>*, and *ROSA26<sup>Nepn-Cherry</sup>; Ptf1a<sup>YFP/+</sup>* embryos in cold 1X PBS. Tissues were dissociated with Accumax (Sigma) at 37°C for up to 1.5 hours with repetitive trituration. The dissociation was quenched with FACS medium (L15 medium containing 1 mg/ml BSA, 10 mM HEPES pH 7.4, 1% penicillin/streptomycin, and 37.5 ng/ml DNase I), and cells were filtered through a 35 µm cell strainer (BD Falcon). Following centrifugation, cells were resuspended in FACS medium or FACS medium with DAPI. Flow cytometry experiments were performed by the VMC Flow Cytometry Shared Resource on a BD Aria III. YFP was excited with a 488 nm laser and emission detected with a 502LP and 530/30 bandpass filter. Cherry was excited with a 561 nm laser and emission detected with a 600LP and 610/20 bandpass filter. Compensation and gates were established using wild type midguts, *Ptf1a<sup>YFP-</sup>* expressing neural tubes and *ROSA26<sup>Nepn-Cherry</sup>* embryos.

### **Whole mount *in situ* hybridization**

For generation of *Nepn* *in situ* hybridization probes, RNA was isolated from E8.5 embryos, cDNA was transcribed, and a 659 bp fragment of *Nepn* sequence was amplified using the following primers: 5' CTAGCAAGCTTGCATTCCCTCCACCAAGAG and 5' CTAGCGAATTCTGACAGGTAAAGATGGGACAGGGTTC. The PCR-amplified fragment was cloned into the pLitmus28i vector. The antisense RNA probe was generated using T7 RNA Polymerase and HindIII-digested plasmid DNA, while the sense RNA probe was generated using T7 RNA Polymerase and XhoI-digested plasmid DNA. Probes were labeled with digoxigenin-UTP (DIG-UTP) using the DIG RNA Labeling Kit (Roche).

Embryos were fixed with 4% formaldehyde, 2 mM EGTA (pH 7.0), and 3 mM NaOH in 1X PBS for at least one hour, dehydrated in a methanol gradient and stored at -20°C. Embryos were rehydrated through a methanol gradient, treated with proteinase K, and post-fixed in 4% formaldehyde and 0.1% glutaraldehyde. Embryos were pre-hybridized for one hour at 65°C in Hybridization Solution (50% Formamide, 1.3X SSC pH 5.0, 5 mM EDTA pH 8.0, 50 µg/ml yeast tRNA, 0.2% Tween-20, 0.5% CHAPS and 100 µg/ml heparin), followed by hybridization overnight at 65°C with an appropriate dilution of each probe in Hybridization Solution. Embryos were subsequently washed repeatedly with Hybridization Solution, followed by washes with TBS-T (1X TBS with 0.1% Tween-20). Embryos were blocked in 2% blocking reagent (Roche) and 20% sheep serum for at least 2 hours and incubated with anti-DIG primary antibody (Roche) for at least 4 hours. Following a series of washes with TBS-T and NTMT (100 mM Tris-HCl pH 9.5, 100 mM NaCl, 50 mM MgCl<sub>2</sub>, 1% Tween-20), embryos were exposed to NBT/BCIP precipitating substrate (Roche). Following visualization of precipitate, embryos were fixed in 4% formaldehyde and 0.1% glutaraldehyde and stored in PBS-T (1X PBS with 0.1% Tween-20) at 4°C.

## Embryo explants

Embryos were dissected in cold 1X PBS, and midguts were cultured in chamber slides (Nunc) at 37°C for 24 hours in DMEM containing 10% calf serum (Hyclone; not heat-inactivated). As indicated, culture medium was supplemented with either recombinant human TGF $\beta$ 2 (R&D Systems) in 4 mM HCl and 0.1% BSA, or an ALK4,5,7 inhibitor, SB431542 hydrate (Sigma) in DMSO. Fluorescence imaging was performed using a Leica MZ 16 FA stereoscope with a QImaging RETIGA 4000R camera.

## Results

### Expression pattern of *Nepn* in *Ptf1a*<sup>YFP</sup>-expressing cells

In the RNA-Seq datasets for *Ptf1a*<sup>YFP</sup>-expressing cells, an analysis of the RPKM values for *Nepn* reveals a meager 1.4-fold increase in expression in *Ptf1a*-deficient cells at E11.5 as compared to *Ptf1a*-expressing MPCs (**Figure 4.1**). Conversely, *Nepn* is upregulated 86.5-fold in *Ptf1a*-expressing MPCs as compared to acinar-specified cells at E15.5. The expression pattern of *Nepn* was further examined in the qRT-PCR profiling of *Ptf1a*<sup>YFP</sup>-expressing and *Ptf1a*-deficient cells. In the temporal and genetic analysis, several genes were identified as highly expressed in both early-stage pancreatic MPCs at E10.5 and in *Ptf1a*-null cells at E11.5 (**Figure 3.5**). Closer examination indicated that *Nepn* is expressed in *Ptf1a*-expressing pancreatic MPCs until E11.5 but virtually absent by E12.5 and later as *Ptf1a*-expressing cells are differentiating into acinar cells (**Figure 4.1**). By qRT-PCR, *Nepn* is 50.5-fold more abundant in *Ptf1a*-deficient cells compared to *Ptf1a*-expressing cells at E11.5 and 35.5-fold more abundant in *Ptf1a*-expressing cells at E11.5 compared to E15.5. While the range of expression differs between the RNA-Seq and qRT-PCR analysis, both datasets show that *Nepn* is most abundant in *Ptf1a*-deficient cells at E11.5 and is expressed in early (E10.5

– E11.5) pancreatic progenitors, while its expression is virtually absent as the *Ptf1a*-expressing cells become specified to the acinar lineage.

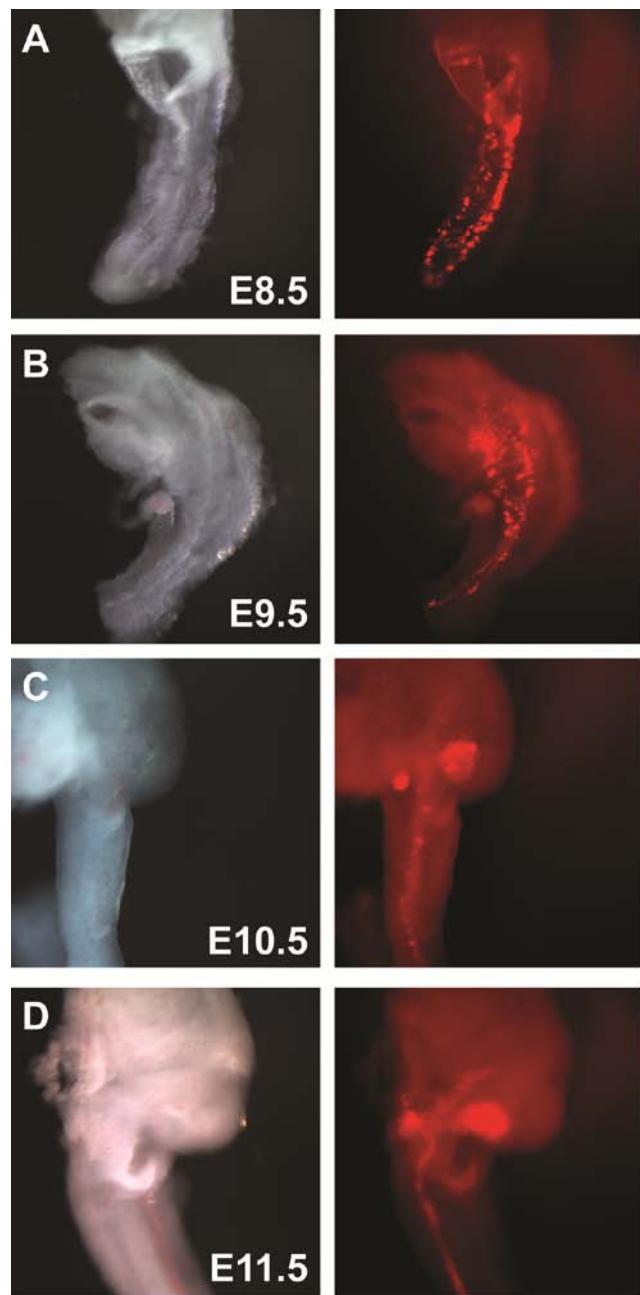
### Generation of a *ROSA26<sup>Nepn-Cherry</sup>* reporter allele

To further explore the expression pattern of *Nepn* during embryogenesis, a three-step procedure was utilized to generate mice containing a *Nepn* reporter in the *ROSA26* locus (**Figure 4.3**) (Zambrowicz et al., 1997). First, a *ROSA26* loxed cassette acceptor (LCA) allele was generated. Gene targeting of the *ROSA26* locus was performed to replace a 5.165 kb region, containing both the promoter and first exon of *ROSA26*, with a lox71/lox2272-flanked dual selection cassette consisting of a fusion of *puromycin resistance* and a mutant *thymidine kinase* driven by the mouse *phosphoglycerol kinase* promoter (*pgk-puΔtk*) and an EM7-driven *neomycin* resistance (EM7-*Neo*<sup>R</sup>) (Chen et al., 2011). Three clones were identified by Southern analysis as correctly targeted (**Figure 4.3**). Second, a lox61/lox2272-flanked exchange cassette containing an FRT-flanked *pgk*/EM7-*Neo*<sup>R</sup> selection cassette and a fusion gene consisting of 9.365 kb of *Nepn* 5'-regulatory sequences (**Figure 4.3**) driving the expression of a monomeric red fluorescent protein, mCherry (Shaner et al., 2004), was generated. The inclusion of 9.365 kb of *Nepn* 5' regulatory sequences was determined by an analysis of homology of the *Nepn* locus among species (**Figure 4.2**). Utilizing RMCE, the *Nepn* reporter cassette was exchanged into the *ROSA26<sup>LCA</sup>* allele and resulted in three correctly exchanged clones (**Figure 4.3**). Lastly, mice containing the *ROSA26<sup>Nepn-Cherry+NeoR</sup>* allele were bred with FLPe-expressing transgenic mice (Rodriguez et al., 2000) to remove the FRT-flanked *pgk*/EM7-*Neo*<sup>R</sup> selection cassette and resulted in mice carrying the *ROSA26<sup>Nepn-Cherry</sup>* allele.

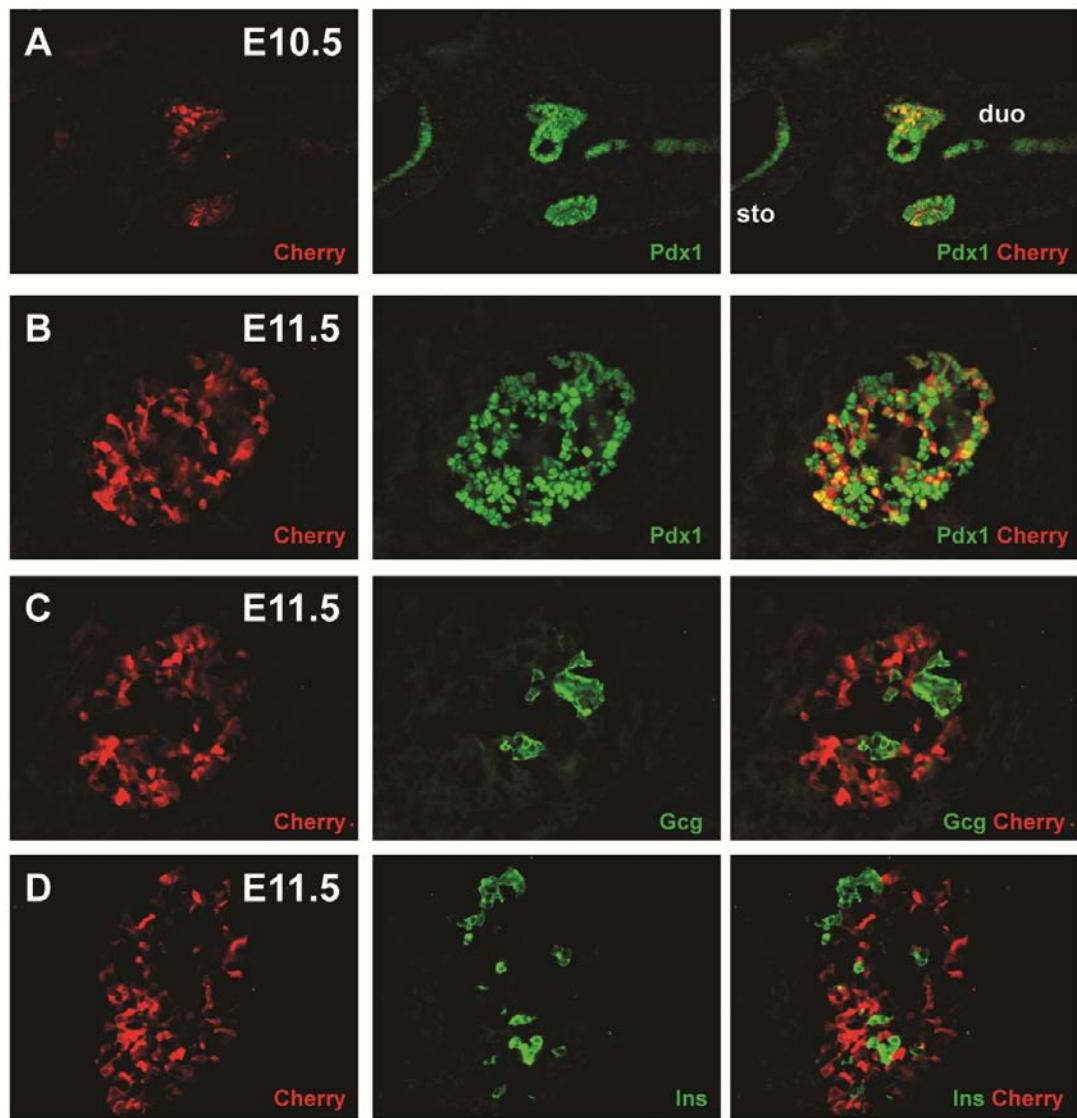
## Expression of *ROSA26<sup>Nepn-Cherry</sup>* transgene

Expression of the *Nepn-Cherry* transgene was first assessed in embryos by whole mount fluorescence microscopy (**Figure 4.4**). Cherry expression was observed in a punctate pattern throughout the midgut and hindgut region of E8.5 (7 – 8 somites) embryos, and the punctate pattern was maintained throughout the midgut region at E9.5. By E10.5, the expression domain of the *Nepn-Cherry* transgene was more clearly defined with stronger expression observed in the dorsal and ventral pancreatic buds and weaker expression present in the caudal stomach, pyloric antrum, duodenum and intestine. A similar pattern was seen at E11.5, with strongest expression present in the dorsal and ventral pancreatic buds. From E10.5 to E11.5, expression of the *ROSA26<sup>Nepn-Cherry</sup>* allele is strikingly similar to the spatial profile of *Pdx1* in the developing midgut (Offield et al., 1996). As development proceeds, Cherry expression is diminished in the pancreas with virtually no expression observed in the pancreas after E15.5. Conversely, as the expression of *ROSA26<sup>Nepn-Cherry</sup>* is diminished in the pancreas, its expression becomes more evident in the embryonic kidneys (data not shown).

To further examine expression of the *ROSA26<sup>Nepn-Cherry</sup>* allele, we performed immunolabeling to identify the population of cells in the developing pancreatic buds that expressed the transgene (**Figure 4.5**). This analysis revealed the co-localization of Cherry with a subpopulation of *Pdx1*-positive pancreatic progenitor cells from E10.5 – E11.5. Additionally, weak Cherry immunolabeling was observed in the caudal stomach, pyloric antrum, and small intestine in a subpopulation of *Pdx1*-positive cells at E10.5 with stronger Cherry immunolabeling observed in the pyloric antrum at E11.5. Interestingly, *ROSA26<sup>Nepn-Cherry</sup>* expression was greatly reduced as cells differentiated towards specific lineages, as evident by the exclusion of Cherry expression in the early insulin- and glucagon-positive cells (**Figure 4.5**). Similarly, as observed in the RNA-Seq and qRT-PCR datasets, the expression of *Nepn* was diminished as *Ptf1a*-expressing cells were specified toward the



**Figure 4.4 Brightfield and fluorescent imaging of *ROSA26*<sup>Nepn-Cherry</sup> embryos.** A) At E8.5, *ROSA26*<sup>Nepn-Cherry</sup> expression is observed in a punctate pattern throughout the developing gut tube endoderm. B) At E9.5, *ROSA26*<sup>Nepn-Cherry</sup> expression can be seen in both the dorsal and ventral endoderm as well as in the posterior endoderm. C) At E10.5, the dorsal and ventral pancreatic buds display higher levels of *ROSA26*<sup>Nepn-Cherry</sup> expression while the antral stomach and duodenum have a punctate expression pattern. D) At E11.5, Cherry expression is more visible in the dorsal and ventral pancreatic buds with continued expression in the antral stomach and duodenum.



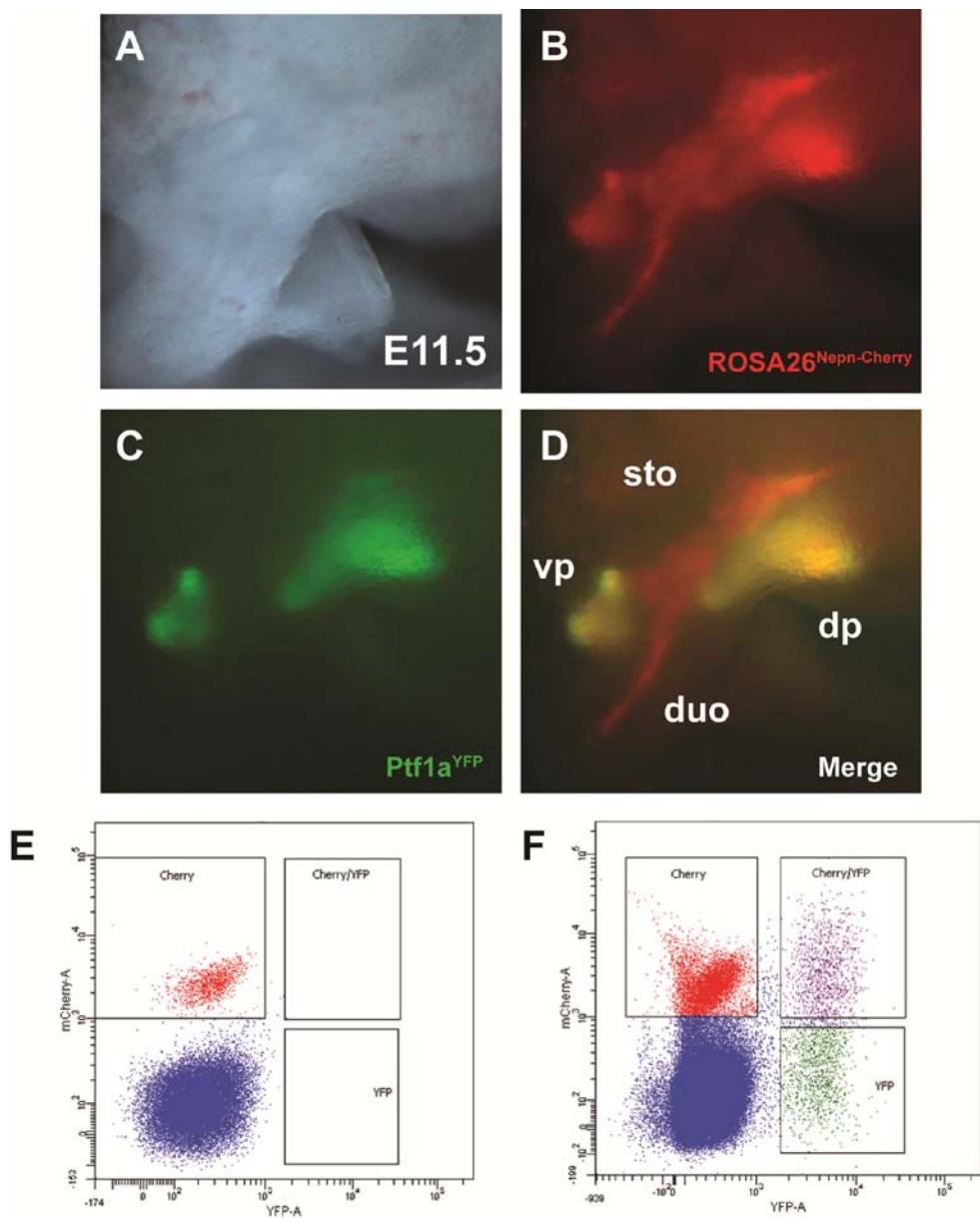
**Figure 4.5 Immunolabeling of *ROSA26<sup>Nepn-Cherry</sup>* embryos.** A) Immunofluorescent analysis of Pdx1 and Cherry expression at E10.5 reveal the co-expression of Pdx1 and Cherry in the dorsal and ventral pancreatic epithelium while lower expression levels of Cherry are observed in the caudal stomach and duodenum. B) Immunofluorescent analysis of Pdx1 and Cherry expression at E11.5 indicates *ROSA26<sup>Nepn-Cherry</sup>* expression in a subpopulation of Pdx1-expressing MPCs in the pancreatic epithelium. C and D) Both glucagon and insulin-expressing endocrine cells are present in the pancreatic epithelium at E11.5 and do not co-express *ROSA26<sup>Nepn-Cherry</sup>*.

acinar lineage (**Figure 4.1**). Based on the temporal pattern of *Nepn* expression during pancreas development, we speculate that *Nepn* plays a role during pancreatic progenitor cell development.

### FACS analysis of *ROSA26<sup>Nepn-Cherry</sup>* cells

To further analyze the *ROSA26<sup>Nepn-Cherry</sup>* cells during development, we crossed mice bearing the *ROSA26<sup>Nepn-Cherry</sup>* allele with *Ptf1a<sup>YFP/+</sup>* mice. At E10.5 and E11.5, whole mount imaging revealed fluorescence of both Cherry and YFP in the dorsal and ventral pancreatic buds with Cherry expression also present in the caudal stomach, pyloric antrum and duodenum (**Figure 4.6** and data not shown). To examine the population of cells which co-express *Ptf1a<sup>YFP</sup>* and *ROSA26<sup>Nepn-Cherry</sup>*, we performed fluorescence-activated cell sorting (FACS). While the optimal excitation wavelengths for Cherry and YFP are 587 nm and 516 nm (Griesbeck et al., 2001; Shaner et al., 2004), the 561 nm and 488 nm lasers were utilized due to instrument configuration limitations.

Consistent with the expression pattern observed by immunolabeling, FACS analysis of *ROSA26<sup>Nepn-Cherry</sup>*; *Ptf1a<sup>YFP/+</sup>* embryos revealed that a subpopulation of the *Ptf1a<sup>YFP</sup>*-expressing MPCs is positive for *Nepn-Cherry* (**Figure 4.6**). While we observed a population of Cherry single-positive cells, which would represent the Cherry-positive cells in the caudal stomach and intestine or other Cherry-single positive cells, we observed a significant Cherry-positive population in wild type embryos. While these wild type embryos are not Cherry-positive by whole mount microscopy or genotyping, the dissected tissues exhibited fluorescence in the Cherry parameters by FACS. This undesirable fluorescence presents a limitation when analyzing and attempting to isolate Cherry single-positive cell populations; however, when the *ROSA26<sup>Nepn-Cherry</sup>* allele is used in combination with another fluorescently-



**Figure 4.6 FACS analysis of *ROSA26<sup>Nepn-Cherry</sup>*; *Ptf1a<sup>YFP</sup>* embryos.** A – D) At E11.5, *ROSA26<sup>Nepn-Cherry</sup>* expression is observed in the caudal stomach, intestine and the dorsal and ventral pancreatic buds (B) while *Ptf1a<sup>YFP</sup>* expression is restricted to the pancreatic buds in the developing midgut (C). Overlay displays the co-incidence of *Ptf1a<sup>YFP</sup>* and *ROSA26<sup>Nepn-Cherry</sup>* in the pancreatic buds (D). Dorsal pancreas (dp), ventral pancreas (vp), stomach (sto), and duodenum (duo). E) FACS analysis of wild type embryos revealed non-specific fluorescence in the Cherry single-positive gate. F) Analysis of *ROSA26<sup>Nepn-Cherry</sup>*; *Ptf1a<sup>YFP</sup>* embryos reveals that *Ptf1a*-expressing MPCs can be divided into two populations: *Ptf1a<sup>YFP</sup>*-positive (YFP) and *ROSA26<sup>Nepn-Cherry</sup>*; *Ptf1a<sup>YFP</sup>* double-positive cells (Cherry/YFP). Additionally, Cherry single-positive cells are identified which represent either non-specific fluorescence or *ROSA26<sup>Nepn-Cherry</sup>* positive cells in the caudal stomach and intestine.

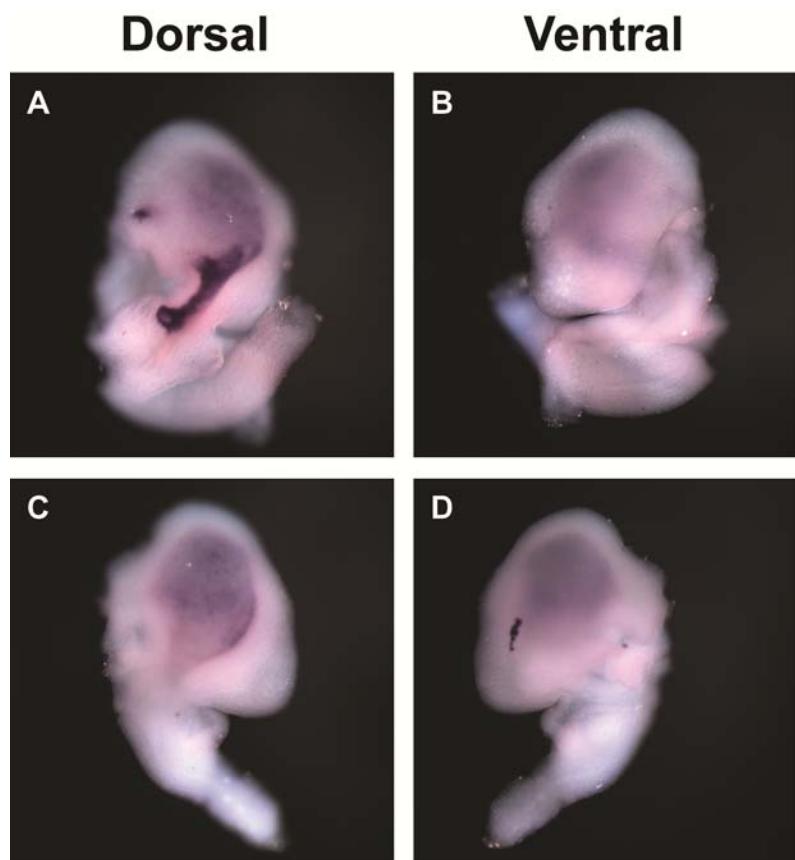
tagged allele, such as *Ptf1a*<sup>YFP</sup>, then dual-positive cells are able to be distinguished and isolated without significant contamination.

### **Analysis of *Nephrocan* expression by *in situ* hybridization**

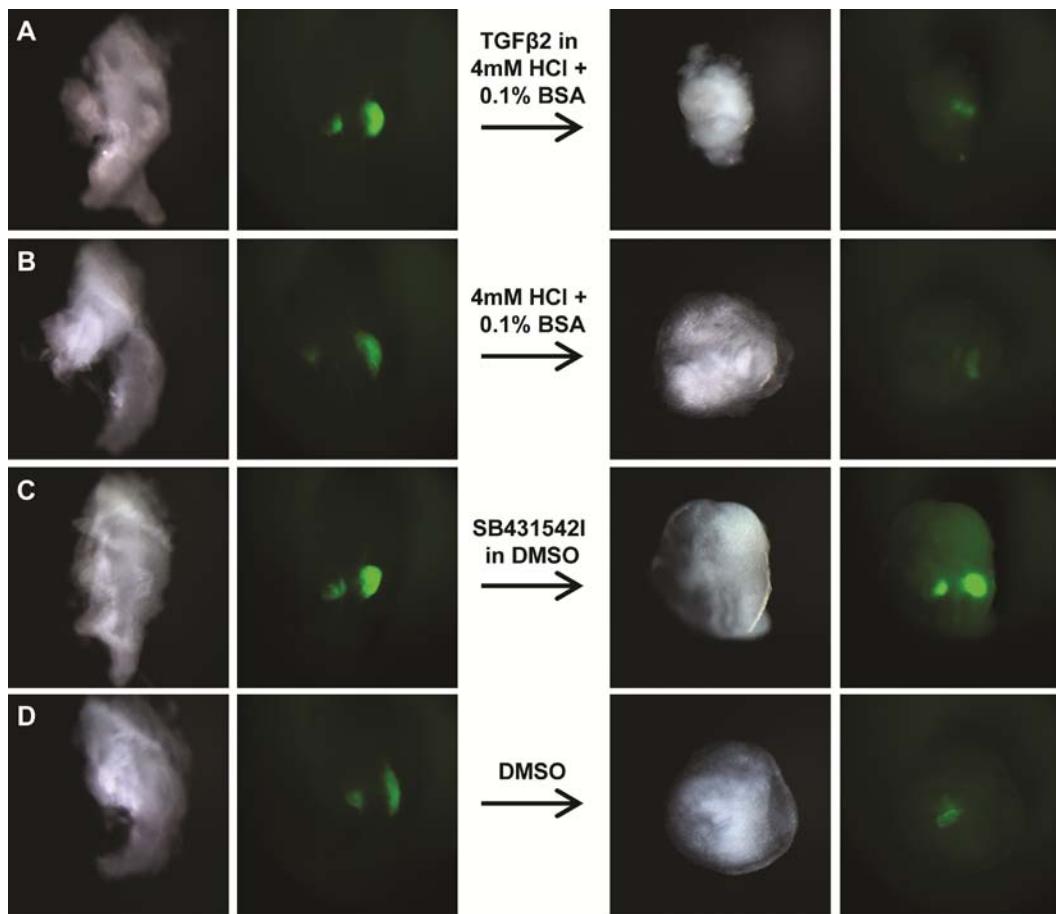
To assess endogenous expression of *Nepn* during embryogenesis, we performed *in situ* hybridization on E11.5 embryos (**Figure 4.7**). Given that *Nepn* expression was detected in pancreatic progenitor cells by both RNA-Sequencing and qRT-PCR, we anticipated that *Nepn* would be detected in the pancreatic epithelium by *in situ* hybridization. However, when hybridized with an RNA probe for *Nepn*, expression was only visualized in the pyloric epithelium, consistent with previous reports (Li et al., 2009). Similar to the expression pattern previously observed, *Nepn* exhibited expression predominantly in the anterior portion of the pylorus; however, there is no clear boundary of expression across the pylorus. Although the *in situ* hybridization did not reveal expression of *Nepn* in the pancreatic epithelium, it is possible that the transcript level of *Nepn* in the pancreatic epithelium is below detection threshold for this assay.

### **Analysis of TGF $\beta$ signaling in embryos**

Given that *Nepn* is a secreted modulator of TGF $\beta$  signaling, we sought to further assess the role of TGF $\beta$  signaling during the development and differentiation of early pancreatic progenitor cells. To examine the impact of both stimulation and inhibition of TGF $\beta$  signaling, we utilized embryos at E9.5 from a fluorescently-tagged *Pdx1* reporter allele (discussed in Chapter V) (Potter et al., 2011). Tissue explants containing the pancreatic epithelium were cultured with either TGF $\beta$ 2 or SB431542, an inhibitor of the TGF $\beta$  receptors (ALK4, 5, and 7). The expression of *Pdx1* was assessed by whole mount fluorescence microscopy following either stimulation or inhibition of TGF $\beta$  (**Figure 4.8**). Following 24



**Figure 4.7** *In situ* analysis of endogenous *Neprocan*. A – B) Dorsal and ventral views following *in situ* hybridization with an anti-sense probe for *Nepn* at E11.5. Expression was predominantly detected at the pylorus with expression evident in the caudal stomach and duodenum. Endogenous *Nepn* expression was not observed in the pancreatic epithelium by the whole mount *in situ* hybridization; however, the expression level of *Nepn* in the developing pancreatic buds may be below the detection threshold of this assay. C – D) Dorsal and ventral views following *in situ* hybridization with a sense probe for *Nepn* at E11.5 as a control.



**Figure 4.8 Embryo explants to study TGF $\beta$  signaling in specification of progenitor cells.** Midgut regions containing both the dorsal and ventral endoderm were isolated from *Pdx1*<sup>CFP/+</sup> embryos and imaged by fluorescent microscopy. Explants were cultured for 24 hours with either a stimulant, TGF $\beta$ 2, or an inhibitor of the TGF $\beta$  signaling pathway, SB43154I, (A and C) as well as corresponding control conditions (B and D). Following 24 hours of culture, explants were imaged by fluorescent microscopy. The maintained fluorescence in the SB43154I-treated *Pdx1*<sup>CFP</sup> explants suggests that TGF $\beta$  inhibition promotes *Pdx1* expression in the dorsal and ventral endoderm at this developmental stage.

hours of culture, treatment of the tissue explants with TGF $\beta$ 2 showed little alteration in the induction of the early pancreatic epithelium marker *Pdx1*, whereas exposure to the TGF $\beta$  inhibitor slightly enhanced induction of *Pdx1* as demonstrated by the increased fluorescence of the reporter allele as compared to the control and TGF $\beta$ 2-stimulated explants.

## Discussion

### Generation of a reporter allele for *Nephrocan* by RMCE

Given that there are few studies concerning the expression pattern of *Nepn* and a reporter allele has not been documented, we sought to generate a fluorescent reporter mouse line to enable the identification of *Nepn*-expressing cells. To achieve this, we utilized RMCE to insert a single copy of a *Nepn-Cherry* transgene into a transcriptionally-disabled *ROSA26* locus. This method was advantageous specifically for two reasons. First, RMCE is more efficient than traditional gene targeting. In RMCE, a DNA cassette can be exchanged rapidly into an acceptor allele and simply screened via PCR, eliminating the need to screen by Southern analysis which is more laborious. Second, this method allowed for a single-copy insertion of the *Nepn-Cherry* transgene into a specified locus, *ROSA26*, which presents advantages over standard BAC transgenic mice which can exhibit multiple insertions and transgene silencing.

Although the method employed to generate the *ROSA26*<sup>*Nepn-Cherry*</sup> allele has certain advantages, our approach is not without certain disadvantages. First, *ROSA26* is known to be ubiquitously expressed, thus insertion into this locus could result in a ubiquitously expressed transgene. However, given that the design of the *ROSA26*<sup>LCA</sup> allele resulted in the removal of both the promoter region and exon 1 of *ROSA26*, the potential for ubiquitous expression of an inserted transgene is small. Second, while our *Nepn-Cherry* reporter incorporated 9.365 kb

of sequence upstream and 833 bp downstream, the proximal and distal regulatory elements driving *Nepn* transcription are unknown. Thus, it is possible that the *ROSA26<sup>Nepn-Cherry</sup>* reporter allele does not accurately mimic *Nepn* expression or the allele may display variegation. However, the construct incorporated all the 5' sequence up to the next gene located upstream (**Figure 4.2**) and the expression pattern seen by whole mount imaging during early endoderm development was similar to previous reports (**Figure 4.4**).

While the maximum size of an exchange cassette for RMCE has not been fully elucidated, the Nepn-Cherry exchange cassette was one of the largest cassette exchanges we have performed. The exchange cassette was 13.8 kb in size and resulted in 3.4% exchange efficiency, a percentage lower than what we typically observe for RMCE. In addition to this exchange cassette, an alternative exchange cassette that was 26.4 kb in size, containing an additional 12.9 kb of *Nepn* 3' sequence, yielded zero positively exchanged clones from two separate electroporations.

### ***ROSA26<sup>Nepn-Cherry</sup>* is expressed in a subset of pancreatic MPCs**

Our analyses indicate that *Nepn* is variably expressed in a subpopulation of pancreatic MPCs from E10.5 to E11.5, with a subset of *Pdx1*-expressing MPCs co-expressing *Nepn* as observed by immunolabeling, and a subpopulation of *Ptf1a*-expressing MPCs expressing *Nepn* as observed by FACS analysis. The generation of a fluorescently-tagged reporter allele for *Nepn* permits isolation of two subpopulations of pancreatic MPCs by FACS: 1) MPCs co-expressing both *Ptf1a<sup>YFP</sup>* and *ROSA26<sup>Nepn-Cherry</sup>* and 2) MPCs expressing only *Ptf1a<sup>YFP</sup>* (**Figure 4.6**). Using these cells, we will be able to examine and characterize potential differences and similarities between the two cell populations. We speculate that MPCs co-expressing both *Ptf1a<sup>YFP</sup>* and *ROSA26<sup>Nepn-Cherry</sup>* represent cells in an earlier progenitor cell state, while MPCs expressing only *Ptf1a<sup>YFP</sup>* represent cells in a more transitional phase of

differentiation and specification. This hypothesis is based on a previous report utilizing single-cell profiling of E10.5 dorsal pancreatic epithelium which identified at least six types of progenitor cells based on combinations of *Pdx1*, *Nkx2-2*, *Nkx6-1*, *Ptf1a*, *Neurog3*, *Pax6*, *Pax4*, *Isl1*, *Neurod1*, *Gcg*, *Ins*, *Sst* and *Ppy* expression (Chiang and Melton, 2003). Of the six progenitor cell types identified, two types (32/60 cells) expressed *Ptf1a* and were subdivided based on expression of *Neurog3*. Given the two different *Ptf1a*-expressing populations classified in that study, we speculate that the MPCs expressing only *Ptf1a*<sup>YFP</sup> may represent cells that have become further specified and have activated transcription for specific pre-endocrine progenitor markers, such as *Neurog3*, while MPCs co-expressing both *Ptf1a*<sup>YFP</sup> and *ROSA26*<sup>Nepn-Cherry</sup> may represent earlier progenitor cells that express markers such as *Pdx1*, *Nkx2-2*, and *Nkx6-1*.

While *Nepn* is expressed in a population of *Ptf1a*-expressing cells, it is unlikely *Nepn* itself is regulated by *Ptf1a* given that it is also very highly expressed in *Ptf1a*-null cells at E11.5 (**Figure 4.1**). Given that *Nepn* is also expressed in other posterior foregut derivatives, the increased expression of *Nepn* in *Ptf1a*-null cells may also reflect reprogramming of the cells to alternative posterior foregut fates. Interestingly, *Nepn* has also been found to be upregulated in *Nkx2-2*-null embryos (Anderson et al., 2009), an observation that may reflect developmental arrest of *Nkx2-2*-deficient cells at a pancreatic MPC state. In this study, the expression of *Nepn* was most abundant in E12.5 and E13.5 *Nkx2-2*-mutant embryos with significant downregulation by E15.5. Thus, it is plausible that at later stages of development when the requirement for *Nkx2-2* is the greatest, the *Nkx2-2*-mutant cells are “held” in a progenitor-like state, thus expressing *Nepn*.

### ***ROSA26<sup>Nepn-Cherry</sup>* expression is downregulated as MPCs differentiate**

The possible requirement for Nepn in regulating pancreas development is likely to be temporally-dependent as evident by its diminished expression as MPCs become further specified. This was first observed in the RNA-Seq datasets when the expression of *Nepn* in *Ptf1a*-expressing cells was downregulated over 86-fold from E11.5 to E15.5 (**Figure 4.1**). Similarly, in the qRT-PCR datasets, the expression of *Nepn* is downregulated over 35-fold from E11.5 to E15.5 and 12,000-fold from E10.5 to E18.5 (**Figure 4.1**). In addition, by immunohistochemical analysis at E11.5, the expression of the *Nepn<sup>Cherry</sup>* transgene is virtually absent in the first-wave hormone-positive cells as they are specified from early progenitors (**Figure 4.5**).

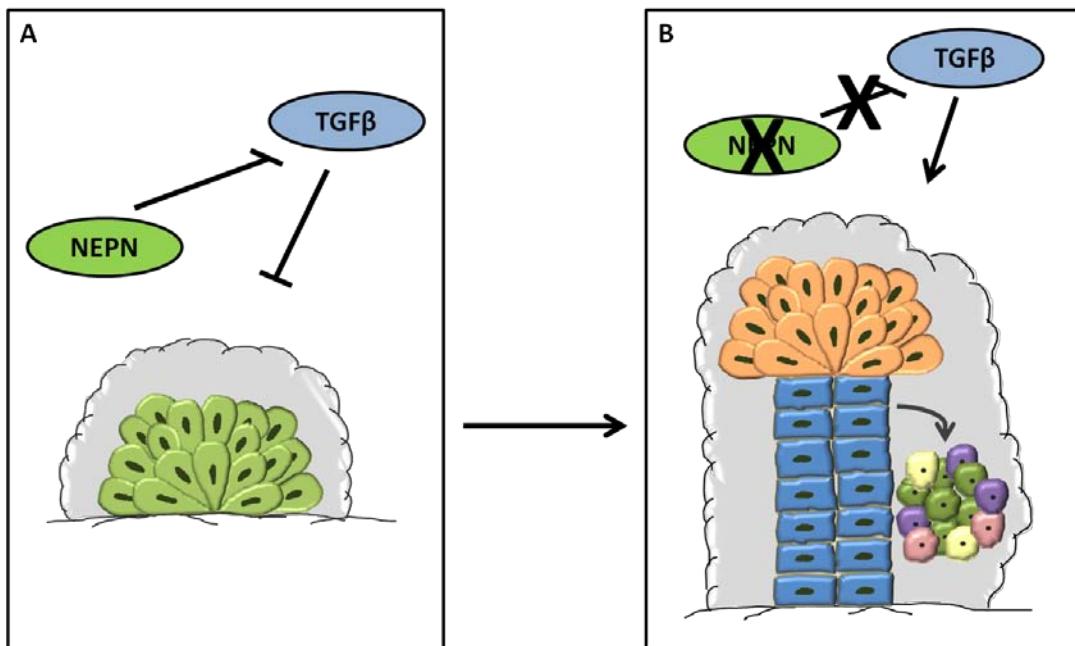
### ***ROSA26<sup>Nepn-Cherry</sup>* future directions**

In addition to analyzing *Nepn*-expressing pancreatic progenitor cells in combination with the *Ptf1a<sup>YFP</sup>* allele, the *ROSA26<sup>Nepn-Cherry</sup>* reporter is advantageous for use in two other studies. First, given that the *Nepn<sup>Cherry</sup>* transgene is expressed in a punctate pattern throughout the midgut and hindgut endoderm, the *ROSA26<sup>Nepn-Cherry</sup>* reporter mice can be used in combination with mice containing a *Sox17<sup>GFP-Cre</sup>* allele (unpublished). The fluorescently-tagged *Sox17* allele allows for the identification of the definitive endoderm during early embryogenesis, as well as the foregut, midgut and hindgut at E8.5. The analysis of mice bearing both *ROSA26<sup>Nepn-Cherry</sup>* and *Sox17<sup>GFP-Cre</sup>* will allow for the identification of *Nepn<sup>Cherry</sup>*-expressing cells within the *Sox17*-expressing endoderm. The isolation of these cells and their characterization by RNA-Sequencing would allow us to examine the difference between endodermal cells that express both *Nepn<sup>Cherry</sup>* and *Sox17<sup>GFP</sup>* or simply *Sox17<sup>GFP</sup>*. This comparison, as well as further analyses by immunolabeling and utilizing embryo explants, will help us examine if Nepn plays a role in endoderm patterning through the inhibition of

TGF $\beta$  signaling. Additionally, given that the Nepn transgene is expressed in a pattern strikingly similar to Pdx1, the *ROSA26*<sup>Nepn-Cherry</sup> reporter mice could be used in combination with mice containing a fluorescently-tagged *Pdx1* allele (see Chapter V), which would facilitate examination of similarities and differences in their expression patterns as well as permit the isolation and characterization of dual-positive cells in the pancreas, caudal stomach and duodenum.

### Potential role of *Nephrocan* in pancreatic progenitor development

Given that TGF $\beta$  signaling inhibits specification of ventral pancreatic progenitors (Wandzioch and Zaret, 2009), Nepn functions to inhibit TGF $\beta$  signaling (Mochida et al., 2006), and the *Nepn*<sup>Cherry</sup> transgene is expressed in some, but not all, pancreatic MPCs (**Figure 4.5** and **4.6**), we speculate that Nepn plays a role in early pancreas specification (**Figure 4.9**). Similar to other signaling pathways during embryogenesis, the requirement for TGF $\beta$  signaling is likely to be both temporally and spatially defined. While the inhibition of TGF $\beta$  signaling may promote progenitor cell specification, TGF $\beta$  signaling is critical for endocrine cell specification (Miralles et al., 1998; Sanvito et al., 1994). Consistent with this temporal role is our findings. First, we observed by RNA-Seq that *Nepn* is expressed during early progenitor cell development, thus possibly inhibiting TGF $\beta$  signaling in these cells and thus promoting progenitor specification. Using the *ROSA26*<sup>Nepn-Cherry</sup> reporter, we were also able to observe expression of the *Nepn*<sup>Cherry</sup> transgene in a subset of the pancreatic MPCs. Second, we observed that *Nepn* expression decreased as *Ptf1a*<sup>YFP</sup>-expressing cells were committed to the acinar lineage. Similarly, we observed that the expression of the *Nepn*<sup>Cherry</sup> transgene was diminished as MPCs began to differentiate towards specific lineages, including hormone-positive cells evident during the first wave of endocrine specification (**Figure 4.9**).



**Figure 4.9 Model of potential role of Nepn in pancreas development.** A) During pancreatic progenitor cell specification, TGF $\beta$  signaling inhibits specification of MPCs. Given that Nepn is expressed in a subset of pancreatic MPCs and that Nepn inhibits TGF $\beta$  signaling, we speculate that Nepn plays a role in the specification of pancreatic MPCs. Green: pancreatic MPCs; Grey: mesenchyme. B) During later development, TGF $\beta$  signaling is critical for endocrine cell specification. We observed a downregulation in Nepn expression as MPCs differentiate to both the endocrine and acinar cell lineages. Thus, we speculate that the downregulation of Nepn in these lineages relieves the inhibition of TGF $\beta$  signaling and thereby promotes mature cell specification. Orange: acinar cells; blue: biopotential (endocrine/duct) progenitor cells; cluster: Islet of Langerhans.

To more fully elucidate the plausible role of Nepn in pancreatic progenitor cell development, it would be advantageous to generate a knockout allele (described in Materials and Methods). However, seeing that Nepn is expressed in definitive and midgut endoderm, it is possible that global knockout mice would cease development shortly after endoderm specification. Thus, it would be more advantageous to generate a conditional knockout where the coding sequence for *Nepn* is flanked by lox sites. Using this approach, mice containing the *Nepn*<sup>flox/flox</sup> allele can be mated to Cre-expressing mice to conditionally eliminate *Nepn* at specific temporal stages during development, such as in early endoderm or pancreatic progenitors using *Foxa2-Cre* or *Pdx1-Cre*. These studies would allow for a better assessment of the temporal requirement for Nepn in the developing endoderm and permit further elucidation of the potential role of Nepn in pancreatic progenitor cell specification.

## CHAPTER V

### FLUORESCENT PROTEIN REPORTER ALLELES FOR PDX1

#### Introduction

Numerous studies have revealed the importance of pancreatic and duodenal homeobox 1 (Pdx1) in the specification and subsequent development of the pancreas (Gu et al., 2002; Jonsson et al., 1994; Offield et al., 1996; Stoffers et al., 1997a; Stoffers et al., 1997b). Prior to its analysis in mammalian model organisms, it was identified first in *Xenopus laevis* as a homeobox gene, *XlHbox8*, which was expressed in the epithelial cells of the pancreatic anlagen and duodenum (Wright et al., 1989). Within mammalian model organisms, Pdx1 has a number of aliases including somatostatin-transactivating factor 1 (Stf-1) (Leonard et al., 1993), islet/duodenum homeobox 1 (Idx-1) (Miller et al., 1994), insulin promoter factor 1 (Ipf-1) (Jonsson et al., 1994; Ohlsson et al., 1991), and maturity onset diabetes of the young 4 (MODY4) (Stoffers et al., 1997a).

Mammalian Pdx1 was first identified as a transcription factor that regulates the transcription of *insulin* and *somatostatin* (Leonard et al., 1993; Ohlsson et al., 1993). At the developmental level, *Pdx1* displays restricted expression in the posterior foregut beginning at E8.0, is expressed throughout the pancreatic epithelium at E9.5, and later in development, high levels of *Pdx1* expression are evident primarily in the  $\beta$  cells (Ohlsson et al., 1993). The temporal expression pattern of Pdx1 highlights two distinct functional roles, first as a critical regulator of pancreas development where its early expression is thought to initiate a complex transcriptional network that leads to the formation of the mature pancreas, and second as a regulator of glucose-induced gene expression (Melloul et al., 1993). The critical role of Pdx1 in pancreas development is further made evident by *Pdx1*-null mutations which result in

pancreatic agenesis in both mouse and human (Jonsson et al., 1994; Offield et al., 1996; Stoffers et al., 1997b). Additionally, *Pdx1* haploinsufficiency results in impaired glucose tolerance and early onset diabetes which led to the identification of *Pdx1* as a genetic locus of early-onset diabetes, known as *MODY4* (Ahlgren et al., 1998; Stoffers et al., 1997a).

Within the developing foregut, *Pdx1* expression begins at E8.0 (approximately 7 somites) in a small population of ventral foregut endothelial cells on the left side of the anterior intestinal portal (AIP). By E8.5, expression is observed on both the left and right sides of the AIP, and by E9.0, expression in the dorsal endoderm is observed (Gannon et al., 2000). At E11.5, the broad foregut expression pattern of *Pdx1* is evident by its expression in the dorsal and ventral pancreatic epithelium, as well as expression in the epithelium of the duodenum, cystic duct, common bile duct and antral stomach (Offield et al., 1996). Lineage tracing results using a Cre/loxP system have shown that *Pdx1*-expressing progenitors can give rise to all three lineages of the mature pancreas (Gu et al., 2002; Herrera, 2000). Notably, it has been suggested that the coinciding expression of *Pdx1* and *Ptf1a* in the pancreatic epithelium from E9.5 to E11.5 demarcates the pancreatic MPC population which gives rise to all cell types of the mature pancreas (Burlison et al., 2008; Zhou et al., 2007).

Numerous lines of mice expressing green fluorescent protein (GFP) (Chalfie et al., 1994; Heim et al., 1995; Shimomura et al., 1962) under control of a variety of gene loci, including *Pdx1* (Micallef et al., 2005), have been reported. However, the utility of some of these reporter alleles, including that of the *Pdx1*<sup>GFP</sup> allele, is limited by a single spectral profile which prevents their combinatorial use. Indeed, the isolation of many distinct cell populations by fluorescence-activated cell sorting (FACS) requires the simultaneous use of two or more FPs with spectrally distinct excitation and emission profiles (Heikal et al., 2000; Nagai et al., 2002; Rizzo et al., 2004; Shaner et al., 2004). Given the differing expression patterns of *Pdx1* and *Ptf1a* within the foregut endoderm, comparisons of the different cell

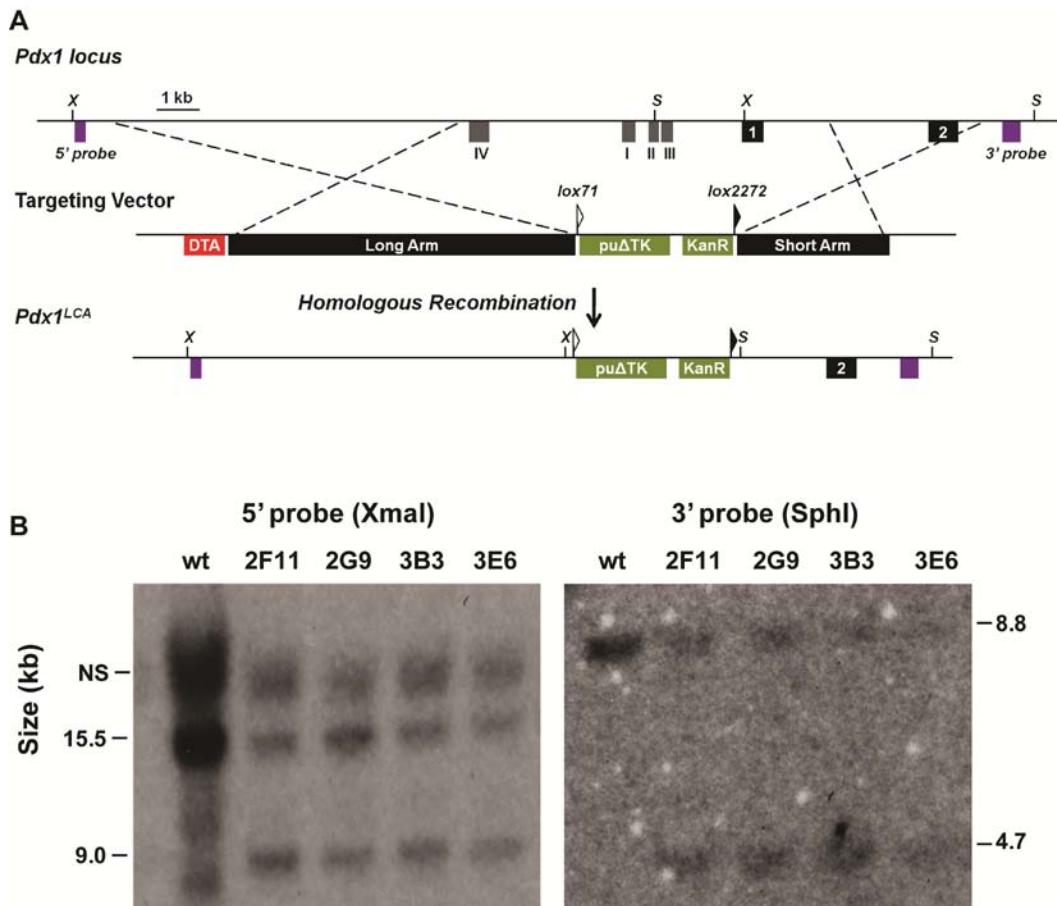
populations will help to define the genetic pathways and transcriptional networks that regulate the development of spatially-defined tissues. In order to achieve these comparisons, it is pertinent to obtain a fluorescently-tagged *Pdx1* allele that can be used in combinatorial fluorescence-activated cell sorting together with GFP- or YFP-expressing reporter alleles.

## Materials and Methods

### Generation of a loxed cassette acceptor allele for *Pdx1*

A gene targeting vector for *Pdx1* was constructed using BAC recombineering and utilized the Cre/lox system for future manipulations by recombinase-mediated cassette exchange (RMCE). The long and short homology arms for *Pdx1*, 8.128 kb and 3.584 kb, respectively, were derived from an RP22 BAC (clone 228-B22). The targeting replaced an 8.6 kb region of the *Pdx1* gene, spanning from -6586 to +2024 bp, with a mouse *phosphoglycerol kinase* promoter (*pgk*) driving expression of a *puromycin resistance*-*Athymidine kinase* fusion gene (*puΔTK*) and a bacterial EM7 promoter driving expression of a *kanamycin resistance* gene (EM7-*Kan*<sup>R</sup>). Both selectable markers were flanked with tandemly-oriented lox71 and lox2272 sites. An MC1-driven *diphtheria toxin A* gene (*DTA*) was placed outside the long homology arm to select against nonrecombinant clones (**Figure 5.1**).

The targeting vector was linearized with NotI prior to electroporation into TL1 mouse ES cells. Four electroporations were performed in which 40 µg of the linearized targeting vector was electroporated into 5.6 x 10<sup>6</sup> TL1 mES cells (passage 13) and cultured with puromycin (1.5 µg/ml) to select for positive insertions. After selection, 192 surviving clones were picked for further expansion and screening. Clones that had undergone the desired homologous recombination events were identified by a band of 9.0 kb after digestion with



**Figure 5.1 Generation of a *Pdx1* floxed cassette acceptor (LCA) allele.** A) Schematic representation of the *Pdx1* locus, targeting vector, and *Pdx1*<sup>LCA</sup> allele. An 8.6 kb region containing the enhancer elements (Areas I-IV) and exon 1 was replaced by a floxed *puΔTK/EM7-Kan<sup>R</sup>* sequence flanked by minimal (34 bp) tandemly oriented lox71 and lox2272 sites. The targeting vector also contains a mouse *pgk*-driven *diphtheria toxin A* gene (*DTA*) outside the long homology arm for negative selection following targeting. Restriction endonucleases (X: XmaI, S: SphI) and probes for Southern hybridization are noted. B) Southern analysis of four ES cell clones following gene targeting. Clones 2F11, 2G9, 3B3, and 3E6 were identified as correctly targeted by a 9.0 kb band following hybridization with a 5' probe and a 4.7 kb band following hybridization with a 3' probe.

XmaI (wild type band: 15.5 kb) and hybridization with a 5' probe and a band of 4.7 kb after digestion with SphI (wild type band: 8.8 kb) and hybridization with a 3' probe (**Table 5.1** and data not shown). Clone 1H5 showed correct targeting at both the 5' and 3' ends of the *Pdx1* allele and was injected into blastocysts. Although clone 1H5 displayed germline competence, subsequent use in RMCE experiments has proven unsuccessful.

Due to the germline incompetence of clone 1H5 in subsequent RMCE experiments, gene targeting was performed for a second time at the Vanderbilt Transgenic Mouse/ESC Shared Resource (TMESCSR). In brief, 200 µg of the linearized targeting vector was electroporated into  $35 \times 10^6$  TL1 mES cells (passage 13) and cultured with puromycin (1.5 µg/ml) to select for positive insertions. After selection, 269 surviving clones were picked for further expansion and screening. DNA was extracted from 172 clones and screened by Southern hybridization as noted above. Clones 2F11, 2G9, 3B3, and 3E6 all displayed correct targeting at both the 5' and 3' ends of the *Pdx1* allele (**Figure 5.1**).

### **Generation of a GFP/Puro<sup>R</sup>-expressing reporter for RMCE**

A basal exchange vector for *Pdx1* (termed pPdx1.Ex1) was constructed to facilitate genetic manipulation of the locus and subsequent insertion by RMCE into mES cells containing the *Pdx1<sup>LCA</sup>* targeted locus. This basal exchange vector includes lox66 and lox2272 sites flanking the 8.6 kb DNA sequence, containing Areas I-IV and exon 1 of the *Pdx1* locus, removed during gene targeting. A fragment from pPdx1.Ex1, obtained by digestion with BglII, was cloned into a modified pGL2 vector (Promega) for construction of the Pdx1<sup>GFP/Puro<sup>R</sup> exchange cassette. The EagI to MluI fragment of the *Pdx1* basal exchange vector was PCR amplified to introduce FseI and MluI sites, and kinased oligos containing sequences for a FLAG-tag and multiple cloning sites (NheI, NdeI, XhoI) were cloned into the modified BglII fragment. The sequences for enhanced GFP (eGFP), three-repeats of a SV40</sup>

**Table 5.1 Southern blot hybridization probes for *Pdx1* gene targeting**

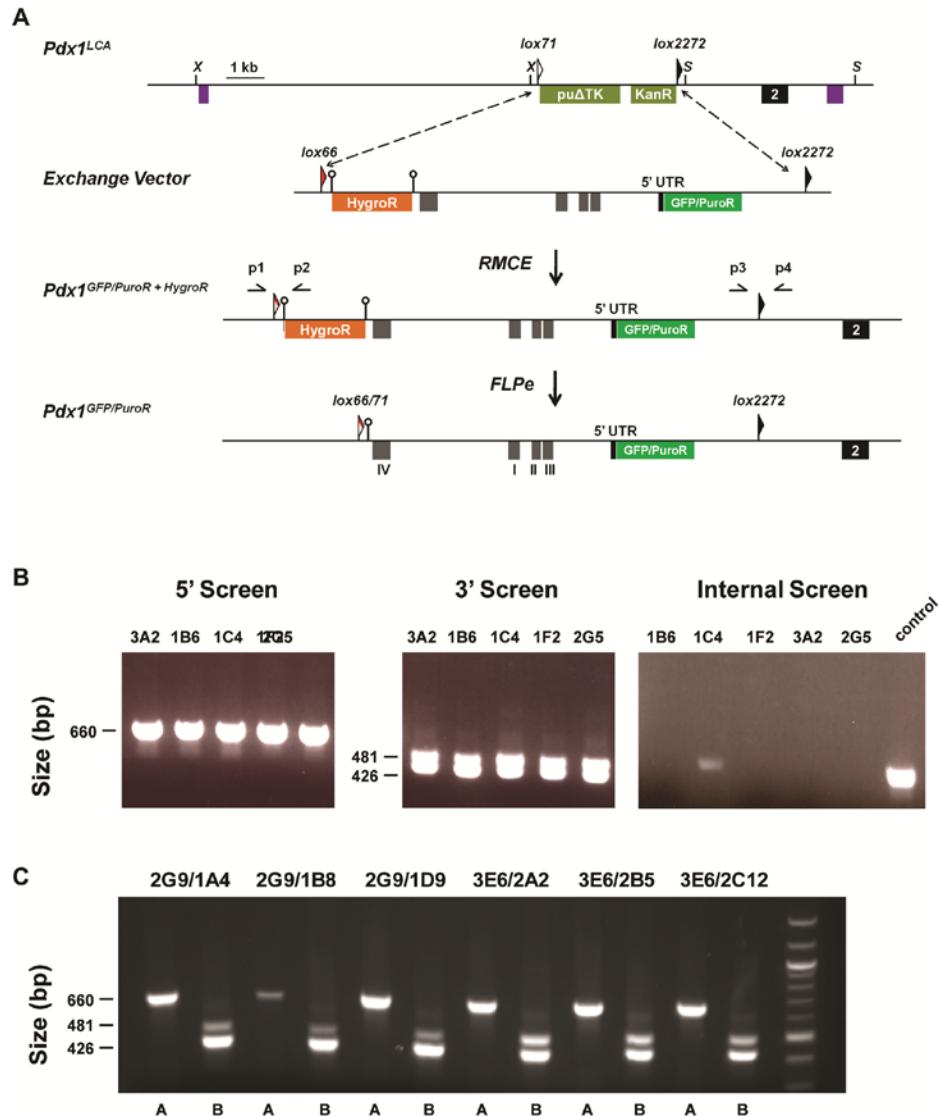
5' Hybridization Enzyme: XmaI Wild type allele: 15.5 kb Targeted allele: 9.0 kb	GAGGCAAGTAGGAACCTTATCTTACATTGGATGATAG CTGTCAAGATGGCACTGGCAAACGTCAAAAGAAC CCAGGGCCAGTGGTTATCTGTGGTGCCTATTAGCATA GAAAGCAGACGCTGACCTCCAGGATCTCTGGGCTTGC AGACACTGGGTACAGAGCTGGCCCCCTACCCCTAAACTT CTCCAGCTCATGGGCTTGCTCCCTTCTCCACTTCTCA TTCATACATAACCATGCCAGTTCAGCCTTGCTTCC
3' Hybridization Enzyme: SphI Wild type allele: 8.8 kb Targeted allele: 4.7 kb	GAGTCACACTGGTAGAAAGAAACTCTCTTAATCAAAA ATGTGTGTTTCCCCCTCTTTAGGGTGAATGCATCC TTTCTGGATGCTGCTCACAGTGTGGCTCTGGATGAG ACAAATCTACATTAGAGATGTCAAAAAAATTCAAGGT CTTCTCCCACTTAATATAGTGGCAAAGTTGTGCTAC CTTCTAGGGTTGCTGTAGGGTTAAATTAAAGATATAA ACCGGCTTACTTAATATAGAAAGGATTCCATCCGACC ATATTATTAGCAAACGTGTTGCTCATCATCATTCTT GAAGAGCCTGGCCTAGAAGCTGCCACTAGCCTCTCC AGTCTCCATGGGCCCCCTGAAGGGTCTCTCCTTAGCAA ACCCCTGTACAGTTGAAGTGATTTCAGGTACCCATT GGTCTTAGCTAAGCA

nuclear localization signal (3X SV40 NLS), an internal ribosome entry site (IRES2), and *puromycin resistance* (*Puro*<sup>R</sup>) were PCR amplified and cloned into the modified BglII fragment. The final fragment was extracted via BglIII digestion and cloned into the basal exchange vector (pPdx1.Ex1). A *pgk*-driven *hygromycin resistance* (*Hygro*<sup>R</sup>) sequence, flanked by tandem flipase recognition target (FRT) sites, was cloned into a NotI site at the 5' end of the exchange vector (**Figure 5.2**).

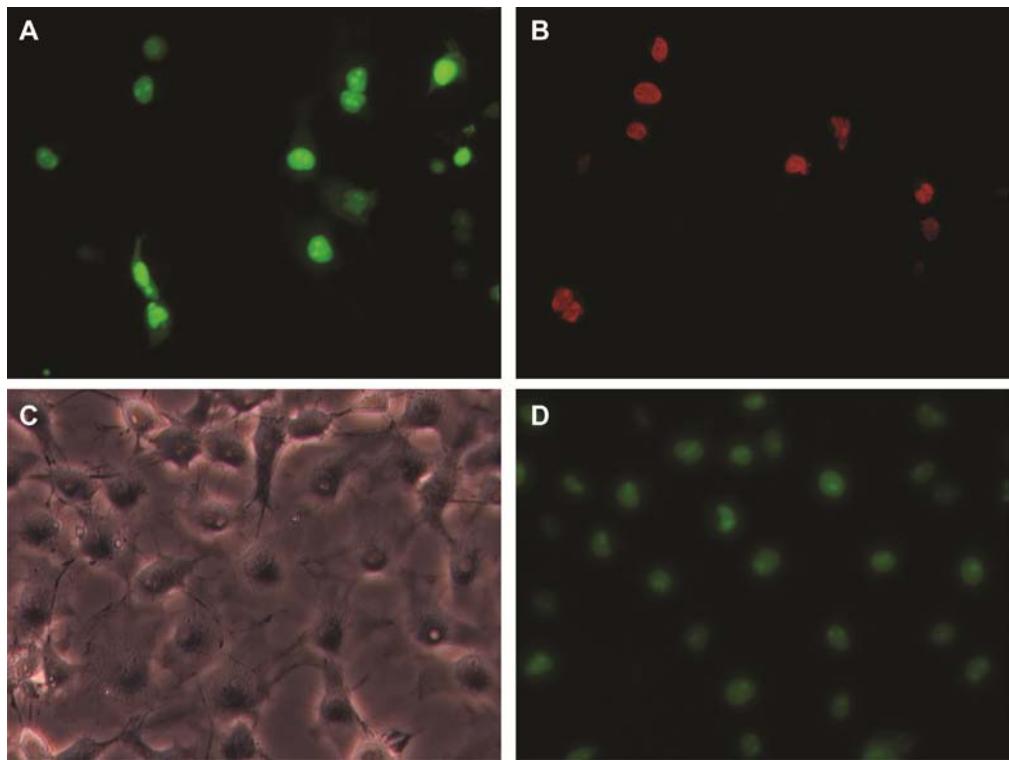
To validate the GFP/Puro<sup>R</sup>-expressing reporter, the FseI to MluI fragment, containing the FLAG-tagged GFP/Puro<sup>R</sup> reporter, was cloned into a *cytomegalovirus* (*CMV*) promoter-driven plasmid. The *CMV*-driven GFP/Puro<sup>R</sup> construct was transfected into COS7 cells using SuperFect (Qiagen). Nuclear-localized eGFP was examined by fluorescence microscopy (**Figure 5.3A**). Immunocytochemistry was used to detect the FLAG-tag using a mouse anti-flag primary antibody and a donkey anti-mouse Cy3 secondary antibody (**Figure 5.3B**). Puromycin resistance was tested by subjecting cells to 1.5 ug/ml puromycin for three weeks (**Figure 5.3C and D**).

### Generation of a CFP-expressing reporter for RMCE

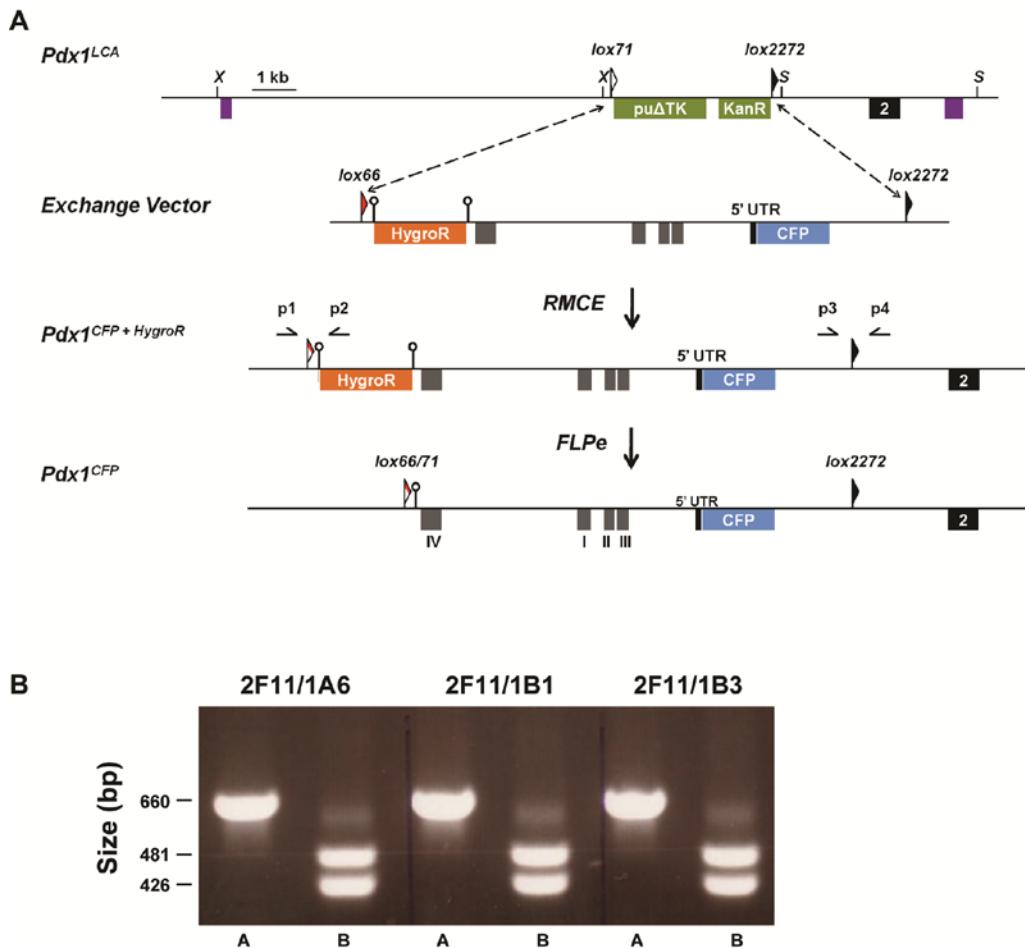
To generate a *Pdx1* reporter expressing a cyan fluorescent protein (CFP), the Pdx1<sup>GFP/PuroR</sup> exchange vector was digested with FseI and MluI to remove the GFP/Puro<sup>R</sup> sequences, and an oligo containing restriction sites FseI, PacI, BbvCI, PmeI, and MluI was inserted to introduce multiple cloning sites. A cyan fluorescent protein, Cerulean, along with a 3X SV40 nuclear localization signal (NLS) was PCR-amplified and cloned into the modified exchange vector. In addition, the rabbit  $\beta$ -globin polyadenylation sequence, containing the splicing region and polyadenylation sequences, was cloned into the plasmid downstream of the nuclear-localized Cerulean. The exchange cassette retained the FRT-flanked *pgk-Hygro*<sup>R</sup> cassette for positive selection of mES cells after RMCE (**Figure 5.4**).



**Figure 5.2 Insertion of GFP/Puro<sup>R</sup> into a *Pdx1* loxed cassette acceptor (LCA) allele by RMCE.** A) Schematic representation of the *Pdx1*<sup>LCA</sup> allele, GFP/Puro<sup>R</sup> exchange vector, and *Pdx1*<sup>GFP/PuroR</sup> allele. Part of the 8.6 kb endogenous *Pdx1* allele, containing the enhancer elements (Areas I-IV) and exon 1, removed in the *Pdx1*<sup>LCA</sup> allele was incorporated in the GFP/Puro<sup>R</sup> exchange vector. The exchange vector included the sequences for *enhanced GFP*, a three-repeat of a SV40 nuclear localization signal, an internal ribosome entry site, and *puromycin resistance* (*Puro*<sup>R</sup>), as well as a *pgk*-driven *hygromycin resistance* (*Hygro*<sup>R</sup>) sequence, flanked by tandem flippase recognition target (FRT) sites, for positive selection following RMCE. Lox66 and lox2272 sites were included to allow recombination with the lox71 and lox2272 sites in the *Pdx1*<sup>LCA</sup> allele. B) PCR analysis of five exchanged ES cell clones following RMCE with targeted clone 1H5. Screening of the 5' and 3' ends reveal all five clones are correctly exchanged by the presence of 660 bp and 481 bp bands; however, the internal screen revealed 1H5/1C4 was a mosaic clone. C) PCR analysis of six exchanged ES cell clones following RMCE with either targeted clone 2G9 or 3E6. Screening of the 5' (lane A) and 3' (lane B) ends reveal all six clones are correctly exchanged.



**Figure 5.3 Validation of the GFP/Puro<sup>R</sup>-expressing construct in COS7 cells.** The GFP/Puro<sup>R</sup>-expressing construct containing a FLAG-tag, enhanced GFP, and puromycin resistance was transfected into COS7 cells. Nuclear-localized eGFP was observed by fluorescent microscopy (A). Presence of the FLAG-tag was determined by immunocytochemistry (B). To assess resistance to puromycin, transfected COS7 cells were subjected to puromycin for three weeks and fluorescent microscopy revealed eGFP expression in surviving cells (C and D).



**Figure 5.4 Insertion of CFP into a *Pdx1* loxed cassette acceptor (LCA) allele by RMCE.**

A) Diagram of the *Pdx1*<sup>LCA</sup> allele, exchange vector, *Pdx1*<sup>CFP+HygroR</sup> allele, and *Pdx1*<sup>CFP</sup> allele. The exchange vector contained lox66 (red triangle) and lox2272 (black triangle) sites flanking a nuclear-localized CFP sequence followed by a rabbit  $\beta$ -globin polyadenylation sequence. A pgk-driven hygromycin resistance (*Hygro*<sup>R</sup>) cassette flanked by FRT sites (open circles) was used as a positive selectable marker during RMCE. Mice containing the *Pdx1*<sup>CFP+HygroR</sup> allele were bred with FLPe-expressing transgenic mice to remove the FRT-flanked *Hygro*<sup>R</sup> cassette. Restriction sites: XmaI (X) and SphI (S). Primer locations: p1, p2, p3 and p4. B) PCR analysis on both the 5' (lane A, p1 and p2) and 3' (lane B, p3 and p4) ends of three *Pdx1*<sup>CFP+HygroR</sup> exchanged clones. Properly exchanged clones were identified by 660 and 481 bp bands on the 5' and 3' ends, respectively.

### Recombinase-mediated cassette exchange for *Pdx1*

To generate the *Pdx1*<sup>GFP/PuroR</sup> reporter, two electroporations were performed in which 40 µg of *Pdx1*<sup>GFP/PuroR</sup> and 40 µg pBS185, a Cre recombinase expression plasmid, were co-electroporated into 5.0 x 10<sup>6</sup> TL1 mES cells that contained the *Pdx1*<sup>LCA</sup> allele (clone 1H5). Selection with 200 µg/ml hygromycin B (Invitrogen) began 48 hours post-electroporation and continued for seven days. 107 surviving colonies were isolated, grown on MEFs in 96-well gelatinized tissue culture plates, and fed complete ES cell medium without hygromycin B for 4-5 days. Cells were then split to two 96-well plates: one for a master plate, which was frozen 3 days later, and one for negative selection using 8 µM gancyclovir (Sigma, St. Louis, MO) for 3-4 days. Following gancyclovir selection, 25 surviving clones were expanded for DNA isolation using proteinaseK (1mg/ml) incubation and phenol/chloroform extraction. DNA was precipitated using 0.3 M sodium acetate in ethanol, and clones surviving selection with both hygromycin and gancyclovir were analyzed for cassette exchange by PCR using primers that spanned either the lox66/71 site on the 5' end or the lox2272 site on the 3' end. On the 5' end, the combination of 5'-TGAGATTGTATATTGCGGTGCA and 5'-ACGAGACTAGTGAGACGTGCTACT primers resulted in a band size of 660 bp after RMCE and on the 3' end, use of 5'-TGAGCAATTCCAAGCAGCTGGA and 5'-ACCTTGCAGTCCTCTGAAGT primers resulted in a 481 bp band for the exchanged allele and a 426 bp band in the wild type allele. Seventeen clones were properly exchanged, and four clones (1H5/3A2, 1H5/1B6, 1H5/1F2, and 1H5/2G5) were expanded (**Figure 5.2**). Clones 1H5/3A2 and 1H5/1F2 were injected into blastocysts but did not display germline competence.

Recombinase-mediated cassette exchange was repeated to generate the *Pdx1*<sup>GFP/PuroR</sup> reporter using *Pdx1*<sup>LCA</sup> clones obtained from the second targeting experiment. For the first repeat, parental clone 3E6 was used and 40 µg of *Pdx1*<sup>GFP/PuroR</sup> and 40 µg pBS185 were co-

electroporated into  $5.0 \times 10^6$  mES cells. A staggered positive-negative selection strategy using hygromycin B (200 µg/ml) then gancyclovir (8 µM) identified 48 hygromycin-resistant and 12 gancyclovir-resistant clones. DNA was extracted from six clones and PCR screening was performed to identify positive exchanges. All six clones were properly exchanged, and three clones (3E6/2A2, 3E6/2B5 and 3E6/2C12) were expanded (**Figure 5.2, C**). All three clones were injected into blastocysts with only 3E6/2A2 displaying germline competence. In parallel, a second repeat was performed where parental clone 2G9 was used and 40 µg of Pdx1<sup>GFP/PuroR</sup> and 40 µg pBS185 were co-electroporated into  $5.0 \times 10^6$  mES cells (passage 24). A staggered positive-negative selection strategy identified 60 hygromycin-resistant and 18 gancyclovir-resistant clones. DNA was extracted from six clones and PCR screening was performed to identify positive exchanges. All six clones were properly exchanged, and three clones (2G9/1A4, 2G9/1B8 and 2G9/1D9) were expanded (**Figure 5.2C**). Clone 2G9/1A4 was injected into blastocysts but did not display germline competence.

For the generation of a CFP-expressing reporter for *Pdx1*, RMCE was performed using clone 2F11 from the second targeting experiment, and 40 µg of Pdx1<sup>CFP</sup> and 40 µg pBS185 were co-electroporated into  $5.0 \times 10^6$  mES cells (passage 24). A staggered positive-negative selection strategy identified 35 hygromycin-resistant and 8 gancyclovir-resistant clones. DNA was extracted from six clones and PCR screening was performed to identify positive exchanges. All six clones were properly exchanged, and three clones (2F11/1A6, 2F11/1B1 and 2F11/1B3) were expanded (**Figure 5.4**). Karyotyping was performed for 2F11/1B1 and 2F11/1A6 and revealed they were 71% and 68% normal, respectively. Clones 2F11/1B1 and 2F11/1A6 were injected into E3.5 C57BL/6 blastocysts and both displayed germline competence. The FRT-flanked *Hygro<sup>R</sup>* sequence was removed by inbreeding with *Tg(ACFLPe)9205Dym* mice, and the resulting *Pdx1<sup>CFP</sup>* allele was maintained on an outbred background. Embryos were isolated from wild type CD-1 females crossed with *Pdx1<sup>CFP/+</sup>*

mice where the presence of vaginal plug at noon was considered E0.5. Experimental protocols were approved by the Vanderbilt Institutional Animal Care and Use Committee.

### **Immunolabeling and imaging**

Immunolabeling was performed as previously reported (Burlison *et al.*, 2008). Primary antibodies were diluted in 1% BSA in 1X PBS as follows: guinea pig anti-Pdx1 (C.V.E. Wright), 1:2500; chicken anti-GFP (Invitrogen), 1:1000; goat anti-amylase (Santa Cruz Biotechnology), 1:1000; guinea pig anti-insulin (Linco), 1:1000. Secondary antibodies were diluted in 1% BSA in 1X PBS as follows: donkey anti-chicken DyLight 488 (Jackson ImmunoResearch), 1:500; donkey anti-guinea pig Cy3 (Jackson ImmunoResearch), 1:1000; donkey anti-goat Alexa Fluor 647 (Invitrogen), 1:1000. Sections were counterstained with DAPI and cover slips mounted using Aqua/Poly Mount (Polysciences, Inc.). Images were acquired using an Axioplan2 microscope (Zeiss) with a QImaging RETIGA EXi camera. Whole mount imaging of *Pdx1*<sup>CFP</sup> embryos was performed using a Leica MZ 16 FA stereoscope with a QImaging RETIGA 4000R camera.

### **Fluorescence-activated cell sorting**

*Pdx1*<sup>CFP/+</sup> embryos were identified by direct fluorescence and dissected tissues were dissociated using Accumax (Sigma). Following filtration through a 35 µm cell strainer (BD Biosciences) and centrifugation, cells were resuspended in FACS medium [L15 medium (Invitrogen) containing 1 mg/ml BSA, 10 mM HEPES pH 7.4, 1% penicillin/streptomycin, and 37.5 ng/ml DNase I] or FACS medium with 7AAD (Invitrogen). Cells were analyzed and isolated using an Aria III (BD Biosciences). CFP was excited using either a 405 nm or 445 nm laser and emission detected with a 470 nm long pass and 510/80 bandpass filter.

## **RNA Isolation and semi-quantitative RT-PCR**

Total RNA was isolated using TRIzol LS (Invitrogen), DNase-treated (Ambion), and column-purified (Zymo Research). Total RNA was reverse transcribed (High Capacity cDNA Archive kit; ABI), and PCR was performed using 1 ng of cDNA template. *Ptf1a* was detected using the following primers: 5'-CGAATTGCCACGGATCACT and 5'-CCCGGAAGGACGAATGG. *Insulin* was detected using the following primers: 5'-CCACCCAGGCTTTGTCAAA and 5'-CCCAGCTCCAGTTGTTCCAC.

## **Results**

### **Generation of a *Pdx1*<sup>LCA</sup> allele**

BAC recombineering and gene targeting were used to generate mES cells with a *Pdx1*<sup>LCA</sup> allele, as shown in **Figure 5.1**. In this allele, an 8.6 kb region of the *Pdx1* locus, containing four previously characterized conserved regulatory regions termed Areas I – IV (Gittes, 2009; Pan and Wright, 2011), as well as exon 1, was replaced with a dual positive-negative selection cassette flanked by lox71 and lox2272 sites (Araki et al., 2002; Chen et al., 2011). Southern blot analysis using probes on both the 5' and 3' ends of the *Pdx1* locus confirmed the desired homologous recombination events (**Figure 5.1**).

### **Derivation of a *Pdx1*<sup>GFP/PuroR</sup> allele by RMCE**

To validate the functionality of the *Pdx1*<sup>LCA</sup> allele and generate a *Pdx1* reporter allele, we next made an exchange vector that replaced *Pdx1* coding sequences in exon one with a bicistronic cassette containing a FLAG-tagged and a nuclear-localized enhanced GFP (eGFP) (Rizzo et al., 2004) followed by an internal ribosome entry site (IRES) to initiate translation of puromycin resistance, thereby generating a *Pdx1*-null allele (**Figure 5.2**). In addition, the

exchange vector contains a *pgk*-driven *hygromycin resistance (Hygro<sup>R</sup>)* cassette, flanked by tandem flippase recognition target (FRT) sites. Prior to RMCE, the GFP/Puro<sup>R</sup>-expressing reporter was validated by cloning the FLAG-tagged GFP/Puro<sup>R</sup> reporter into a *cytomegalovirus (CMV)* promoter-driven plasmid. Following transfection into COS7 cells, nuclear-localized eGFP was detected by fluorescence microscopy (**Figure 5.3, A**), and the cells exhibited immunolabeling with an anti-FLAG antibody (**Figure 5.3, B**). Additionally, cells survived culture with puromycin and exhibited GFP fluorescence (**Figure 5.3, C and D**). Taken together, these results confirmed the functionality of the cassette.

RMCE of the GFP/Puro<sup>R</sup>-expressing reporter into the *Pdx1<sup>LCA</sup>* was achieved through the use of a staggered positive-negative selection strategy (Long et al., 2004) after co-electroporation of the exchange vector and a Cre-expression plasmid. Chimeric mice were generated by injection of an exchanged clone (**Figure 5.2**) into E3.5 mouse blastocysts. After germline transmission, mice containing the *Pdx1<sup>GFP/PuroR+HygroR</sup>* allele were bred with FLP-expressing transgenic mice to remove the FRT-flanked *Hygro<sup>R</sup>* cassette, thereby generating the *Pdx1<sup>GFP/PuroR</sup>* allele. Although the preliminary data suggested that the GFP/Puro<sup>R</sup>-expressing reporter was functional, the targeted mice did not exhibit GFP fluorescence in the expected *Pdx1* expression domains (data not shown). While the exact cause for the lack of GFP fluorescence is unknown, we suspect that the IRES sequence resulted in the destabilization of the protein and diminished fluorescence (Mizuguchi et al., 2000).

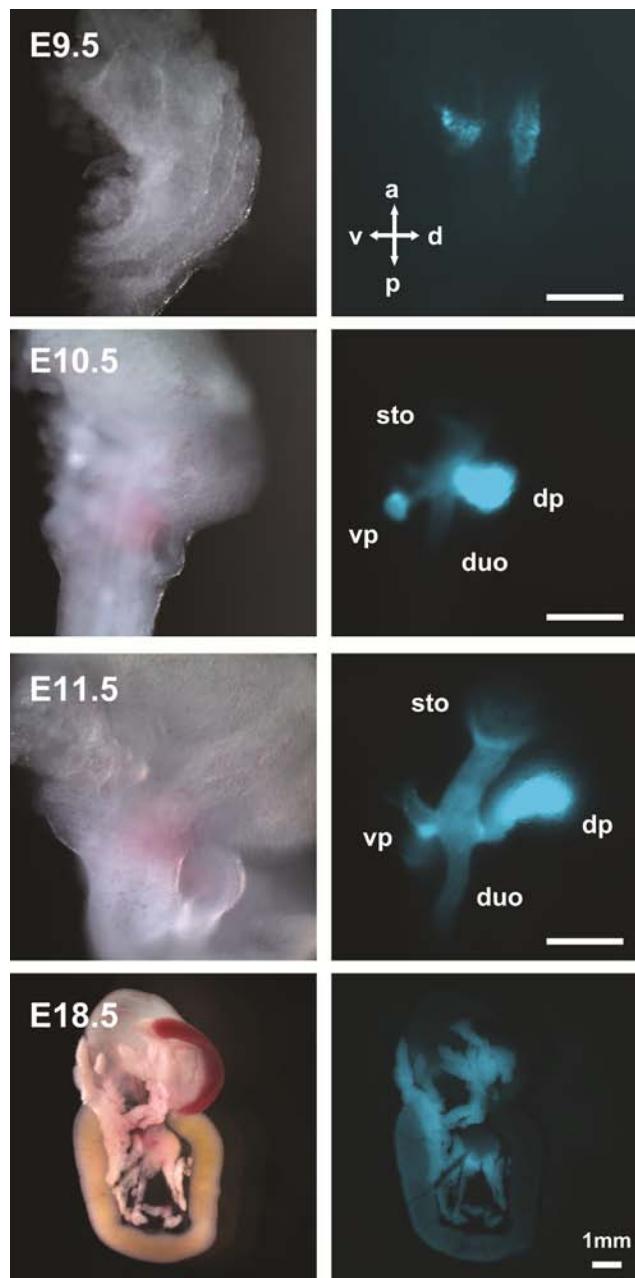
### Derivation of a *Pdx1<sup>CFP</sup>* allele by RMCE

Given the lack of fluorescence observed from the *Pdx1<sup>GFP/PuroR</sup>* allele, we sought to generate another *Pdx1* fluorescent reporter allele. We generated an exchange vector that replaced coding sequences in exon one of the *Pdx1* gene with a nuclear-localized CFP (Cerulean) (Rizzo et al., 2004), thereby generating a *Pdx1*-null allele (**Figure 5.4**). A portion

of the rabbit  $\beta$ -globin gene, containing both intronic and polyadenylation sequences, was placed downstream of the CFP coding sequences (Chen et al., 2011; Westwood et al., 1993). In addition, the exchange vector contains a *pgk-hygromycin resistance (Hygro<sup>R</sup>)* cassette, flanked by tandem flipase recognition target (FRT) sites. RMCE into the *Pdx1<sup>LCA</sup>* was achieved and chimeric mice were generated similar to the methods described above. After germline transmission, mice containing the *Pdx1<sup>CFP+HygroR</sup>* allele were bred with FLPe-expressing transgenic mice to remove the FRT-flanked *Hygro<sup>R</sup>* cassette, thereby generating the *Pdx1<sup>CFP</sup>* allele. While similar in nature to the *Pdx1<sup>GFP/PuroR</sup>* allele, the *Pdx1<sup>CFP</sup>* allele is more advantageous because the excitation and emission profiles for CFP are spectrally distinct from both GFP and YFP, thus permitting the combinatorial use of this allele with GFP- and YFP-expressing reporter alleles.

### **Expression pattern of *Pdx1<sup>CFP</sup>* by whole mount fluorescence microscopy**

To analyze expression of CFP, we first utilized whole mount fluorescence microscopy (**Figure 5.5**). In *Pdx1<sup>CFP/+</sup>* embryos, CFP expression was easily observable beginning at E9.5 in both the dorsal and ventral endoderm, consistent with the pattern previously determined from a *Pdx1<sup>lacZ</sup>* insertion allele (Offield et al., 1996). Between E10.5 to E11.5, CFP expression in both the dorsal and ventral pancreatic buds was brighter than that of the caudal stomach and duodenum, normal domains of *Pdx1* expression. During this developmental stage, the expression domain of *Pdx1* contains a subpopulation of pancreatic MPCs (Gu et al., 2002; Pan and Wright, 2011). As development proceeds, CFP expression persisted at a high level in the pancreatic epithelium and at a lower level in the caudal stomach and duodenum.



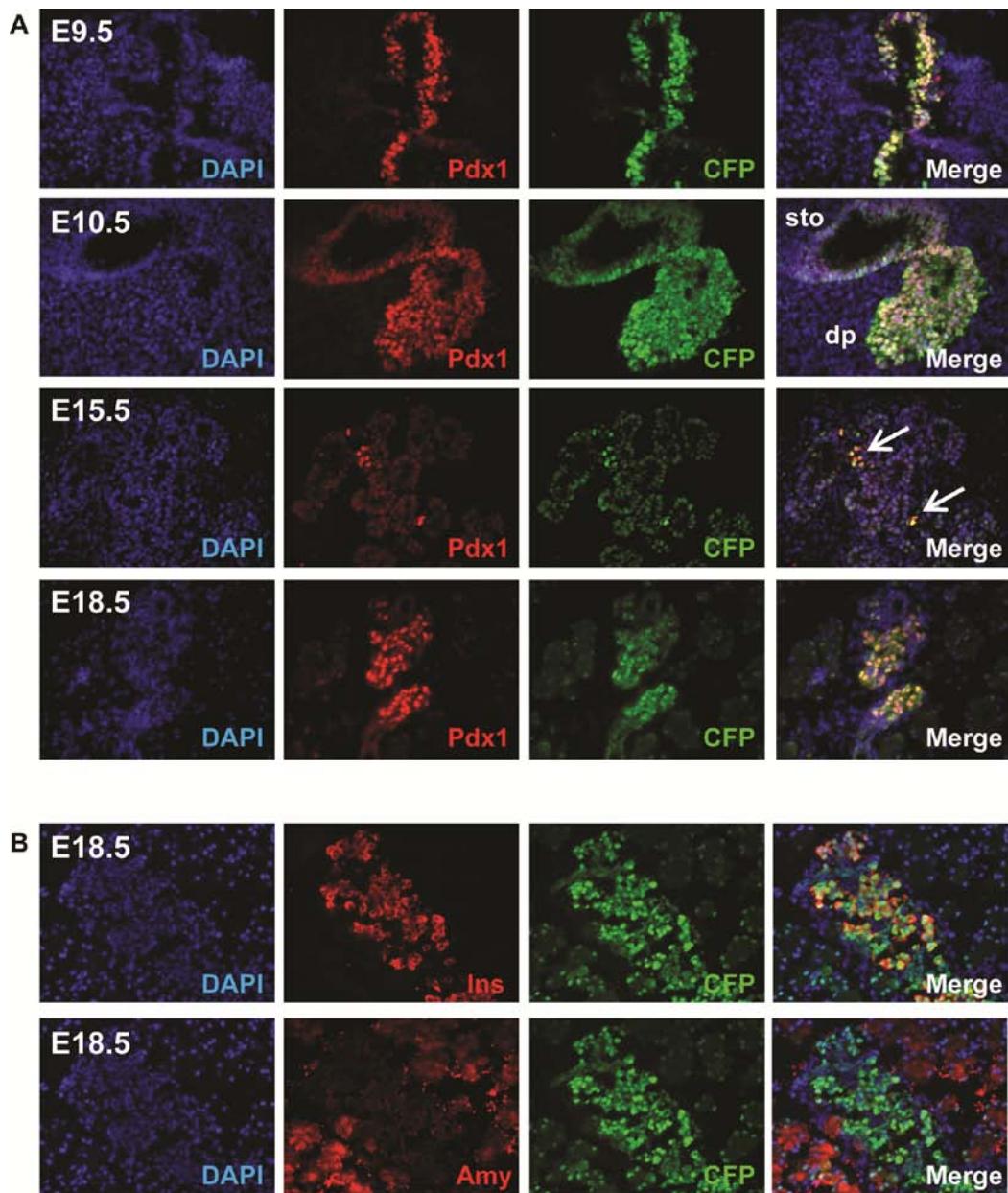
**Figure 5.5 Cerulean fluorescence in *Pdx1*<sup>CFP/+</sup> embryos.** CFP fluorescence was observed in dissected *Pdx1*<sup>CFP/+</sup> embryos as early as E9.5 in the dorsal and ventral endoderm, consistent with previous reports detailing the expression pattern of *Pdx1*. From E10.5 to E11.5, CFP expression was detected in the dorsal and ventral pancreatic buds at a higher intensity as compared to the expression visualized in the caudal stomach and duodenum. Throughout development, CFP expression persisted at higher levels in the pancreatic epithelium than in the caudal stomach and duodenum. Scale bar = 250  $\mu$ m unless otherwise noted. Dorsal pancreas (dp), ventral pancreas (vp), stomach (sto), duodenum (duo), anterior (a), posterior (p), dorsal (d), ventral (v).

### **Expression pattern of *Pdx1*<sup>CFP</sup> by immunohistochemical analysis**

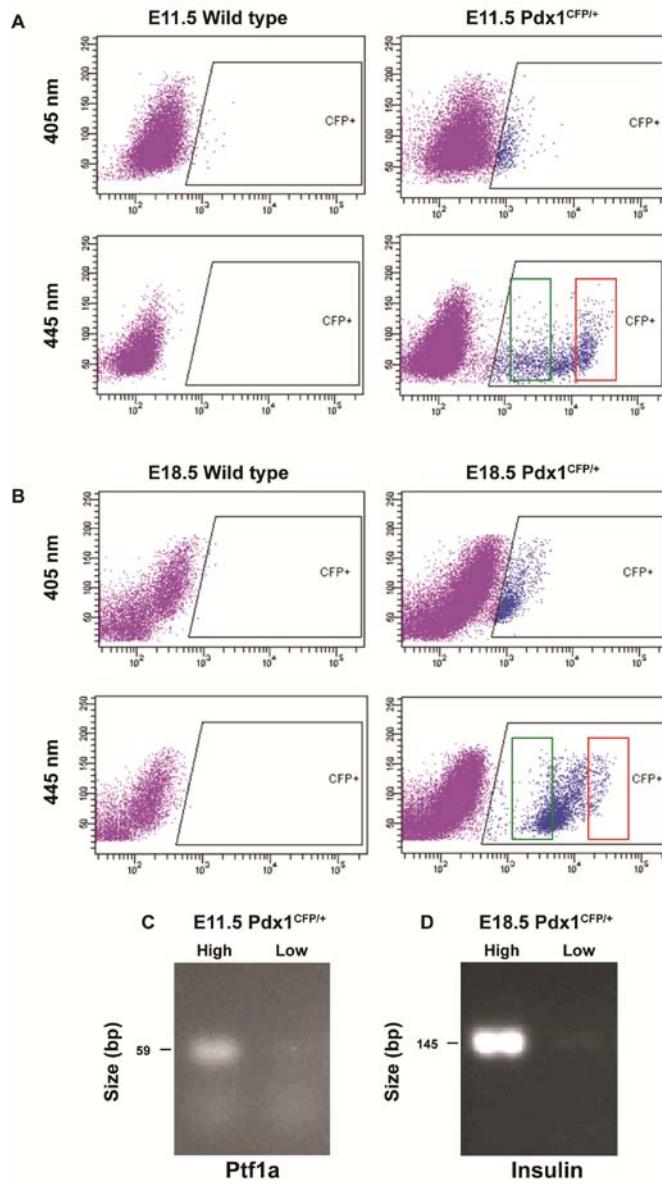
To determine whether CFP and Pdx1 were co-expressed, we performed immunohistochemical analysis using tissues from multiple developmental stages (**Figure 5.6** and data not shown). Both Pdx1 and CFP were detected in the dorsal and ventral endodermal evaginations of the posterior foregut at E9.5. From E10.5 to E11.5, Pdx1 and CFP were detected throughout the pancreatic epithelium, and expression was also observed in the stomach and duodenum. From E12.5 – E14.5, CFP immunofluorescence co-localized with Pdx1 throughout the branching epithelium, and by E15.5, high levels of Pdx1 and CFP expression were evident in periodic clusters throughout the epithelium with lower levels throughout the epithelium. At E18.5, high levels of Pdx1<sup>CFP</sup> expression were restricted primarily to the developing β cells as evident by the co-localization with insulin-expressing cells (**Figure 5.6B**). Additionally, lower levels of CFP expression were co-localized with amylase, indicative of Pdx1 expression in the acinar cells, again in accordance with previous reports (**Figure 5.6B**) (Guz et al., 1995; Wu et al., 1997). At each developmental stage, expression the *Pdx1*<sup>CFP</sup> allele paralleled that of endogenous Pdx1 expression, consistent with the *Pdx1*<sup>CFP</sup> allele faithfully recapitulating expression of *Pdx1* (Guz et al., 1995; Leonard et al., 1993; Miller et al., 1994; Offield et al., 1996; Ohlsson et al., 1993). While perdurance of certain FP reporter alleles has been reported (Burlison et al., 2008; Viotti et al., 2011), this was not observed at any of the time points examined in this study which may simply reflect the maintained expression of Pdx1 in numerous cell types of the pancreatic epithelium.

### **FACS analysis of *Pdx1*<sup>CFP</sup> cells**

Given that the optimal excitation wavelength for Cerulean is 433 nm (Rizzo et al., 2004), we determined the effect of using either a 405 nm or 445 nm laser, both of which are commercially available, for sorting *Pdx1*<sup>CFP</sup>-expressing cells. As shown in **Figure 5.7**, the



**Figure 5.6 Immunolabeling in  $Pdx1^{CFP/+}$  embryos.** A) Co-expression of Pdx1 and CFP at E9.5 throughout the dorsal and ventral endoderm. At E10.5, Pdx1 and CFP expression are co-localized in the pancreatic epithelium with expression also observed in other posterior foregut derivatives, such as the stomach. By E15.5, high levels of Pdx1 and CFP were evident in scattered clusters (arrows) throughout the epithelium with lower levels throughout the epithelium. At E18.5, Pdx1 and CFP are co-localized in cells displaying both high and low levels of expression. Stomach (sto), dorsal pancreas (dp). DAPI: nuclear counterstain. B) At E18.5,  $Pdx1^{CFP}$  expression is restricted primarily to developing  $\beta$  cells, as indicated by co-expression with insulin, with lower levels of CFP expression observed in the acinar cells, as indicated by co-expression with amylase.



**5.7 FACS analysis of *Pdx1*<sup>CFP/+</sup> embryos.** The emission intensity of CFP following excitation with either 405 nm or 445 nm was compared. E11.5 embryos (A) and E18.5 embryos (B) were used to isolate CFP-positive cells. The fluorescence intensity observed following excitation with 405 nm laser (top panels) is minimally distinguishable above cellular autofluorescence seen in the wild type embryos (left panels); whereas a 445 nm laser (bottom panels) provided more optimal excitation of CFP. Both high- and low-expressing cells, indicated by red and green boxes, respectively, were isolated from both E11.5 embryos (A) and E18.5 embryos (B). (C) A higher level of *Ptfla* gene expression was detected in cells showing high- versus low-intensity CFP fluorescence consistent with their origin in the pancreatic epithelium. (D) Similarly, a higher level of *Insulin* gene expression was observed in high- versus low-intensity CFP fluorescence cells at E18.5 consistent with them being pancreatic  $\beta$  cells.

445 nm laser resulted in a broader range of emission intensity using embryonic tissues at E11.5 and E18.5. Cyan fluorescence observed using the 405 nm laser was minimally distinguishable above cellular autofluorescence in wild type embryos. However, CFP-expressing cells excited by the 445 nm laser displayed approximately 50-fold higher fluorescence intensity. At E11.5, both high- and low-intensity CFP fluorescence was observed by FACS (**Figure 5.7A**), which is similar to the high- and low-expression patterns evident by immunolabeling in the pancreatic epithelium and the stomach/duodenal epithelium, respectively. Cells exhibiting higher levels of CFP fluorescence primarily represented cells of the pancreatic epithelium as indicated by *Ptf1a* expression (**Figure 5.7C**). In addition, high- and low-intensity CFP fluorescence was observed at E18.5 (**Figure 5.7B**). An analysis of *Insulin* expression revealed that cells displaying high levels of CFP fluorescence at E18.5 were predominantly pancreatic  $\beta$  cells (**Figure 5.7D**).

## Discussion

An analysis of purified native cell populations from the various intermediates transitioned through during pancreas development (e.g. definitive endoderm, posterior foregut, pancreatic MPC, endocrine progenitor, etc.) may help to identify transcriptional networks and signaling mechanisms that are critical for the growth and maturation of the final organ. Similar to the isolation and transcriptional profiling of *Ptf1a*-expressing cells (discussed in Chapter III), by utilizing the *Pdx1<sup>CFP</sup>*-expressing mice, we are poised to isolate specific cell populations, such as the posterior foregut endoderm, that will be useful for further understanding specific developmental stages within the context of the  $\beta$  cell developmental pathway. While valuable in and of itself, the transcriptional profile is most useful when assessed within the context of the sequential development of the pancreatic  $\beta$  cell (**Figure 1.7**). When analyzed within the context of numerous developmental

intermediates, the expression profile of the single cell population becomes much more valuable and will provide a context in which to discover transcripts and mechanisms that may be critical during numerous steps of  $\beta$  cell development (discussed further in Chapter VI).

In addition to identifying and isolating cell populations based on the expression of a single transcriptional marker, the use of combinatorial cell sorting will prove useful for identifying and isolating specific subsets of cells or cells that are spatially-defined. These combinatorial comparisons will allow for the identification of transcripts whose expression is restricted spatially or more precisely defines specific cell populations. By utilizing the  $Pdx1^{CFP}$  and  $Ptf1a^{YFP}$  alleles, we will be able to specifically discriminate between pancreatic and non-pancreatic foregut endothelial cells which will enable us to identify transcripts that are specifically expressed within the developing pancreatic MPCs (discussed further in Chapter VI). Furthermore, it will allow for the identification of signaling pathways that promote the growth and maturation of the pancreatic progenitors explicitly. The identification of such signaling mechanisms will help to further focus hES cell directed differentiation efforts in directing the differentiation of the posterior foregut endoderm-like cells specifically towards pancreatic fates and not other foregut derivatives.

## CHAPTER VI

### CONCLUSIONS AND FUTURE DIRECTIONS

#### Synopsis

Progress in learning how to direct the differentiation of hES cells to pancreatic fates rests upon a foundation of knowledge gained over the past two decades concerning the mechanisms of pancreas development, and most of this information has been gained from studies in mice. However, as our understanding of pancreas development in the mouse has expanded, it has become apparent that greater efforts must be made to translate this information into new human therapies.

Numerous prior studies have shown the central importance of Pdx1 and Ptf1a in pancreas development (Jonsson et al., 1994; Kawaguchi et al., 2002; Krapp et al., 1998; Offield et al., 1996). Specifically, Ptf1a is a valuable marker of the pancreatic multipotent progenitor cell population within the pancreatic epithelium, and thus many studies have sought to understand how Ptf1a functions during development. During early organogenesis, PTF1 is required for the growth and branching of the epithelium and is composed of Ptf1a, a class A bHLH protein and Rbpj (PTF1-J) (Rose et al., 2001; Roux et al., 1989). At the secondary transition, Ptf1a becomes restricted to the MPCs in the branching tips, and its continued expression is required for the further development of the late multipotent progenitor cells (MPCs). At approximately E13, Rbpl replaces Rbpj in the PTF1 complex (Beres et al., 2006) of pancreatic MPCs that commit to acinar cell fate. This transition to a PTF1-L complex completes acinar differentiation and maintains the acinar phenotype in adult pancreas (Masui et al., 2010).

To explore changes in the gene expression profile of *Ptf1a*-expressing cells, we performed RNA-Seq using a temporal and genetic approach enabled by FACS isolation of three distinct cell populations marked by the *Ptf1a*<sup>YFP</sup> allele (Burlison et al., 2008). To analyze the temporal changes that occur as the *Ptf1a*-expressing pancreatic MPCs transition to a unipotent, committed acinar cell fate, the transcriptional profiles of *Ptf1a*<sup>YFP/+</sup> cells at E11.5 and E15.5 were compared. Additionally, to examine the genetic changes between *Ptf1a*-expressing MPCs and *Ptf1a*-deficient progenitor cells, *Ptf1a*<sup>YFP/YFP</sup> cells, which lack *Ptf1a*, and *Ptf1a*<sup>YFP/+</sup> cells were profiled at E11.5 (**Figure 3.1**). These specific developmental timepoints were chosen because at E11.5 *Ptf1a* is broadly expressed throughout the expanding dorsal and ventral pancreatic epithelium marking the MPC population, while at E15.5 *Ptf1a* expression has become restricted to the acinar lineage (Burlison et al., 2008; Masui et al., 2007). A temporal comparison of the E11.5 *Ptf1a*<sup>YFP/+</sup> and E15.5 *Ptf1a*<sup>YFP/+</sup> profiles identified 2,136 genes whose expression was increased at least 5-fold and another 657 genes that were decreased by a similar amount. Among the upregulated genes were *amylase 2a5* (*Amy2a5*), *chymotrypsin-like elastase family, member 1* (*Cela1/Ela1*), *chymotrypsinogen B* (*Ctrb*), and *carboxypeptidase A1* (*Cpa1*), all of which are known targets for PTF1-L (Beres et al., 2006). Conversely, *neurogenin 3* (*Neurog3*), *insulinoma 1* (*Insm1*), and *glucagon* (*Gcg*) were expressed in the *Ptf1a*-expressing MPCs at E11.5 but not detected in *Ptf1a*-expressing cells at E15.5. Similarly, by comparing the transcriptional profile of the *Ptf1a*-expressing MPCs at E11.5 with those from the *Ptf1a*-null embryos of the same age, we found 997 genes that were expressed at levels exceeding 5-fold in cells in *Ptf1a*-expressing MPCs at E11.5 and 1,105 genes that had the opposite pattern. Interestingly, the expression of genes such as *glucagon* (*Gcg*), *secretin* (*Sct*), *v-maf musculoaponeurotic fibrosarcoma oncogene family, protein B* (*Mafb*), and *synaptophysin* (*Syp*) were more strongly expressed in the *Ptf1a*-deficient progenitor cells, a result that agrees with publications suggesting that

*Ptf1a* acts to prevent the premature differentiation of pancreatic MPCs towards endocrine or intestinal cell fates (Kawaguchi et al., 2002).

Following transcriptional profiling, a deeper analysis of the expression datasets from the various developmental intermediates proves useful for identifying transcription factors and signaling pathways that have undiscovered roles in pancreas progenitor cell specification and pancreatic organogenesis. An analysis of the transcriptional profiles for *Ptf1a*<sup>YFP</sup>-expressing cells as determined by RNA-Seq and qRT-PCR revealed an upregulation of *Nepn* in *Ptf1a*-expressing MPCs at E10.5 and E11.5 as compared to acinar-specified cells at E15.5 and later (**Figure 4.1**). Given these results, we sought to further explore the expression pattern of *Nepn* during embryogenesis, and thus generated mice containing a *Nepn-Cherry* reporter in the *ROSA26* locus which revealed the expression of *ROSA26*<sup>*Nepn-Cherry*</sup> in a subpopulation of pancreatic MPCs from E10.5 to E11.5 (**Figure 4.5 and 4.6**). Given these results and that *Nepn* functions to inhibit TGF $\beta$  signaling (Mochida et al., 2006), further studies using either a global or conditional gene knockout would allow for a better assessment of the temporal requirement for *Nepn* in the developing endoderm and permit further elucidation of the potential role of *Nepn* in pancreatic progenitor cell specification.

In order to generate new reporter alleles, our preference is to perform gene targeting since this method inherently assures that necessary regulatory sequences for proper expression of a fluorescent protein reporter are utilized. This may be important since it is likely that the enhancer sequences for certain genes have not been identified, as is the case with *Nepn*. However, we are also acutely aware that the gene targeting approach is time consuming, especially as it pertains to gene targeting in hES cells, which has proven more difficult than targeting in mES cells. For this reason, generating an optimal reporter line may not always be worth the time or effort, especially when one considers that FP reporter lines generated in this manner are typically rendered haploinsufficient at the locus used to drive

reporter expression. For this reason, our second choice has been to utilize RMCE to place FP reporters driven by 8 – 12 kb fragments of upstream regulatory sequence into the mouse or human *ROSA26* gene locus, as has been done for the generation of the *Nepn-Cherry* transgene (**Figure 4.3**) and the *Pdx1-CFP* reporter human ES cell line (**Figure 2.4**). While not without certain limitations, these reporter transgenes permit identification of genetically-tagged cells. For instance, hES cells containing the *Pdx1-CFP* transgene were used during directed pancreatic differentiation *in vitro*, and expression of the reporter gene can be observed as hES cells differentiated into posterior foregut and pancreatic endoderm-like cell fates (**Figure 2.4**) which provides compelling preliminary data that it will be possible to isolate pancreatic MPCs that have been induced to express *Pdx1* and thus may represent early pancreatic endoderm.

One objective of our future studies is to utilize hES cell lines in which key developmental regulatory genes are marked by distinct FP reporters to generate transcriptional profiles that can be compared to the profiles established for *in vivo* isolated cell populations during mouse development, similar to the studies discussed in Chapter III. However, while the availability of such hES cell lines will enable purification and characterization of specific cellular populations from impure populations and facilitate the development of more robust directed differentiation protocols capable of generating functional human  $\beta$  cells, the generation of reporter alleles via RMCE into the *ROSA26<sup>LCA</sup>* is not currently feasible. In response to Executive Order 13505, issued on March 9, 2009, the NIH established new guidelines under which they will fund research pertaining to the area of hES cells. Based on these guidelines, each hES cell line must adhere to the new policy standards and only eligible hES cell lines may be used for NIH-supported research. The hES2 hES cell line, which is currently the only cell line in which the targeted *ROSA26<sup>LCA</sup>* allele

exists, has not been federally approved based on these new guidelines; therefore NIH-supported research using these cells is not permitted.

### **Transcriptional profiling of pancreatic developmental intermediates**

During mouse development, endocrine cells form as the result of a sequential process that is best distinguished by a combination of marker genes that are direct transcriptional effectors of the developmental program. The transcriptional profiling of the *Ptf1a*-expressing cells documented here represents only a small portion of a larger developmental scheme. We have critically assessed and identified cell populations that are likely to yield the greatest understanding of the lineage from definitive endoderm to mature  $\beta$  cells, which in turn will be used to mirror the cellular stages through which hES cells must transit to become  $\beta$  cells. As shown in **Figure 1.7**, these populations span a developmental window ranging from E8.0 (definitive endoderm) to P60 (mature pancreatic  $\beta$  cells), but primarily consist of pancreatic progenitor cell populations that will be isolated from dissected mouse embryonic tissues. To characterize these developmental intermediates, we are using specific combinations of fluorescently-tagged mouse alleles to obtain highly defined FACS-purified cell populations. Similar to the profiling performed for the *Ptf1a*-expressing cells, we will obtain a detailed RNA expression profile for each specific purified cell population through RNA-Sequencing, as well as transcription factor-promoter binding data through ChIP-Seq. These data will be used to precisely ascertain changes in gene expression at different times and in very discrete populations of cells in the developing mouse pancreas, including those formed in mutant backgrounds, such as the *Ptf1a*<sup>YFP/YFP</sup> progenitor cells reported here. By comparing different RNA profiles using RNA-Seq and protein/DNA interactions using ChIP-Seq, we hope to identify gene clusters essential for the formation of pancreatic  $\beta$  cells from normal and

genetically modified pathways. Identification of transcription factors and cell surface receptors in these cell populations will be useful for further dissecting the subpopulations of pancreatic MPCs *in vivo*. Parallels and analogies can then be drawn for identifying and isolating human MPC-like cells and for distinguishing different progenitor cells during the appropriate and inappropriate directed differentiation of hES cells towards a  $\beta$  cell fate.

Our approach to obtain the transcriptional profiles of these developmental intermediates has many advantages over prior studies (Chiang and Melton, 2003; Gu et al., 2004; White et al., 2008). First, by utilizing fluorescently-tagged alleles, we are able to isolate genetically defined cell populations. This is a major improvement upon previous studies that often relied on manual dissection of pancreatic regions resulting in populations that are inescapably contaminated by other cell types. While numerous researchers have sought to characterize the transcriptome of specific cell populations present during pancreatic organogenesis during mouse development, one of the major drawbacks has been the inability to identify precisely and isolate, with minimal contamination, distinct cell populations marking specific developmental intermediates. The method of manually dissecting desired tissues has been one of the major contributors to contamination by other cell types, such as the mesoderm or mesenchyme. Therefore, the use of fluorescently-tagged reporter alleles permits the isolation of a cell population of interest by FACS without risking contamination of the cell population by undesired tissues. However, one consideration to take into account with fluorescently-tagged alleles is fluorescent protein perdurance. For example, the perdurance of a *Neurog3-eGFP* transgene led to the labeling of both Neurog3-positive endocrine progenitors and their hormone-positive progeny (Sugiyama et al., 2007). Therefore, when identifying a cell population to isolate, it is pertinent to examine whether the fluorescent reporter faithfully mimics the gene's endogenous expression. Second, by utilizing the recently developed RNA-Sequencing approach for whole transcriptome profiling, instead

of DNA microarrays, we obtain a less biased and more quantitative view of the gene regulatory changes occurring (Marioni et al., 2008; Mortazavi et al., 2008). Third, by collecting information from closely related cell types, we will be able to more accurately assess the spatial and temporal alteration of genetic networks that regulate the step-wise production of  $\beta$  cells, which was not achieved by previous efforts (Gu et al., 2004; White et al., 2008). Finally, as part of the future directions for the transcriptional profiling of the various developmental intermediates, we plan to utilize ChIP-Seq to directly define binding of transcription factors, such as Neurog3, Nkx2-2, and Pax4, to regulatory sites in chromatin, which will reveal how these factors regulate gene expression in cell type specific manners for proper cell differentiation and function.

In addition to analyzing the transcriptional profile of these developmental intermediates, the fluorescently-tagged reporter alleles permit FACS isolation of distinct cell populations which can then be used in cell transplantation assays to assess the lineage potential of specific purified progenitor cell populations. Previously, it has been shown that the injection of mouse Pdx1-positive progenitor cells into E12.5 dorsal mouse pancreas, and subsequent culturing of these cells *in vitro*, results in proliferation and differentiation of the cells into endocrine cells, including those of the  $\beta$  cell lineage (Xu et al., 2008). Similarly, studies have demonstrated that hES cell-derived unpurified pancreatic progenitors can be differentiated into  $\beta$  cell clusters when transplanted into an E12.5 mouse dorsal pancreas (Brolen et al., 2005). Based on these results, we hypothesize that hES cell-derived Pdx1-positive progenitors will respond to signals within the *in vitro* developing dorsal pancreas by differentiating into insulin-positive  $\beta$  cells and other pancreatic cell types. To test this directly, hES cell-derived Pdx1-positive cells can be isolated by FACS and injected into the dorsal pancreas from E12.5 *Neurog3<sup>-/-</sup>* embryos, which lack the ability to generate endogenous endocrine cells. Subsequently, explants can be analyzed by both immunolabeling

and RT-PCR for markers characteristic of human cells and mature functional endocrine and exocrine cells. This analysis will reveal if these hES cell-derived progenitors are, in fact, *bona fide*  $\beta$  cell progenitors and competent to mature *in vivo* without the undesired teratoma formation evident in previous studies (Kroon et al., 2008).

### **Examining complex gene regulatory networks and elucidating signaling pathways through transcriptional profiling of developmental intermediates**

As transcriptional profiles are obtained, we expect to identify transcripts that are not previously known to be expressed in each cell population, including components of growth factor pathways, juxtacrine or intercellular signaling pathways, transcription factors, structural proteins, channels, and other necessary proteins for hormone production, transport, and secretion. Thus, we will seek to identify a signature group of transcripts that define each intermediate cell type and then examine whether specific hES cell differentiation methods favor the derivation of each cell type from hES cells. However, while experiments using RNA-Seq will reveal transcriptional profiles for a number of critical cell populations, they do not directly reveal how these profiles are established. Such knowledge can be obtained only by identifying transcription factor binding sites. Thus, to facilitate identification of critical regulatory molecules for each stage of pancreas development process, it would be advantageous to determine the direct transcriptional targets of factors whose function is well established for pancreatic cell differentiation. While such studies can be performed using antibody-based ChIP analysis, antibodies that are suitable for this purpose are sometimes unavailable. Therefore, mouse lines in which a gene of interest has been epitope-tagged may be necessary to further examine the complex regulatory networks that underlie each developmental intermediate.

The RNA-Seq datasets provide a significant amount of information pertaining to each

cellular population analyzed, and these datasets are rich for downstream analyses of specific transcription factors with novel roles in pancreatic organogenesis, for elucidating gene regulatory networks that promote cell specification, and for examining signaling pathways that direct cell differentiation. By performing pair-wise analyses of the datasets, differentially expressed transcripts can be identified. Performing such a comparison, the analysis of *Ptf1a*<sup>YFP</sup> expressing MPCs and *Ptf1a*<sup>YFP</sup>-expressing acinar specified cells led to the identification of *Nepn*, an inhibitor of TGF $\beta$  signaling, as a potential novel marker of the pancreatic MPCs (**Figure 4.1**). While the role of Nepn in pancreas development and MPC specification still remains unclear, the use of a *ROSA26*<sup>Nepn-Cherry</sup> transgene permitted the identification of *Nepn*-expressing cells during embryogenesis (**Figure 4.4**). However, the *ROSA26*<sup>Nepn-Cherry</sup> transgene has certain drawbacks, therefore generating a targeted reporter allele for Nepn will be essential for more accurately assessing its expression pattern during development and for identifying and isolating *Nepn*-expressing cells.

While a targeted reporter allele for *Nepn* will permit the identification and isolation of *Nepn*-expressing cells during development and elucidate the developmental role of Nepn in the context of a null background, the current design of the targeted allele will not permit the generation of a conditional knockout using spatiotemporally restricted Cre expression. Given that we would like to analyze the role of Nepn in pancreatic MPC development, if the global null allele displays a phenotype much earlier during development than pancreatic progenitor specification it may be necessary to generate a conditional floxed allele. Taking this into consideration, as investigators move forward analyzing the RNA-Seq datasets, identifying differentially expressed transcripts, and selecting transcripts of interest, the generation of a conditional allele may be more advantageous than simply a reporter allele for its analysis at certain developmental timepoints.

In addition to identifying transcription factors with potential novel roles in pancreas

development, the RNA-Seq datasets hold the possibility of uncovering signaling pathways critical for endoderm, MPC, and/or endocrine progenitor specification. Signaling pathways are the main foundation upon which a cell becomes specified, and these pathways activate and/or repress certain transcriptional regulatory networks, thus governing the differentiation of cells. Therefore, an analysis of which receptors and signaling molecules are differentially expressed between two various timepoints will elucidate which signaling pathways may be upregulated or downregulated at a particular stage and thus be critical for the specification of a certain developmental intermediate.

### **Analysis of cell populations using combinatorial sorting**

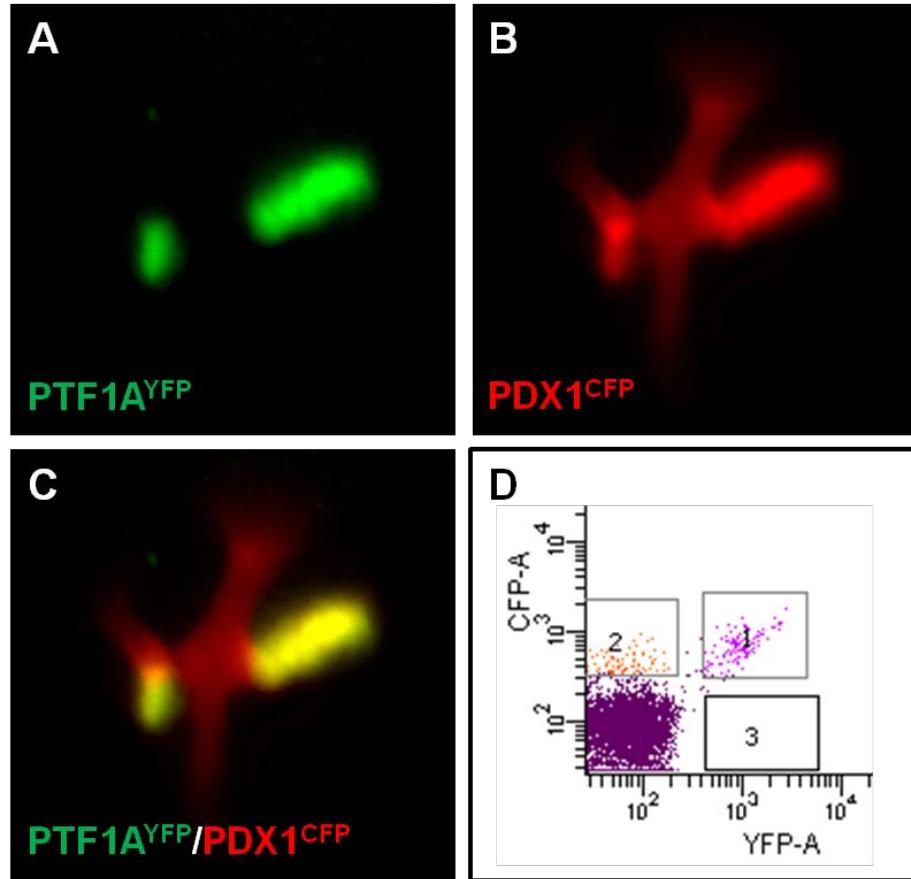
In addition to obtaining transcriptional profiles for *Ptf1a<sup>YFP</sup>*-expressing cells, the generation of a new fluorescently-tagged *Pdx1* reporter permits isolation of additional distinct developmental intermediates. Although *Pdx1* is expressed in other early endoderm populations such as the epithelium of the duodenum, cystic duct, common bile duct and antral stomach, it is a valuable marker of the pancreatic MPCs and will facilitate isolation of spatially-defined progenitor cells within the developing foregut. When used in combination with other fluorescently-tagged alleles, *Pdx1<sup>CFP</sup>* allows for the isolation of cell populations that will provide important comparisons following transcriptional profiling of the cells. For example, when used in combination with *Ptf1a<sup>YFP</sup>*, *Pdx1<sup>CFP</sup>* allows for the isolation of two critical populations: 1) non-pancreatic foregut cells (*Pdx1<sup>CFP/+</sup>* cells minus *Ptf1a<sup>YFP/+</sup>* cells at E11.5) and 2) late pancreatic MPCs (*Ptf1a<sup>YFP/+</sup>*; *Pdx1<sup>CFP/+</sup>* double-positive cells at E12.5).

During early stages of development (E10.5 – E11.5), *Pdx1* is broadly expressed throughout the posterior foregut endoderm, serving to mark the dorsal and ventral pancreatic epithelium and the antral stomach, common bile duct, and duodenum (**Figure 6.1**). Conversely, *Ptf1a* is primarily restricted to the dorsal and ventral pancreatic epithelium

within the posterior foregut (**Figure 6.1**). Thus, by isolating  $Pdx1^{CFP/+}$  single-positive cells and comparing them with  $Ptf1a^{YFP/+}; Pdx1^{CFP/+}$  double-positive cells, we will be able to examine the transcriptional differences between cells that adopt a gastrointestinal or pancreatic fate (**Figure 6.1**). Indeed, given the differing expression patterns of  $Pdx1$  and  $Ptf1a$  within the foregut endoderm, this combinatorial comparison will allow for the identification of signaling pathways that promote the growth and maturation of the pancreatic progenitors explicitly and can in turn be used to enhance hES cell directed differentiation efforts whereby posterior foregut endoderm-like cells are specifically differentiated towards pancreatic fates and not other foregut derivatives.

After E11.5, the pancreatic epithelium undergoes extensive proliferation and branching leading to the formation of an epithelial ductal tree.  $Ptf1a$  is predominantly expressed in the periphery of the epithelial tree (termed the “tip” cells) and restricted from the central domain of the epithelial tree (termed the “trunk” domain) (Zhou et al., 2007). Conversely,  $Pdx1$  expression is evident in both the periphery and central domain of the epithelial tree. Importantly, the periphery of the epithelial tree is thought to maintain the pancreatic MPC population for a period of time, while the central domain contains a bipotential cell population that arose from the MPCs and can subsequently give rise to endocrine and ductal cells (Gu et al., 2002; Zhou et al., 2007). Thus,  $Pdx1$  and  $Ptf1a$  can be used in combination to distinguish the late pancreatic MPCs in the periphery of the epithelium ( $Ptf1a^{YFP/+}; Pdx1^{CFP/+}$  double-positive cells) and the bipotential progeny of the central domain ( $Pdx1^{CFP/+}$  single-positive cells).

However, the full utility of reporter alleles in combination is only realized when the FPs have spectrally distinct excitation and emission profiles (Heikal et al., 2000; Nagai et al., 2002; Rizzo et al., 2004; Shaner et al., 2004). When selecting FP reporters to be used in

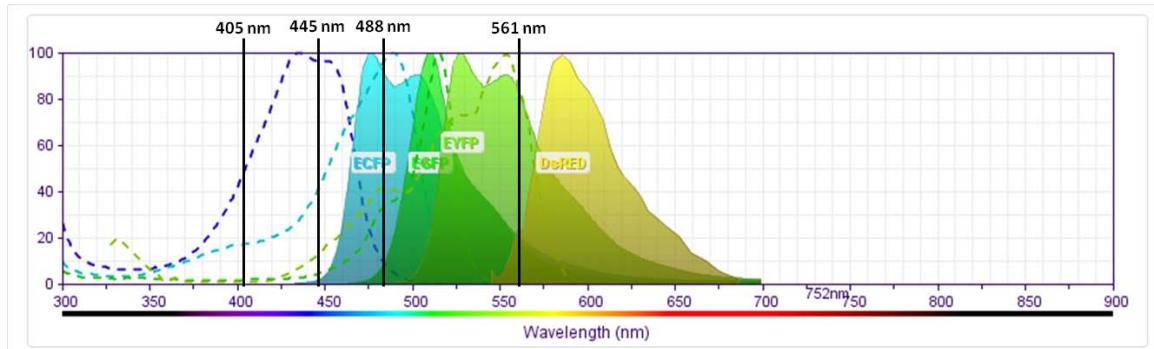


**Figure 6.1 Combinatorial analysis of *Ptf1a*<sup>YFP/+</sup>; *Pdx1*<sup>CFP/+</sup> embryos by fluorescence-activated cell sorting (FACS).** A) Whole mount fluorescent imaging of an E11.5 *Ptf1a*<sup>YFP/+</sup> dissected embryo which displays fluorescence in the dorsal and ventral pancreatic buds. B) Conversely, *Pdx1*<sup>CFP/+</sup> dissected embryos display expression in the caudal stomach, common bile duct, and duodenum in addition to the dorsal and ventral pancreatic epithelium. C) Merged fluorescence revealing co-expression of *Ptf1a*<sup>YFP</sup> and *Pdx1*<sup>CFP</sup> in the dorsal and ventral pancreatic buds. D) When used in combination *Pdx1*<sup>CFP</sup> and *Ptf1a*<sup>YFP</sup> permit the isolation of *Ptf1a*<sup>YFP/+</sup>; *Pdx1*<sup>CFP/+</sup> double-positive pancreatic MPCs (gate 1) and *Pdx1*<sup>CFP/+</sup> single-positive non-pancreatic foregut cells (gate 2); however the latter contains significant contamination due to cellular autofluorescence.

combination, it is important to assess both the wavelength at which they are optimally excited and the range of emission wavelengths they exhibit. An analysis of the spectral profiles for cyan, green, yellow and red fluorescent protein variants is critical for assessing which combinations of fluorescent reporters can be discriminated by FACS (**Figure 6.2**). Although CFP can be excited with the 405 nm laser, this suboptimal excitation resulted in the inability to distinguish between *CFP*-expressing cells and the cellular auto-fluorescence seen in wild type cells (**Figure 5.7**). While the use of the 445 nm laser alleviated this problem, currently it is not possible to utilize both the 445 nm and 488 nm lasers simultaneously, thus not permitting concurrent sorting of CFP and YFP. However, one alternative approach is to utilize a recently generated *Ptf1a<sup>tdTomato</sup>*-expressing mouse (unpublished). The tandem dimer red fluorescent protein, tdTomato, is optimally excited at 554 nm (Shaner et al., 2004). Thus, it is possible to utilize the 561 nm laser, for tdTomato excitation, in combination with the 445 nm laser, for Cerulean excitation, to perform combinatorial cell sorting and obtain the desired double-positive and single-positive cell populations.

### **Improving the directed pancreatic differentiation of human ES cells**

How can you know you are on the right track if you do not even know where you want to end up? While philosophical in nature, that question holds true even for the directed differentiation of hES cells to pancreatic fates. An understanding of the fundamental characteristics of authentic, *in vivo*-isolated mature pancreatic  $\beta$  cells will certainly provide a “landmark” to strive towards as we attempt to produce cells that resemble certain cellular intermediates. In support of this theory, the transcriptional analyses of FACS-purified hematopoietic progenitor cells have accelerated the attempts to generate these cells from hES cells (Bu et al., 2009; Yang et al., 2008). Seeing that the premise underlying *in vitro* pancreatic directed differentiation is to both closely mimic the normal developmental



**Figure 6.2 Spectral profile for cyan fluorescent protein (ECFP), green fluorescent protein (EGFP), yellow fluorescent protein (EYFP) and red fluorescent protein (DsRED).** Excitation (dashed lines) and emission (filled peaks) profiles are shown for ECFP (dark blue dashed line/blue filled emission peak), EGFP (light blue/green), EYFP (dark green/light green) and DsRED (light green/yellow). Excitation wavelengths at 405 nm, 445 nm, 488 nm, and 561 nm are noted. The spectral profiles for ECFP, EYFP and dsRED are similar to the profiles for Cerulean, Citrine and tdTomato, respectively. By using different fluorescently-tagged alleles in the mouse, we are able to isolate by fluorescence activated cell sorting (FACS) various cell populations marked by specific genetic reporters. However, when determining fluorescently-tagged alleles to use in combination, it is critical to examine the spectral overlap displayed by the excitation and emission profiles.

signaling mechanisms for cell specification and establish the proper gene regulatory networks at each developmental stage, a deeper understanding of the signaling pathways and transcriptional networks established at each developmental stage will further accelerate how functional  $\beta$  cells can be replicated in cultured cells. While progress has been made in learning how to generate specific  $\beta$  cell forebearers such as definitive, foregut and pancreatic endoderm, the differentiation towards these intermediates still results in a heterogeneous population of cells. The contributions of said heterogeneity is debatable with a number of investigators recognizing that the undesired cell populations provide signals to instruct the other cells towards pancreatic fates while other investigators feel that the heterogeneity simply decreases the overall efficiency and any potential signals provided by these undesired cells should be elucidated to achieve more robust differentiation. Recognizing this problem, we assert that the best path forward is to 1) isolate and characterize specific pancreatic progenitor cell populations, 2) profile gene expression changes at the level of the entire transcriptome, and 3) learn how to translate this information into better protocols for the directed differentiation of hES cells into hormone-producing cells. Thus, the transcriptional profiling reported in this dissertation is only one of the high quality datasets that will be obtained from defined pancreatic progenitor cell populations and subsequently used to develop new bioinformatics strategies that will enable the discovery of pro- $\beta$  cell signals in the mouse and other key information for guiding improvements of protocols for the directed differentiation of hES cells into  $\beta$  cells.

While the genes and developmental pathways are likely to be highly conserved between mice and humans, we recognize that they are unlikely to be identical. Thus, there are major challenges for performing meaningful comparisons of gene expression profiles in a setting that is highly dynamic and temporally-specific. For this reason further strategies need to be developed for 1) distinguishing natural species-related differences in development

between mouse and human and 2) distinguishing the differences in gene expression profiles that arise *in vivo* with those that are the result of *ex vivo* manipulations. Finally, even when cross-species differences can be interpreted and managed, we will need to determine what extra- and intracellular signals invoke the expression of specific gene networks and clusters. However, in spite of the limitations, a deeper understanding of the transcriptional networks established at each developmental intermediate is likely to guide the directed differentiation efforts concerning the differentiation of hES cells to pancreatic fates and thus, in turn, expedite the process by which we can develop a cellular therapy for diabetes.

## REFERENCES

- Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., *et al.* (1991). Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* *252*, 1651-1656.
- Affolter, M., Zeller, R., and Caussinus, E. (2009). Tissue remodelling through branching morphogenesis. *Nat Rev Mol Cell Biol* *10*, 831-842.
- Ahlgren, U., Jonsson, J., Jonsson, L., Simu, K., and Edlund, H. (1998). beta-cell-specific inactivation of the mouse Ipf1/Pdx1 gene results in loss of the beta-cell phenotype and maturity onset diabetes. *Genes Dev* *12*, 1763-1768.
- Alba, R., Fei, Z., Payton, P., Liu, Y., Moore, S.L., Debbie, P., Cohn, J., D'Ascenzo, M., Gordon, J.S., Rose, J.K., *et al.* (2004). ESTs, cDNA microarrays, and gene expression profiling: tools for dissecting plant physiology and development. *Plant J* *39*, 697-714.
- Anderson, K.R., White, P., Kaestner, K.H., and Sussel, L. (2009). Identification of known and novel pancreas genes expressed downstream of Nkx2.2 during development. *BMC Dev Biol* *9*, 65.
- Ang, S.L., Wierda, A., Wong, D., Stevens, K.A., Cascio, S., Rossant, J., and Zaret, K.S. (1993). The formation and maintenance of the definitive endoderm lineage in the mouse: involvement of HNF3/forkhead proteins. *Development* *119*, 1301-1315.
- Araki, K., Araki, M., and Yamamura, K. (2002). Site-directed integration of the cre gene mediated by Cre recombinase using a combination of mutant lox sites. *Nucleic Acids Res* *30*, e103.
- Assady, S., Maor, G., Amit, M., Itskovitz-Eldor, J., Skorecki, K.L., and Tzukerman, M. (2001). Insulin production by human embryonic stem cells. *Diabetes* *50*, 1691-1697.
- Baharvand, H., Jafary, H., Massumi, M., and Ashtiani, S.K. (2006). Generation of insulin-secreting cells from human embryonic stem cells. *Dev Growth Differ* *48*, 323-332.
- Ballinger, W.F., and Lacy, P.E. (1972). Transplantation of intact pancreatic islets in rats. *Surgery* *72*, 175-186.
- Barbacci, E., Reber, M., Ott, M.O., Breillat, C., Huetz, F., and Cereghini, S. (1999). Variant hepatocyte nuclear factor 1 is required for visceral endoderm specification. *Development* *126*, 4795-4805.
- Bayha, E., Jorgensen, M.C., Serup, P., and Grapin-Botton, A. (2009). Retinoic acid signaling organizes endodermal organ specification along the entire antero-posterior axis. *PLoS One* *4*, e5845.
- Beddington, R.S., and Robertson, E.J. (1989). An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. *Development* *105*, 733-737.

- Bennett, S.T., Barnes, C., Cox, A., Davies, L., and Brown, C. (2005). Toward the 1,000 dollars human genome. *Pharmacogenomics* *6*, 373-382.
- Beres, T.M., Masui, T., Swift, G.H., Shi, L., Henke, R.M., and MacDonald, R.J. (2006). PTF1 is an organ-specific and Notch-independent basic helix-loop-helix complex containing the mammalian Suppressor of Hairless (RBP-J) or its parologue, RBP-L. *Mol Cell Biol* *26*, 117-130.
- Berne, R.M., and Levy, M.N. (1993). *Physiology*, 3rd edn (St. Louis, Mosby Year Book).
- 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.
- Biben, C., Stanley, E., Fabri, L., Kotecha, S., Rhinn, M., Drinkwater, C., Lah, M., Wang, C.C., Nash, A., Hilton, D., *et al.* (1998). Murine cerberus homologue mCer-1: a candidate anterior patterning molecule. *Dev Biol* *194*, 135-151.
- Brennan, J., Lu, C.C., Norris, D.P., Rodriguez, T.A., Beddington, R.S., and Robertson, E.J. (2001). Nodal signalling in the epiblast patterns the early mouse embryo. *Nature* *411*, 965-969.
- Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., *et al.* (2000). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol* *18*, 630-634.
- Brolen, G.K., Heins, N., Edsbagge, J., and Semb, H. (2005). Signals from the embryonic mouse pancreas induce differentiation of human embryonic stem cells into insulin-producing beta-cell-like cells. *Diabetes* *54*, 2867-2874.
- Brubaker, P.L., and Drucker, D.J. (2004). Minireview: Glucagon-like peptides regulate cell proliferation and apoptosis in the pancreas, gut, and central nervous system. *Endocrinology* *145*, 2653-2659.
- Bu, L., Jiang, X., Martin-Puig, S., Caron, L., Zhu, S., Shao, Y., Roberts, D.J., Huang, P.L., Domian, I.J., and Chien, K.R. (2009). Human ISL1 heart progenitors generate diverse multipotent cardiovascular cell lineages. *Nature* *460*, 113-117.
- Burlison, J.S., Long, Q., Fujitani, Y., Wright, C.V., and Magnuson, M.A. (2008). Pdx-1 and Ptfla concurrently determine fate specification of pancreatic multipotent progenitor cells. *Dev Biol* *316*, 74-86.
- Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., and Prasher, D.C. (1994). Green fluorescent protein as a marker for gene expression. *Science* *263*, 802-805.
- Champy, M.F., Le Voci, L., Selloum, M., Peterson, L.B., Cumiskey, A.M., and Blom, D. (2010). Reduced body weight in male Tspan8-deficient mice. *Int J Obes (Lond)*.

Chen, S.X., Osipovich, A.B., Ustione, A., Potter, L.A., Hipkens, S., Gangula, R., Yuan, W., Piston, D.W., and Magnuson, M.A. (2011). Quantification of factors influencing fluorescent protein expression using RMCE to generate an allelic series in the ROSA26 locus in mice. *Dis Model Mech.*

Chiang, M.K., and Melton, D.A. (2003). Single-cell transcript analysis of pancreas development. *Dev Cell* 4, 383-393.

Cho, Y.M., Lim, J.M., Yoo, D.H., Kim, J.H., Chung, S.S., Park, S.G., Kim, T.H., Oh, S.K.,

Choi, Y.M., Moon, S.Y., *et al.* (2008). Betacellulin and nicotinamide sustain PDX1 expression and induce pancreatic beta-cell differentiation in human embryonic stem cells. *Biochem Biophys Res Commun* 366, 129-134.

Choi, M.Y., Romer, A.I., Hu, M., Lepourcelet, M., Mechoor, A., Yesilaltay, A., Krieger, M., Gray, P.A., and Shivdasani, R.A. (2006). A dynamic expression survey identifies transcription factors relevant in mouse digestive tract development. *Development* 133, 4119-4129.

Chu, K., and Tsai, M.J. (2005). Neuronatin, a downstream target of BETA2/NeuroD1 in the pancreas, is involved in glucose-mediated insulin secretion. *Diabetes* 54, 1064-1073.

Clayton, E., Doupe, D.P., Klein, A.M., Winton, D.J., Simons, B.D., and Jones, P.H. (2007). A single type of progenitor cell maintains normal epidermis. *Nature* 446, 185-189.

Cloonan, N., Forrest, A.R., Kolle, G., Gardiner, B.B., Faulkner, G.J., Brown, M.K., Taylor, D.F., Steptoe, A.L., Wani, S., Bethel, G., *et al.* (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat Methods* 5, 613-619.

Coffinier, C., Barra, J., Babinet, C., and Yaniv, M. (1999). Expression of the vHNF1/HNF1beta homeoprotein gene during mouse organogenesis. *Mech Dev* 89, 211-213.

Conlon, F.L., Lyons, K.M., Takaesu, N., Barth, K.S., Kispert, A., Herrmann, B., and Robertson, E.J. (1994). A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. *Development* 120, 1919-1928.

Copeland, N.G., Jenkins, N.A., and Court, D.L. (2001). Recombineering: a powerful new tool for mouse functional genomics. *Nat Rev Genet* 2, 769-779.

Costa, M., Dottori, M., Sourris, K., Jamshidi, P., Hatzistavrou, T., Davis, R., Azzola, L., Jackson, S., Lim, S.M., Pera, M., *et al.* (2007). A method for genetic modification of human embryonic stem cells using electroporation. *Nat Protoc* 2, 792-796.

D'Amour, K.A., Agulnick, A.D., Eliazer, S., Kelly, O.G., Kroon, E., and Baetge, E.E. (2005). Efficient differentiation of human embryonic stem cells to definitive endoderm. *Nat Biotechnol* 23, 1534-1541.

- 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. *Nat Biotechnol* 24, 1392-1401.
- Davis, R.P., Ng, E.S., Costa, M., Mossman, A.K., Sourris, K., Elefanty, A.G., and Stanley, E.G. (2008). Targeting a GFP reporter gene to the MIXL1 locus of human embryonic stem cells identifies human primitive streak-like cells and enables isolation of primitive hematopoietic precursors. *Blood* 111, 1876-1884.
- Derynck, R., and Zhang, Y. (1996). Intracellular signalling: the mad way to do it. *Curr Biol* 6, 1226-1229.
- Dessimoz, J., Opoka, R., Kordich, J.J., Grapin-Botton, A., and Wells, J.M. (2006). FGF signaling is necessary for establishing gut tube domains along the anterior-posterior axis in vivo. *Mech Dev* 123, 42-55.
- Deutsch, G., Jung, J., Zheng, M., Lora, J., and Zaret, K.S. (2001). A bipotential precursor population for pancreas and liver within the embryonic endoderm. *Development* 128, 871-881.
- Duncan, S.A., Manova, K., Chen, W.S., Hoodless, P., Weinstein, D.C., Bachvarova, R.F., and Darnell, J.E., Jr. (1994). Expression of transcription factor HNF-4 in the extraembryonic endoderm, gut, and nephrogenic tissue of the developing mouse embryo: HNF-4 is a marker for primary endoderm in the implanting blastocyst. *Proc Natl Acad Sci U S A* 91, 7598-7602.
- Ema, H., and Nakauchi, H. (2003). Self-renewal and lineage restriction of hematopoietic stem cells. *Curr Opin Genet Dev* 13, 508-512.
- Ewens, W.J., and Grant, G.R. (2004). In *Statistical Methods in Bioinformatics: An Introduction*, M. Gail, K. Krickeberg, J. Samet, A. Tsiatis, and W. Wong, eds. (New York, Springer), pp. 139-142.
- Fausto, N. (2000). Liver regeneration. *J Hepatol* 32, 19-31.
- Feldman, A.L., Costouros, N.G., Wang, E., Qian, M., Marincola, F.M., Alexander, H.R., and Libutti, S.K. (2002). Advantages of mRNA amplification for microarray analysis. *Biotechniques* 33, 906-912, 914.
- Fischer, Y., Ganic, E., Ameri, J., Xian, X., Johannesson, M., and Semb, H. (2010). NANOG reporter cell lines generated by gene targeting in human embryonic stem cells. *PLoS One* 5.
- Fuchs, E., Tumbar, T., and Guasch, G. (2004). Socializing with the neighbors: stem cells and their niche. *Cell* 116, 769-778.
- Fukuda, A., Kawaguchi, Y., Furuyama, K., Kodama, S., Horiguchi, M., Kuhara, T., Kawaguchi, M., Terao, M., Doi, R., Wright, C.V., et al. (2008). Reduction of Ptf1a gene dosage causes pancreatic hypoplasia and diabetes in mice. *Diabetes* 57, 2421-2431.

- Gannon, M., Gerrish, K., Stein, R., and Wright, C. (2000). Role of Pdx-1 in Pancreatic Development. In Molecular Pathogenesis of MODYs, F. Matschinsky, and M. Magnuson, eds. (Basel, Karger), pp. 166-180.
- Ghosh, B., and Leach, S.D. (2006). Interactions between hairy/enhancer of split-related proteins and the pancreatic transcription factor Ptf1-p48 modulate function of the PTF1 transcriptional complex. *Biochem J* 393, 679-685.
- Gittes, G.K. (2009). Developmental biology of the pancreas: a comprehensive review. *Dev Biol* 326, 4-35.
- Gordon, J.W., and Ruddle, F.H. (1981). Integration and stable germ line transmission of genes injected into mouse pronuclei. *Science* 214, 1244-1246.
- Gowda, M., Jantasuriyarat, C., Dean, R.A., and Wang, G.L. (2004). Robust-LongSAGE (RL-SAGE): a substantially improved LongSAGE method for gene discovery and transcriptome analysis. *Plant Physiol* 134, 890-897.
- Gradwohl, G., Dierich, A., LeMeur, M., and Guillemot, F. (2000). neurogenin3 is required for the development of the four endocrine cell lineages of the pancreas. *Proc Natl Acad Sci U S A* 97, 1607-1611.
- Griesbeck, O., Baird, G.S., Campbell, R.E., Zacharias, D.A., and Tsien, R.Y. (2001). Reducing the environmental sensitivity of yellow fluorescent protein. Mechanism and applications. *J Biol Chem* 276, 29188-29194.
- Gu, G., Dubauskaite, J., and Melton, D.A. (2002). Direct evidence for the pancreatic lineage: NGN3+ cells are islet progenitors and are distinct from duct progenitors. *Development* 129, 2447-2457.
- Gu, G., Wells, J.M., Dombkowski, D., Preffer, F., Aronow, B., and Melton, D.A. (2004). Global expression analysis of gene regulatory pathways during endocrine pancreatic development. *Development* 131, 165-179.
- Guz, Y., Montminy, M.R., Stein, R., Leonard, J., Gamer, L.W., Wright, C.V., and Teitelman, G. (1995). Expression of murine STF-1, a putative insulin gene transcription factor, in beta cells of pancreas, duodenal epithelium and pancreatic exocrine and endocrine progenitors during ontogeny. *Development* 121, 11-18.
- Han, J.H., Rall, L., and Rutter, W.J. (1986). Selective expression of rat pancreatic genes during embryonic development. *Proc Natl Acad Sci U S A* 83, 110-114.
- Harmon, E.B., Apelqvist, A.A., Smart, N.G., Gu, X., Osborne, D.H., and Kim, S.K. (2004). GDF11 modulates NGN3+ islet progenitor cell number and promotes beta-cell differentiation in pancreas development. *Development* 131, 6163-6174.
- Harrison, K.A., Druey, K.M., Deguchi, Y., Tuscano, J.M., and Kehrl, J.H. (1994). A novel human homeobox gene distantly related to proboscipedia is expressed in lymphoid and pancreatic tissues. *J Biol Chem* 269, 19968-19975.

- Harrison, K.A., Thaler, J., Pfaff, S.L., Gu, H., and Kehrl, J.H. (1999). Pancreas dorsal lobe agenesis and abnormal islets of Langerhans in Hlxb9-deficient mice. *Nat Genet* 23, 71-75.
- Hart, A., Papadopoulou, S., and Edlund, H. (2003). Fgf10 maintains notch activation, stimulates proliferation, and blocks differentiation of pancreatic epithelial cells. *Dev Dyn* 228, 185-193.
- Hebrok, M., Kim, S.K., and Melton, D.A. (1998). Notochord repression of endodermal Sonic hedgehog permits pancreas development. *Genes Dev* 12, 1705-1713.
- Heikal, A.A., Hess, S.T., Baird, G.S., Tsien, R.Y., and Webb, W.W. (2000). Molecular spectroscopy and dynamics of intrinsically fluorescent proteins: coral red (dsRed) and yellow (Citrine). *Proc Natl Acad Sci U S A* 97, 11996-12001.
- Heim, R., Cubitt, A.B., and Tsien, R.Y. (1995). Improved green fluorescence. *Nature* 373, 663-664.
- Hering, B., Schultz, A., Schultz, B., Brendel, M., and Bretzel, R.e. (1996). International Islet Transplant Registry Newsletter no. 7 (Giesen, Germany).
- Herrera, P.L. (2000). Adult insulin- and glucagon-producing cells differentiate from two independent cell lineages. *Development* 127, 2317-2322.
- Herrera, P.L., Huarte, J., Zufferey, R., Nichols, A., Mermilliod, B., Philippe, J., Muniesa, P., Sanvito, F., Orci, L., and Vassalli, J.D. (1994). Ablation of islet endocrine cells by targeted expression of hormone-promoter-driven toxigenes. *Proc Natl Acad Sci U S A* 91, 12999-13003.
- Hildebrand, A., Romaris, M., Rasmussen, L.M., Heinegard, D., Twardzik, D.R., Border, W.A., and Ruoslahti, E. (1994). Interaction of the small interstitial proteoglycans biglycan, decorin and fibromodulin with transforming growth factor beta. *Biochem J* 302 ( Pt 2), 527-534.
- Hirshberg, B., Rother, K.I., Digon, B.J., 3rd, Lee, J., Gaglia, J.L., Hines, K., Read, E.J., Chang, R., Wood, B.J., and Harlan, D.M. (2003). Benefits and risks of solitary islet transplantation for type 1 diabetes using steroid-sparing immunosuppression: the National Institutes of Health experience. *Diabetes Care* 26, 3288-3295.
- Hocking, A.M., Shinomura, T., and McQuillan, D.J. (1998). Leucine-rich repeat glycoproteins of the extracellular matrix. *Matrix Biol* 17, 1-19.
- Hou, J., Charters, A.M., Lee, S.C., Zhao, Y., Wu, M.K., Jones, S.J., Marra, M.A., and Hoodless, P.A. (2007). A systematic screen for genes expressed in definitive endoderm by Serial Analysis of Gene Expression (SAGE). *BMC Dev Biol* 7, 92.
- Hudson, C., Clements, D., Friday, R.V., Stott, D., and Woodland, H.R. (1997). Xsox17alpha and -beta mediate endoderm formation in Xenopus. *Cell* 91, 397-405.

- Iozzo, R.V. (1997). The family of the small leucine-rich proteoglycans: key regulators of matrix assembly and cellular growth. *Crit Rev Biochem Mol Biol* 32, 141-174.
- Iozzo, R.V. (1998). Matrix proteoglycans: from molecular design to cellular function. *Annu Rev Biochem* 67, 609-652.
- Iozzo, R.V. (1999). The biology of the small leucine-rich proteoglycans. Functional network of interactive proteins. *J Biol Chem* 274, 18843-18846.
- Irion, S., Luche, H., Gadue, P., Fehling, H.J., Kennedy, M., and Keller, G. (2007). Identification and targeting of the ROSA26 locus in human embryonic stem cells. *Nat Biotechnol* 25, 1477-1482.
- Jacquemin, P., Yoshitomi, H., Kashima, Y., Rousseau, G.G., Lemaigre, F.P., and Zaret, K.S. (2006). An endothelial-mesenchymal relay pathway regulates early phases of pancreas development. *Dev Biol* 290, 189-199.
- Jain, K.K. (2005). Ethical and regulatory aspects of embryonic stem cell research. *Expert Opin Biol Ther* 5, 153-162.
- Jensen, J., Heller, R.S., Funder-Nielsen, T., Pedersen, E.E., Lindsell, C., Weinmaster, G., Madsen, O.D., and Serup, P. (2000). Independent development of pancreatic alpha- and beta-cells from neurogenin3-expressing precursors: a role for the notch pathway in repression of premature differentiation. *Diabetes* 49, 163-176.
- Jiang, J., Au, M., Lu, K., Eshpeter, A., Korbutt, G., Fisk, G., and Majumdar, A.S. (2007a). Generation of insulin-producing islet-like clusters from human embryonic stem cells. *Stem Cells* 25, 1940-1953.
- Jiang, W., Shi, Y., Zhao, D., Chen, S., Yong, J., Zhang, J., Qing, T., Sun, X., Zhang, P., Ding, M., et al. (2007b). In vitro derivation of functional insulin-producing cells from human embryonic stem cells. *Cell Res* 17, 333-344.
- Jonsson, J., Carlsson, L., Edlund, T., and Edlund, H. (1994). Insulin-promoter-factor 1 is required for pancreas development in mice. *Nature* 371, 606-609.
- Jorgensen, M.C., Ahnfelt-Ronne, J., Hald, J., Madsen, O.D., Serup, P., and Hecksher-Sorensen, J. (2007). An illustrated review of early pancreas development in the mouse. *Endocr Rev* 28, 685-705.
- Jung, J., Zheng, M., Goldfarb, M., and Zaret, K.S. (1999). Initiation of mammalian liver development from endoderm by fibroblast growth factors. *Science* 284, 1998-2003.
- Kanai-Azuma, M., Kanai, Y., Gad, J.M., Tajima, Y., Taya, C., Kurohmaru, M., Sanai, Y., Yonekawa, H., Yazaki, K., Tam, P.P., et al. (2002). Depletion of definitive gut endoderm in Sox17-null mutant mice. *Development* 129, 2367-2379.

- Kawaguchi, Y., Cooper, B., Gannon, M., Ray, M., MacDonald, R.J., and Wright, C.V. (2002). The role of the transcriptional regulator Ptf1a in converting intestinal to pancreatic progenitors. *Nat Genet* 32, 128-134.
- Kim, S.K., Hebrok, M., and Melton, D.A. (1997). Notochord to endoderm signaling is required for pancreas development. *Development* 124, 4243-4252.
- Kim, S.K., and Melton, D.A. (1998). Pancreas development is promoted by cyclopamine, a hedgehog signaling inhibitor. *Proc Natl Acad Sci U S A* 95, 13036-13041.
- Kimelman, D., and Griffin, K.J. (2000). Vertebrate mesendoderm induction and patterning. *Curr Opin Genet Dev* 10, 350-356.
- Kinkel, M.D., Sefton, E.M., Kikuchi, Y., Mizoguchi, T., Ward, A.B., and Prince, V.E. (2009). Cyp26 enzymes function in endoderm to regulate pancreatic field size. *Proc Natl Acad Sci U S A* 106, 7864-7869.
- Kodzius, R., Kojima, M., Nishiyori, H., Nakamura, M., Fukuda, S., Tagami, M., Sasaki, D., Imamura, K., Kai, C., Harbers, M., *et al.* (2006). CAGE: cap analysis of gene expression. *Nat Methods* 3, 211-222.
- Krapp, A., Knofler, M., Ledermann, B., Burki, K., Berney, C., Zoerkler, N., Hagenbuchle, O., and Wellauer, P.K. (1998). The bHLH protein PTF1-p48 is essential for the formation of the exocrine and the correct spatial organization of the endocrine pancreas. *Genes Dev* 12, 3752-3763.
- 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. *Nat Biotechnol* 26, 443-452.
- Krusius, T., and Ruoslahti, E. (1986). Primary structure of an extracellular matrix proteoglycan core protein deduced from cloned cDNA. *Proc Natl Acad Sci U S A* 83, 7683-7687.
- Kubo, A., Shinozaki, K., Shannon, J.M., Kouskoff, V., Kennedy, M., Woo, S., Fehling, H.J., and Keller, G. (2004). Development of definitive endoderm from embryonic stem cells in culture. *Development* 131, 1651-1662.
- Kumar, M., Jordan, N., Melton, D., and Grapin-Botton, A. (2003). Signals from lateral plate mesoderm instruct endoderm toward a pancreatic fate. *Dev Biol* 259, 109-122.
- Kwon, G.S., Viotti, M., and Hadjantonakis, A.K. (2008). The endoderm of the mouse embryo arises by dynamic widespread intercalation of embryonic and extraembryonic lineages. *Dev Cell* 15, 509-520.
- Labosky, P.A., Barlow, D.P., and Hogan, B.L. (1994). Mouse embryonic germ (EG) cell lines: transmission through the germline and differences in the methylation imprint of

insulin-like growth factor 2 receptor (Igf2r) gene compared with embryonic stem (ES) cell lines. *Development* 120, 3197-3204.

Landry, C., Clotman, F., Hioki, T., Oda, H., Picard, J.J., Lemaigre, F.P., and Rousseau, G.G. (1997). HNF-6 is expressed in endoderm derivatives and nervous system of the mouse embryo and participates to the cross-regulatory network of liver-enriched transcription factors. *Dev Biol* 192, 247-257.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, R25.

Lau, J., Kawahira, H., and Hebrok, M. (2006). Hedgehog signaling in pancreas development and disease. *Cell Mol Life Sci* 63, 642-652.

Lee, E.C., Yu, D., Martinez de Velasco, J., Tessarollo, L., Swing, D.A., Court, D.L., Jenkins, N.A., and Copeland, N.G. (2001). A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics* 73, 56-65.

Lee, G., and Saito, I. (1998). Role of nucleotide sequences of loxP spacer region in Cre-mediated recombination. *Gene* 216, 55-65.

Lee, Y.C., Damholt, A.B., Billestrup, N., Kisbye, T., Galante, P., Michelsen, B., Kofod, H., and Nielsen, J.H. (1999). Developmental expression of proprotein convertase 1/3 in the rat. *Mol Cell Endocrinol* 155, 27-35.

Leonard, J., Peers, B., Johnson, T., Ferreri, K., Lee, S., and Montminy, M.R. (1993). Characterization of somatostatin transactivating factor-1, a novel homeobox factor that stimulates somatostatin expression in pancreatic islet cells. *Mol Endocrinol* 7, 1275-1283.

Li, H., Arber, S., Jessell, T.M., and Edlund, H. (1999). Selective agenesis of the dorsal pancreas in mice lacking homeobox gene Hlxb9. *Nat Genet* 23, 67-70.

Li, L., and Xie, T. (2005). Stem cell niche: structure and function. *Annu Rev Cell Dev Biol* 21, 605-631.

Li, X., Udager, A.M., Hu, C., Qiao, X.T., Richards, N., and Gumucio, D.L. (2009). Dynamic patterning at the pylorus: formation of an epithelial intestine-stomach boundary in late fetal life. *Dev Dyn* 238, 3205-3217.

Li, Y., Rankin, S.A., Sinner, D., Kenny, A.P., Krieg, P.A., and Zorn, A.M. (2008). Sfrp5 coordinates foregut specification and morphogenesis by antagonizing both canonical and noncanonical Wnt11 signaling. *Genes Dev* 22, 3050-3063.

Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. (2008). Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* 133, 523-536.

- Long, Q., Shelton, K.D., Lindner, J., Jones, J.R., and Magnuson, M.A. (2004). Efficient DNA cassette exchange in mouse embryonic stem cells by staggered positive-negative selection. *Genesis* 39, 256-262.
- Lowe, L.A., Yamada, S., and Kuehn, M.R. (2001). Genetic dissection of nodal function in patterning the mouse embryo. *Development* 128, 1831-1843.
- Mackay, D.J., Callaway, J.L., Marks, S.M., White, H.E., Acerini, C.L., Boonen, S.E., Dayanikli, P., Firth, H.V., Goodship, J.A., Haemers, A.P., *et al.* (2008). Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57. *Nat Genet* 40, 949-951.
- Malone, J.H., and Oliver, B. (2011). Microarrays, deep sequencing and the true measure of the transcriptome. *BMC Biol* 9, 34.
- Marguerat, S., and Bahler, J. (2010). RNA-seq: from technology to biology. *Cell Mol Life Sci* 67, 569-579.
- Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., *et al.* (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376-380.
- Marin, F., and Nieto, M.A. (2006). The expression of Scratch genes in the developing and adult brain. *Dev Dyn* 235, 2586-2591.
- Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., and Gilad, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* 18, 1509-1517.
- Martin, M., Gallego-Llamas, J., Ribes, V., Kedinger, M., Niederreither, K., Chambon, P., Dolle, P., and Gradwohl, G. (2005). Dorsal pancreas agenesis in retinoic acid-deficient Raldh2 mutant mice. *Dev Biol* 284, 399-411.
- Massague, J., Blain, S.W., and Lo, R.S. (2000). TGF $\beta$  signaling in growth control, cancer, and heritable disorders. *Cell* 103, 295-309.
- Masui, T., Long, Q., Beres, T.M., Magnuson, M.A., and MacDonald, R.J. (2007). Early pancreatic development requires the vertebrate Suppressor of Hairless (RBPJ) in the PTF1 bHLH complex. *Genes Dev* 21, 2629-2643.
- Masui, T., Swift, G.H., Deering, T., Shen, C., Coats, W.S., Long, Q., Elsasser, H.P., Magnuson, M.A., and MacDonald, R.J. (2010). Replacement of Rbpj with Rbpjl in the PTF1 complex controls the final maturation of pancreatic acinar cells. *Gastroenterology* 139, 270-280.
- Matsumura, H., Ito, A., Saitoh, H., Winter, P., Kahl, G., Reuter, M., Kruger, D.H., and Terauchi, R. (2005). SuperSAGE. *Cell Microbiol* 7, 11-18.

McCulloch, E.A., and Till, J.E. (2005). Perspectives on the properties of stem cells. *Nat Med* *11*, 1026-1028.

McDowell, D.G., Burns, N.A., and Parkes, H.C. (1998). Localised sequence regions possessing high melting temperatures prevent the amplification of a DNA mimic in competitive PCR. *Nucleic Acids Res* *26*, 3340-3347.

McGrath, K.E., Koniski, A.D., Maltby, K.M., McGann, J.K., and Palis, J. (1999). Embryonic expression and function of the chemokine SDF-1 and its receptor, CXCR4. *Dev Biol* *213*, 442-456.

McLean, A.B., D'Amour, K.A., Jones, K.L., Krishnamoorthy, M., Kulik, M.J., Reynolds, D.M., Sheppard, A.M., Liu, H., Xu, Y., Baetge, E.E., *et al.* (2007). Activin a efficiently specifies definitive endoderm from human embryonic stem cells only when phosphatidylinositol 3-kinase signaling is suppressed. *Stem Cells* *25*, 29-38.

McLin, V.A., Rankin, S.A., and Zorn, A.M. (2007). Repression of Wnt/beta-catenin signaling in the anterior endoderm is essential for liver and pancreas development. *Development* *134*, 2207-2217.

Melloul, D., Ben-Neriah, Y., and Cerasi, E. (1993). Glucose modulates the binding of an islet-specific factor to a conserved sequence within the rat I and the human insulin promoters. *Proc Natl Acad Sci U S A* *90*, 3865-3869.

Micallef, S.J., Janes, M.E., Knezevic, K., Davis, R.P., Elefanty, A.G., and Stanley, E.G. (2005). Retinoic acid induces Pdx1-positive endoderm in differentiating mouse embryonic stem cells. *Diabetes* *54*, 301-305.

Miller, C.P., McGehee, R.E., Jr., and Habener, J.F. (1994). IDX-1: a new homeodomain transcription factor expressed in rat pancreatic islets and duodenum that transactivates the somatostatin gene. *Embo J* *13*, 1145-1156.

Miralles, F., Czernichow, P., and Scharfmann, R. (1998). Follistatin regulates the relative proportions of endocrine versus exocrine tissue during pancreatic development. *Development* *125*, 1017-1024.

Mizuguchi, H., Xu, Z., Ishii-Watabe, A., Uchida, E., and Hayakawa, T. (2000). IRES-dependent second gene expression is significantly lower than cap-dependent first gene expression in a bicistronic vector. *Mol Ther* *1*, 376-382.

Mochida, Y., Parisuthiman, D., Kaku, M., Hanai, J., Sukhatme, V.P., and Yamauchi, M. (2006). Nephrocan, a novel member of the small leucine-rich repeat protein family, is an inhibitor of transforming growth factor-beta signaling. *J Biol Chem* *281*, 36044-36051.

Molotkov, A., Molotkova, N., and Duester, G. (2005). Retinoic acid generated by Raldh2 in mesoderm is required for mouse dorsal endodermal pancreas development. *Dev Dyn* *232*, 950-957.

- Monaghan, A.P., Kaestner, K.H., Grau, E., and Schutz, G. (1993). Postimplantation expression patterns indicate a role for the mouse forkhead/HNF-3 alpha, beta and gamma genes in determination of the definitive endoderm, chordamesoderm and neuroectoderm. *Development* 119, 567-578.
- Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5, 621-628.
- Nagai, T., Ibata, K., Park, E.S., Kubota, M., Mikoshiba, K., and Miyawaki, A. (2002). A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat Biotechnol* 20, 87-90.
- Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., and Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320, 1344-1349.
- Obermuller, S., Calegari, F., King, A., Lindqvist, A., Lundquist, I., Salehi, A., Francolini, M., Rosa, P., Rorsman, P., Huttner, W.B., *et al.* (2010). Defective secretion of islet hormones in chromogranin-B deficient mice. *PLoS One* 5, e8936.
- Offield, M.F., Jetton, T.L., Labosky, P.A., Ray, M., Stein, R.W., Magnuson, M.A., Hogan, B.L., and Wright, C.V. (1996). PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. *Development* 122, 983-995.
- Ohlsson, H., Karlsson, K., and Edlund, T. (1993). IPF1, a homeodomain-containing transactivator of the insulin gene. *Embo J* 12, 4251-4259.
- Ohlsson, H., Thor, S., and Edlund, T. (1991). Novel insulin promoter- and enhancer-binding proteins that discriminate between pancreatic alpha- and beta-cells. *Mol Endocrinol* 5, 897-904.
- Oikawa, T., Kamiya, A., Kakinuma, S., Zeniya, M., Nishinakamura, R., Tajiri, H., and Nakauchi, H. (2009). Sall4 regulates cell fate decision in fetal hepatic stem/progenitor cells. *Gastroenterology* 136, 1000-1011.
- Orkin, S.H., and Zon, L.I. (2002). Hematopoiesis and stem cells: plasticity versus developmental heterogeneity. *Nat Immunol* 3, 323-328.
- Pan, F.C., and Wright, C. (2011). Pancreas organogenesis: from bud to plexus to gland. *Dev Dyn* 240, 530-565.
- Pang, K., Mukonoweshuro, C., and Wong, G.G. (1994). Beta cells arise from glucose transporter type 2 (Glut2)-expressing epithelial cells of the developing rat pancreas. *Proc Natl Acad Sci U S A* 91, 9559-9563.
- Phillips, B.W., Hentze, H., Rust, W.L., Chen, Q.P., Chipperfield, H., Tan, E.K., Abraham, S., Sadasivam, A., Soong, P.L., Wang, S.T., *et al.* (2007). Directed differentiation of human embryonic stem cells into the pancreatic endocrine lineage. *Stem Cells Dev* 16, 561-578.

- Potter, L.A., Choi, E., Hipkens, S.B., Wright, C.V., and Magnuson, M.A. (2011). An RMCE-derived cyan fluorescent protein reporter allele for Pdx1. *Genesis*.
- Rausa, F., Samadani, U., Ye, H., Lim, L., Fletcher, C.F., Jenkins, N.A., Copeland, N.G., and Costa, R.H. (1997). The cut-homeodomain transcriptional activator HNF-6 is coexpressed with its target gene HNF-3 beta in the developing murine liver and pancreas. *Dev Biol* 192, 228-246.
- Rizzo, M.A., Springer, G.H., Granada, B., and Piston, D.W. (2004). An improved cyan fluorescent protein variant useful for FRET. *Nat Biotechnol* 22, 445-449.
- Robertson, J.A. (2010). Embryo stem cell research: ten years of controversy. *J Law Med Ethics* 38, 191-203.
- Rodaway, A., and Patient, R. (2001). Mesendoderm. an ancient germ layer? *Cell* 105, 169-172.
- Rodriguez, C.I., Buchholz, F., Galloway, J., Sequerra, R., Kasper, J., Ayala, R., Stewart, A.F., and Dymecki, S.M. (2000). High-efficiency deleter mice show that FLPe is an alternative to Cre-loxP. *Nat Genet* 25, 139-140.
- Rose, S.D., Swift, G.H., Peyton, M.J., Hammer, R.E., and MacDonald, R.J. (2001). The role of PTF1-P48 in pancreatic acinar gene expression. *J Biol Chem* 276, 44018-44026.
- Rossi, J.M., Dunn, N.R., Hogan, B.L., and Zaret, K.S. (2001). Distinct mesodermal signals, including BMPs from the septum transversum mesenchyme, are required in combination for hepatogenesis from the endoderm. *Genes Dev* 15, 1998-2009.
- Roux, E., Strubin, M., Hagenbuchle, O., and Wellauer, P.K. (1989). The cell-specific transcription factor PTF1 contains two different subunits that interact with the DNA. *Genes Dev* 3, 1613-1624.
- Ruby, K.M., and Zheng, B. (2009). Gene targeting in a HUES line of human embryonic stem cells via electroporation. *Stem Cells* 27, 1496-1506.
- Ryan, E.A., Lakey, J.R., Rajotte, R.V., Korbutt, G.S., Kin, T., Imes, S., Rabinovitch, A., Elliott, J.F., Bigam, D., Kneteman, N.M., et al. (2001). Clinical outcomes and insulin secretion after islet transplantation with the Edmonton protocol. *Diabetes* 50, 710-719.
- Ryan, E.A., Paty, B.W., Senior, P.A., Bigam, D., Alfadhli, E., Kneteman, N.M., Lakey, J.R., and Shapiro, A.M. (2005). Five-year follow-up after clinical islet transplantation. *Diabetes* 54, 2060-2069.
- Saha, S., Sparks, A.B., Rago, C., Akmaev, V., Wang, C.J., Vogelstein, B., Kinzler, K.W., and Velculescu, V.E. (2002). Using the transcriptome to annotate the genome. *Nat Biotechnol* 20, 508-512.
- Sancho, E., Batlle, E., and Clevers, H. (2004). Signaling pathways in intestinal development and cancer. *Annu Rev Cell Dev Biol* 20, 695-723.

- Sanvito, F., Herrera, P.L., Huarte, J., Nichols, A., Montesano, R., Orci, L., and Vassalli, J.D. (1994). TGF-beta 1 influences the relative development of the exocrine and endocrine pancreas in vitro. *Development* 120, 3451-3462.
- Sasaki, H., and Hogan, B.L. (1993). Differential expression of multiple fork head related genes during gastrulation and axial pattern formation in the mouse embryo. *Development* 118, 47-59.
- Schaefer, L., and Iozzo, R.V. (2008). Biological functions of the small leucine-rich proteoglycans: from genetics to signal transduction. *J Biol Chem* 283, 21305-21309.
- Scharp, D.W., Lacy, P.E., Santiago, J.V., McCullough, C.S., Weide, L.G., Falqui, L., Marchetti, P., Gingerich, R.L., Jaffe, A.S., Cryer, P.E., et al. (1990). Insulin independence after islet transplantation into type I diabetic patient. *Diabetes* 39, 515-518.
- Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467-470.
- Scheres, B. (2007). Stem-cell niches: nursery rhymes across kingdoms. *Nat Rev Mol Cell Biol* 8, 345-354.
- Schonherr, E., Broszat, M., Brandan, E., Bruckner, P., and Kresse, H. (1998). Decorin core protein fragment Leu155-Val260 interacts with TGF-beta but does not compete for decorin binding to type I collagen. *Arch Biochem Biophys* 355, 241-248.
- 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.
- Segev, H., Fishman, B., Ziskind, A., Shulman, M., and Itsikovitz-Eldor, J. (2004). Differentiation of human embryonic stem cells into insulin-producing clusters. *Stem Cells* 22, 265-274.
- 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. *Proc Natl Acad Sci U S A* 104, 1865-1870.
- Shaner, N.C., Campbell, R.E., Steinbach, P.A., Giepmans, B.N., Palmer, A.E., and Tsien, R.Y. (2004). Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat Biotechnol* 22, 1567-1572.
- Shapiro, A.M., Lakey, J.R., Ryan, E.A., Korbett, 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. *N Engl J Med* 343, 230-238.
- Shendure, J., Porreca, G.J., Reppas, N.B., Lin, X., McCutcheon, J.P., Rosenbaum, A.M., Wang, M.D., Zhang, K., Mitra, R.D., and Church, G.M. (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* 309, 1728-1732.

- Sherwood, R.I., Jitianu, C., Cleaver, O., Shaywitz, D.A., Lamenzo, J.O., Chen, A.E., Golub, T.R., and Melton, D.A. (2007). Prospective isolation and global gene expression analysis of definitive and visceral endoderm. *Dev Biol* *304*, 541-555.
- Shim, J.H., Kim, S.E., Woo, D.H., Kim, S.K., Oh, C.H., McKay, R., and Kim, J.H. (2007). Directed differentiation of human embryonic stem cells towards a pancreatic cell fate. *Diabetologia* *50*, 1228-1238.
- Shimomura, O., Johnson, F.H., and Saiga, Y. (1962). Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea*. *J Cell Comp Physiol* *59*, 223-239.
- Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., Kodzius, R., Watahiki, A., Nakamura, M., Arakawa, T., et al. (2003). Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc Natl Acad Sci U S A* *100*, 15776-15781.
- Smithies, O., Gregg, R.G., Boggs, S.S., Koralewski, M.A., and Kucherlapati, R.S. (1985). Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. *Nature* *317*, 230-234.
- Sommer, L., Hagenbuchle, O., Wellauer, P.K., and Strubin, M. (1991). Nuclear targeting of the transcription factor PTF1 is mediated by a protein subunit that does not bind to the PTF1 cognate sequence. *Cell* *67*, 987-994.
- Stafford, D., Hornbruch, A., Mueller, P.R., and Prince, V.E. (2004). A conserved role for retinoid signaling in vertebrate pancreas development. *Dev Genes Evol* *214*, 432-441.
- Stanger, B.Z., Tanaka, A.J., and Melton, D.A. (2007). Organ size is limited by the number of embryonic progenitor cells in the pancreas but not the liver. *Nature* *445*, 886-891.
- Stoffers, D.A., Ferrer, J., Clarke, W.L., and Habener, J.F. (1997a). Early-onset type-II diabetes mellitus (MODY4) linked to IPF1. *Nat Genet* *17*, 138-139.
- Stoffers, D.A., Zinkin, N.T., Stanojevic, V., Clarke, W.L., and Habener, J.F. (1997b). Pancreatic agenesis attributable to a single nucleotide deletion in the human IPF1 gene coding sequence. *Nat Genet* *15*, 106-110.
- Su, A.I., Cooke, M.P., Ching, K.A., Hakak, Y., Walker, J.R., Wiltshire, T., Orth, A.P., Vega, R.G., Sapinoso, L.M., Moqrich, A., et al. (2002). Large-scale analysis of the human and mouse transcriptomes. *Proc Natl Acad Sci U S A* *99*, 4465-4470.
- 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. *Proc Natl Acad Sci U S A* *104*, 175-180.
- Sultan, M., Schulz, M.H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., Seifert, M., Borodina, T., Soldatov, A., Parkhomchuk, D., et al. (2008). A global view of gene activity

and alternative splicing by deep sequencing of the human transcriptome. *Science* 321, 956-960.

Sumazaki, R., Shiojiri, N., Isoyama, S., Masu, M., Keino-Masu, K., Osawa, M., Nakuchi, H., Kageyama, R., and Matsui, A. (2004). Conversion of biliary system to pancreatic tissue in Hes1-deficient mice. *Nat Genet* 36, 83-87.

Svensson, P., Williams, C., Lundeberg, J., Ryden, P., Bergqvist, I., and Edlund, H. (2007). Gene array identification of *Ipf1/Pdx1*-/- regulated genes in pancreatic progenitor cells. *BMC Dev Biol* 7, 129.

Takeuchi, Y., Kodama, Y., and Matsumoto, T. (1994). Bone matrix decorin binds transforming growth factor-beta and enhances its bioactivity. *J Biol Chem* 269, 32634-32638.

Thomas, K.R., Folger, K.R., and Capecchi, M.R. (1986). High frequency targeting of genes to specific sites in the mammalian genome. *Cell* 44, 419-428.

Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., and Narechania, A. (2003). PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res* 13, 2129-2141.

Thomson, J.A., Itsikovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. *Science* 282, 1145-1147.

Tremblay, K.D., Hoodless, P.A., Bikoff, E.K., and Robertson, E.J. (2000). Formation of the definitive endoderm in mouse is a Smad2-dependent process. *Development* 127, 3079-3090.

Urbach, A., Schuldiner, M., and Benvenisty, N. (2004). Modeling for Lesch-Nyhan disease by gene targeting in human embryonic stem cells. *Stem Cells* 22, 635-641.

Velculescu, V.E., Zhang, L., Vogelstein, B., and Kinzler, K.W. (1995). Serial analysis of gene expression. *Science* 270, 484-487.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., *et al.* (2001). The sequence of the human genome. *Science* 291, 1304-1351.

Vincent, S.D., Dunn, N.R., Hayashi, S., Norris, D.P., and Robertson, E.J. (2003). Cell fate decisions within the mouse organizer are governed by graded Nodal signals. *Genes Dev* 17, 1646-1662.

Viotti, M., Nowotschin, S., and Hadjantonakis, A.K. (2011). Afp::mCherry, a red fluorescent transgenic reporter of the mouse visceral endoderm. *Genesis* 49, 124-133.

- Wagner, E.F., Stewart, T.A., and Mintz, B. (1981). The human beta-globin gene and a functional viral thymidine kinase gene in developing mice. *Proc Natl Acad Sci U S A* 78, 5016-5020.
- Wandzioch, E., and Zaret, K.S. (2009). Dynamic signaling network for the specification of embryonic pancreas and liver progenitors. *Science* 324, 1707-1710.
- Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10, 57-63.
- Watanabe, K., Ueno, M., Kamiya, D., Nishiyama, A., Matsumura, M., Wataya, T., Takahashi, J.B., Nishikawa, S., Muguruma, K., and Sasai, Y. (2007). A ROCK inhibitor permits survival of dissociated human embryonic stem cells. *Nat Biotechnol* 25, 681-686.
- Wells, J.M., and Melton, D.A. (1999). Vertebrate endoderm development. *Annu Rev Cell Dev Biol* 15, 393-410.
- Wells, J.M., and Melton, D.A. (2000). Early mouse endoderm is patterned by soluble factors from adjacent germ layers. *Development* 127, 1563-1572.
- Westwood, J.A., Jones, I.M., and Bishop, D.H. (1993). Analyses of alternative poly(A) signals for use in baculovirus expression vectors. *Virology* 195, 90-99.
- 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.
- Wiebe, P.O., Kormish, J.D., Roper, V.T., Fujitani, Y., Alston, N.I., Zaret, K.S., Wright, C.V., Stein, R.W., and Gannon, M. (2007). Ptfla binds to and activates area III, a highly conserved region of the Pdx1 promoter that mediates early pancreas-wide Pdx1 expression. *Mol Cell Biol* 27, 4093-4104.
- Wilson, M.E., Kalamaras, J.A., and German, M.S. (2002). Expression pattern of IAPP and prohormone convertase 1/3 reveals a distinctive set of endocrine cells in the embryonic pancreas. *Mech Dev* 115, 171-176.
- Wilson, M.E., Scheel, D., and German, M.S. (2003). Gene expression cascades in pancreatic development. *Mech Dev* 120, 65-80.
- Wright, C.V., Schnegelsberg, P., and De Robertis, E.M. (1989). Xlhbox 8: a novel Xenopus homeo protein restricted to a narrow band of endoderm. *Development* 105, 787-794.
- Wu, K.L., Gannon, M., Peshavaria, M., Offield, M.F., Henderson, E., Ray, M., Marks, A., Gamer, L.W., Wright, C.V., and Stein, R. (1997). Hepatocyte nuclear factor 3beta is involved in pancreatic beta-cell-specific transcription of the pdx-1 gene. *Mol Cell Biol* 17, 6002-6013.
- Xu, X., D'Hoker, J., Stange, G., Bonne, S., De Leu, N., Xiao, X., Van de Casteele, M., 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.

- Yanez, R.J., and Porter, A.C. (2002). A chromosomal position effect on gene targeting in human cells. *Nucleic Acids Res* 30, 4892-4901.
- Yang, L., Soonpaa, M.H., Adler, E.D., Roepke, T.K., Kattman, S.J., Kennedy, M., Henckaerts, E., Bonham, K., Abbott, G.W., Linden, R.M., *et al.* (2008). Human cardiovascular progenitor cells develop from a KDR+ embryonic-stem-cell-derived population. *Nature* 453, 524-528.
- Yasunaga, M., Tada, S., Torikai-Nishikawa, S., Nakano, Y., Okada, M., Jakt, L.M., Nishikawa, S., Chiba, T., and Era, T. (2005). Induction and monitoring of definitive and visceral endoderm differentiation of mouse ES cells. *Nat Biotechnol* 23, 1542-1550.
- Ye, F., Duvillie, B., and Scharfmann, R. (2005). Fibroblast growth factors 7 and 10 are expressed in the human embryonic pancreatic mesenchyme and promote the proliferation of embryonic pancreatic epithelial cells. *Diabetologia* 48, 277-281.
- Yoshitomi, H., and Zaret, K.S. (2004). Endothelial cell interactions initiate dorsal pancreas development by selectively inducing the transcription factor Ptf1a. *Development* 131, 807-817.
- 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. *Proc Natl Acad Sci U S A* 94, 3789-3794.
- Zheng, K., Wu, X., Kaestner, K.H., and Wang, P.J. (2009). The pluripotency factor LIN28 marks undifferentiated spermatogonia in mouse. *BMC Dev Biol* 9, 38.
- Zhou, Q., Law, A.C., Rajagopal, J., Anderson, W.J., Gray, P.A., and Melton, D.A. (2007). A multipotent progenitor domain guides pancreatic organogenesis. *Dev Cell* 13, 103-114.
- Zwaka, T.P., and Thomson, J.A. (2003). Homologous recombination in human embryonic stem cells. *Nat Biotechnol* 21, 319-321.