The role of transcription factor *GATA6* in the development of the human pancreas

Crystal Ying Chia

Homerton College

University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy

July 2017

Acknowledgements

"Follow your dreams", my parents always reminded me.

Growing up, I had never imagined I would do a PhD, much less in Cambridge. Being enrolled at the University of Cambridge to read a doctoral degree is a dream come true for me. The path to getting to where I am today is attributed to a mixture of hard work and luck in meeting the right people who not only opened doors and made this once in a lifetime opportunity possible, but also believed in me and supported me throughout this journey. There are many people for whom I have heartfelt gratitude, but sadly, it is only possible to particularly thank some here.

To Professor Ludovic Vallier, my Cambridge supervisor, who has been instrumental in driving the progress of my PhD project from start to finish. Thank you for accepting me as your student and for providing me with this valuable opportunity. You have been a great mentor who has taught me so much. Thank you for all the guidance, understanding and support you have shown me over the years, and for moulding my development not only in stem cell biology, but as a scientist. You have provided me with such an all-rounded PhD experience, from teaching me important set of skills for my scientific career, to exploring Europe and USA on the annual lab retreats and conferences.

To Dr Norris Ray Dunn, my Singapore supervisor, who has moved mountains for me. Thank you for accepting me as your student and for providing me with this valuable opportunity. I am truly lucky and blessed to have you as a mentor, and you have been nothing but supportive, encouraging and caring. Thank you for imparting valuable skills to me not only in research, but also in writing and presentations. Thank you for making my PhD experience such an amazing and wholesome one and for teaching me so much that is not only important for my scientific career, but also in life.

To my family, thank you for all the encouragement, support, care and concern. To my parents, for giving me the freedom and opportunities to pursue all my desires in life. To my brother and sister-in-law, for always being there no matter the distance.

To my friends and colleagues in Cambridge, Li Meng, Kasia, Sapna, Hiroko- you have been there for me since day one. Thank you for being my pillars of support through the good and the bad times, and for making my life in Cambridge so wonderful.

To my friends in Singapore and Malaysia who have not forgotten me despite being away for two years, Atee, Jieying, Daphne, Xuefang and Anita. Thank you for your friendship over the many, many years and for always pushing me to excel in the things I do.

To CJ, thanks for all the motivation, lessons of patience and for being there for me during the final push of my PhD.

To A*STAR, for providing me with the opportunity to pursue a PhD degree by funding my course of study.

To members of the Vallier and NRD lab who have created a fun and lively place to work in. Thank you for all the help, support and friendship.

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

Crystal Ying Chia

July 2017

Statement of Length

This dissertation does not exceed the word limit of 60,000 words excluding figures, photographs, tables, appendices and bibliography for the Biology Degree Committee.

List of Publications and Presentations

Work from this thesis contributed to the following publications and presentations:

Academic publication:

Chia, C.Y., Madrigal, P., Denil, S.L.I.J., El-Khairi, R., Chhatriwala, M., Hattersley, A.T., N. Dunn, N.R., and Vallier, L.. GATA6 cooperates with EOMES/SMAD2/3 to deploy the gene regulatory network governing human definitive endoderm and pancreas formation. To be submitted (2017).

Poster presentations:

Chia, C.Y., Chhatriwala, M., Dunn, N.R., and Vallier, L.. (2014). The role of transcription factor GATA6 in human pancreas and liver. Poster presented at the Keystone Symposia Conference: Emerging Concepts and Targets in Islet Biology held from April 6 – 11, 2014, Colorado, USA.

Chia, C.Y., Chhatriwala, M., Dunn, N.R., and Vallier, L.. (2014). The role of transcription factor GATA6 in human pancreas and liver. Poster presented at the Company of Biologist Workshops: From Stem Cells to Human Development held from September 21 – 24, 2014, Wotton House, Surrey, UK.

Chia, C.Y., Chhatriwala, M., Dunn, N.R., and Vallier, L.. (2014). The role of transcription factor GATA6 in human pancreas and liver. Poster presented at the Wellcome Trust - Medical Research Council Cambridge Stem Cell Institute: SCI PhD Day held on July 22, 2014, Cambridge, UK.

Summary

While there has been an opulence of data and studies surrounding the study of the developing pancreas in mammals and other vertebrates, the focus has largely been in mice. The paucity of research in the development of the human pancreas has led to diminished knowledge in the area, compared to other species. Recent discoveries provide growing evidence for discrepancies between mouse and human pancreatic development and diseases and highlight the fact that developmental studies of the pancreas in humans are imperative. The need to develop therapies for diabetes, a growing and one of the leading health problems worldwide, further compels more exploration in this area to deepen our understanding in the different aspects of diabetes in humans and its underlying causes.

Research involving modelling human diseases *in vitro* enables the investigation of the cellular and molecular mechanisms underlying these diseases as well as the development of therapies for treating them. The availability of hPSCs brings with it the advantage of overcoming the limitations of animal models for certain disorders such as pancreatic agenesis, the focus of my project. The use of site-specific nucleases such as TALENs for such a purpose represents a paradigm shift in disease modelling, where TALENs are capable of directly correcting disease-causing mutations, therefore permanently eliminating the symptoms with precise genome modifications. Alternatively, TALENs can also be used to inactivate specific genes by inducing site-specific mutations.

Using these tools, I found that *GATA6* is required for the formation of the definitive endoderm and pancreas in humans; hPSCs harbouring homozygous *GATA6* mutations fail to form the definitive endoderm, and consequently the pancreas, whereas hPSCs harbouring heterozygous *GATA6* mutations exhibited impairment in definitive endoderm development, although it remains unclear if this is a protocol-dependent defect. At the pancreatic stage, heterozygous *GATA6* mutations consistently compromised pancreas formation regardless of protocol used. I also found that *GATA6* transcriptionally activates the development of the definitive

endoderm and pancreatic endoderm, and possibly represses the development of mesoderm. Furthermore, I also established that *GATA6* directly interacts with key definitive endoderm markers *CXCR4* and *SOX17*, and pancreatic marker *PDX1*.

Taken together, the work herein demonstrates the successful use of hPSCs coupled with the TALEN genome editing technology as a unique *in vitro* system for disease modelling. These findings also establish two developmental windows, the DE and pancreatic progenitor stages, where *GATA6* haploinsufficiency can result in the impairment of pancreatic development leading to pancreatic hypoplasia observed in human *GATA6* heterozygous patients. Lastly, my work also provides the molecular mechanism by which *GATA6* regulates pancreatic development.

Overall, this study provided new insights in the role of *GATA6* during development of the human pancreas. These results will be important in developing new methods of differentiation for hPSCs and understanding the interconnection between early organogenesis and late onset of diabetes.

Table of Contents

List of Abbreviations

ABCC8	ATP-binding cassette, sub-family C, member 8
ADV-BSA	Advanced DMEM/F-12 containing bovine serum albumin
AIP	Anterior intestinal portal
bFGF	Basic fibroblast growth factor
BNZ	6-Bnz-cAMP sodium salt
DE	Definitive endoderm
CDM	Chemically defined media
CDM-PVA	Chemically defined medium-poly vinyl alcohol
CEL	Carboxyl ester lipase
ChIP	Chromatin immunoprecipitation
CRISPR	Clustered regulatory interspaced short palindromic repeats
CS	Carnegie stages
CXCR4	Chemokine (C-X-C motif) receptor 4
DE	Definitive endoderm
Dpc	Days post-conception
DSB	Double-stranded break
E8	Essential 8
EIF2AK3	Eukaryotic translation initiation factor 2 alpha kinase 3
ELISA	Enzyme linked immunosorbent assay
EMT	Epithelial-mesenchymal transition
EP	Endocrine progenitors
ESC	Embryonic stem cell
FACS	Fluorescence activated cell sorting
FLASH	Fast Ligation-based Automatable Solid-phase High-throughput

FOX	Forkhead box
GATA	GATA binding protein
GCK	Glucokinase
GLIS3	GLIS family zinc finger 3
GO	Gene ontology
GSC	Goosecoid
GSIS	Glucose-stimulated insulin secretion
GRN	Gene regulatory network
hESC	Human embryonic stem cell
HGF	Hepatocyte growth factor
hiPSC	Human induced pluripotent stem cell
HIPSCI	Human induced pluripotent stem cell initiative
hPSC	Human pluripotent stem cell
HMG	High mobility group
HNF	Hepatocyte nuclear factor
HR	Homologous recombination
HRP	Horseradish peroxidase
ICA	Iterative Capped Assembly
ICC	Immunocytochemistry
ICM	Inner cell mass
INS	Insulin
iPSC	Induced pluripotent stem cells
KCNJ11	Potassium inwardly-rectifying channel, subfamily J, member 11
LIF	Leukaemia inhibitory factor
MEF	Mouse embryonic fibroblasts
MODY	Maturity onset diabetes of the young

NDM	Neonatal diabetes mellitus
NEUROD1	Neuronal differentiation 1
NGN	Neurogenin
NHEJ	Non-homologous end-joining
NKX	Nirenberg and Kim homeobox factor
ORF	Open reading frame
OSM	Oncostatin M
PBMC	Peripheral blood mononuclear cell
PCR	Polymerase chain reaction
PDX1	Pancreatic and duodenal homeobox factor 1
PE	Pancreatic endoderm
PFA	Paraformaldehyde
PMSF	Phenylmethylsulfonylfluoride
PNDM	Permanent neonatal diabetes mellitus
PPP	Partial protein product
PSC	Pluripotent stem cell
PTF1A	Pancreas transcription factor 1A
qRT-PCR	Quantitative real time polymerase chain reaction
RA	Retinoic acid
RFX6	Regulatory factor X6
RVD	Repeat variable di-residues
SCT	Stem Cell Technologies
SHH	Sonic hedgehog
SOX	SRY (sex determining region Y)-box
T1D	Type 1 diabetes
T2D	Type 2 diabetes

TALEN	Transcription activator-like effector nuclease
TF	Transcription factor
TGF-β	Transforming growth factor-β
TNDM	Transient neonatal diabetes mellitus
TSS	Transcription start site
Wp	Well plate
Wpc	Weeks post-conception
ZFN	Zinc finger nuclease

List of Figures

List of Tables

CHAPTER 1 INTRODUCTION

1.1. Overview of the pancreas

The pancreas is a glandular organ originating from two separate primordia, the dorsal and ventral buds that arise from either side of the distal foregut endoderm. The organ is made up of a variety of distinct cell types and has a mix of both exocrine and endocrine functions.

The exocrine gland serves as a digestive organ with its acinar cells performing gastrointestinal functions by secreting digestive enzymes and a duct system that allows these digestive enzymes to drain into the intestine.

The endocrine gland is arranged as cell clusters known as islets of Langerhans which functions in regulating blood glucose homeostasis and other hormone secretions. Each cluster comprises of multiple distinct cell types, each secreting unique hormones into the body's circulation (α-cells, glucagon to increase blood glucose; β-cells, insulin to decrease blood glucose; δ-cells, somatostatin which regulates α-cells and β-cells; ε-cells, ghrelin which stimulates hunger and functions as a neuropeptide in the central nervous system; and γ or pancreatic polypeptide (Rashid et al.)-cells, pancreatic polypeptide which regulates pancreatic secretions, hepatic glycogen levels and gastrointestinal secretions). (Figure 1, adapted from (Shih et al., 2013)).

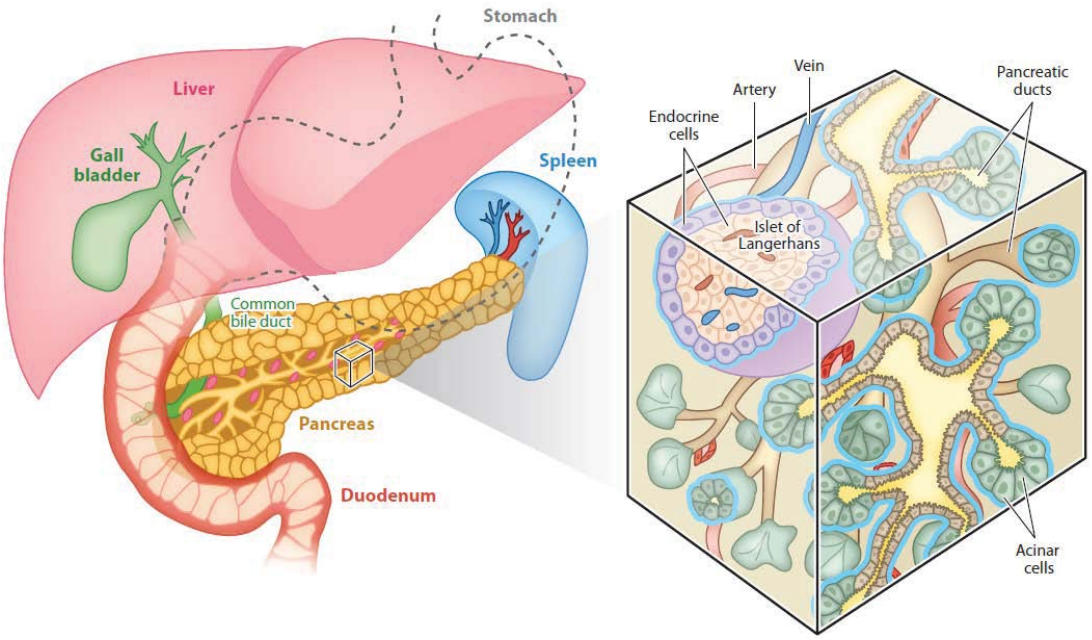


Figure 1. **The pancreas as a mixed exocrine and endocrine organ.** The mature pancreas lies behind the stomach and is adjacent to the duodenum. Acinar cells form the exocrine pancreas. The endocrine pancreas consists of small cell clusters, called islets of Langerhans, containing five endocrine cell types. Adapted from (Shih et al., 2013).

1.1.1. Development of the human pancreas

Human embryogenesis spans from fertilisation to approximately 8 weeks post-conception. After which, the embryo is referred to as a foetus. During the development of the embryo, specification of the three germ layers: ectoderm, mesoderm and endoderm occur, from which all adult tissues are formed. A recent publication by O'Rahilly and Müller has proposed a staging classification covering embryonic development (O'Rahilly and Müller, 2010). Based on a morphological scheme and staged by extension of time i.e. days post-conception (dpc), human embryonic development was divided into 23 different Carnegie Stages (CS). The key developmental stages of the pancreas during human embryogenesis, along with the approximate equivalent stage of mouse development, are mapped onto the 23 different CS stages (CS12 to CS23) (Table 1).

Human embryonic stage	Approximate days post-conception (dpc)	Examples of morphological features	Key events in human embryonic pancreas development	Approximate equivalent stage of mouse development*	Key transcription factors expressed
CS12	29-31	Lens and otic placodes, caudal neuropore closing, 1st-3rd pharyngeal arches	First detection of PDX1 in presumptive pancreatic endoderm	E9-E9.5	GATA6, GATA4, FOXA2, PDX1
CS13	30-33	Early sign of upper limb bud	Clear dorsal and ventral pancreatic buds	E9.5-E10	
CS14	33-35	Upper and lower limb buds clearly visible		E10-E11.5	
CS15	35-37	Hand plate now visible		E11.5-E12.25	
CS16	37-40	Clear retinal pigment, auricular hillocks, foot plate visible	Growth of organ and proliferation of multipotent pancreatic progenitors	E12.25-E12.75	GATA6, GATA4, FOXA2, PDX1, SOX9, NKX6-1
CS17	39-42	Digital rays first visible in hand plate		E12.75-E13.25	
CS18	42-45	Digital rays first visible in foot plate		E13.25-E14	
CS19	45-47	Clearly notched hand plate	Distinction possible between central trunk cells and peripheral tip cells, e.g. GATA4 levels	E14-E14.5	
CS20	47-50	Clearly notched foot plate, webbed fingers, scalp vascular plexus visible		E14.5-E15	
CS21	49-52	Visible fingers, webbed toes, scalp vascular plexus halfway up head	Onset of detection of *NEUROG3* and first detection of insulin-positive cell (i.e. signs of endocrine commitment)	E15-E15.5	NGN3 (transient)
CS22	52-55	Scalp plexus two-thirds of the way up head, separated fingers	Ventral bud largely rotated around the gut and becomes opposed with the dorsal bud	E15.5-E16	PDX1, MAFA, NKX6-1, ISL1
CS23	53-58	Scalp vascular plexus at top of head, separated toes		E16-E16.5	

Table 1. Stages of human pancreas development and their respective Carnegie Stages (CS). CS stages are shown together with their estimates of corresponding days post-conception (dpc). Key events in human embryonic pancreas development along with the approximate equivalent stage of mouse development are mapped to the CS stages. Table edited and adapted from O'Rahilly and Müller (2010).

Pancreas organogenesis is a highly complex and orchestrated process, comprising of coordinated signalling events that occur in a step-wise manner, as well as transcriptional networks that result in a cascade of transcription factors driving pancreatic specification (Figure 2 adapted from (Jennings et al., 2015)).

Pancreas induction occurs at CS9, where the definitive endoderm (DE) maintains communication with the visceral endoderm of the yolk sac (Jennings et al., 2013), and ventral and dorsal thickenings of the epithelial cells in the distal foregut occurs (Piper et al., 2002, Piper et al., 2004).

At CS10, endodermal folding gives rise to the foregut and hindgut, thus restricting the opening of the yolk sac to the intervening midgut (Jennings et al., 2013). The anterior end of this opening, known as the anterior intestinal portal (AIP), constitutes the foregut-midgut boundary and is the site of pancreatic specification. In other species, early specification of the pancreas within the gut endoderm occurs in the absence of sonic hedgehog (Shh) signalling (Apelqvist et al., 1997, Kim and Melton, 1998, Hebrok et al., 2000). In chick embryos, Activin secreted from the notochord and the close proximity of the resulting dorsal foregut endoderm to it causes exclusion of Shh expression, allowing for the expression of key transcription factor pancreatic and duodenal homeobox factor 1 (*Pdx1*) (Kim and Melton, 1998, Hebrok et al., 1998). In humans, patterning was similarly observed where *PDX1* was first detected at CS12, even though SHH could still be detected at CS10, which suggests a slightly later timing for the exclusion of SHH in humans (Jennings et al., 2013). The dorsal foregut endoderm subsequently develops into the dorsal pancreatic bud. One difference worth noting that was observed between human embryos and mouse or chick embryos is that in humans, there has been no detection of early pancreatic endocrine differentiation (Villasenor et al., 2008, Jennings et al., 2013), whereas early pro-endocrine patterning has been observed in mouse and chick (Lammert et al., 2001, Bonal and Herrera, 2008). This observation could possibly be explained by the lack of proximity of the paired dorsal aortae to the early pancreatic endoderm, thus reducing the opportunity for early pro-endocrine patterning.

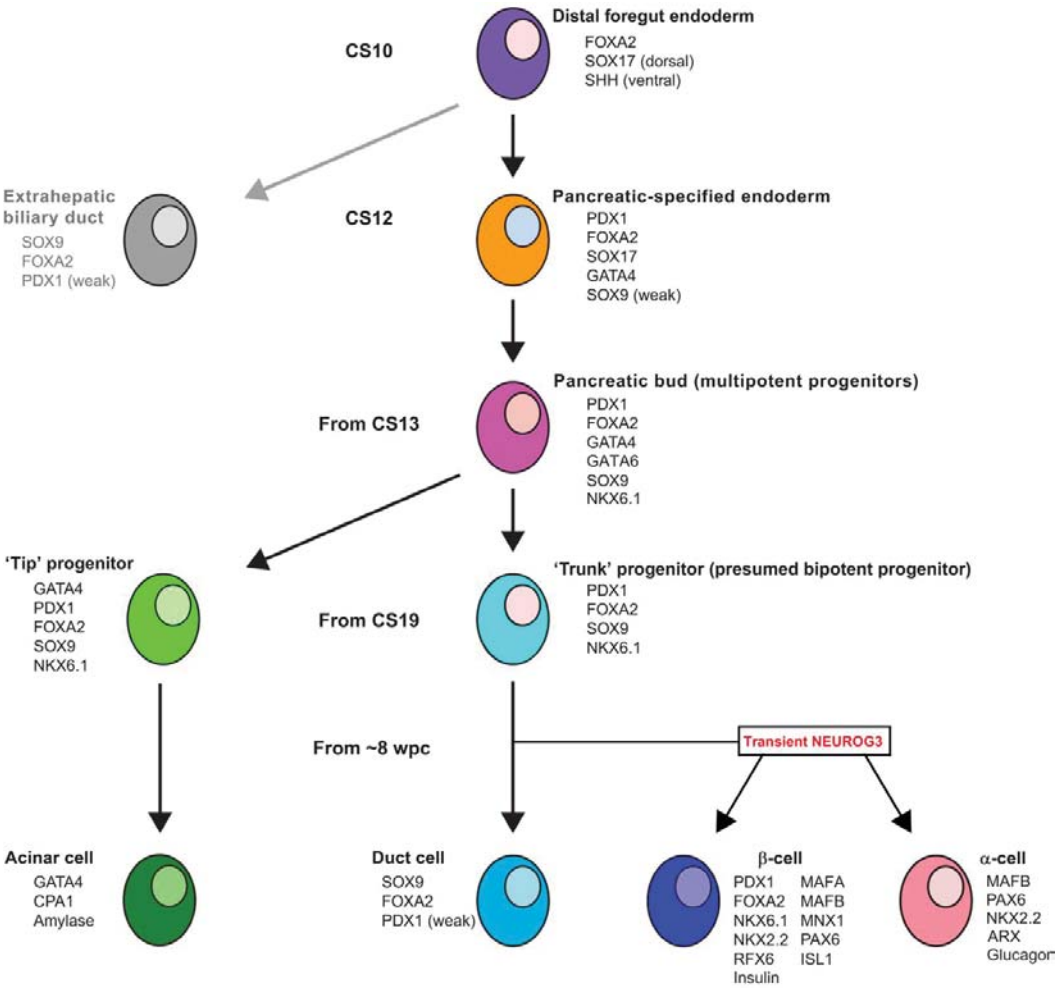


Figure 2. Developmental stages of human pancreas development and its respective transcription factor network. Illustration of transcription factors and key markers that identify the various key stages of early pancreas specification in the dorsal pancreas and commitment to subsequent lineages.

At CS13, the dorsal and ventral pancreatic buds are clearly defined and are marked by the transcription factors SRY (sex-determining region Y)-box 9 (*SOX9*), *PDX1*, GATA binding protein 4 (*GATA4*) and *GATA6* (Piper et al., 2004, Jennings et al., 2013), all of which play a pivotal role in promoting human pancreatic growth (Stoffers et al., 1997, Piper et al., 2002, Allen et al., 2012, Shaw-Smith et al., 2014). The human pancreas continues its expansion of proliferative progenitor cells for the remainder of the embryonic period. Another notable difference between the human and mouse pancreatic development is that transcription factor Nirenberg and Kim homeobox factor (NKX) 2.2 (*Nkx2-2*) is detected in these progenitor cells of the mouse but not human (Jennings et al., 2013).

At CS19, divergence into "tip" or "trunk" progenitor cells is marked by the expression levels of *GATA4* (Jennings et al., 2013). "Trunk" cells, which commit to central duct-like structures, express lower levels of *GATA4* as compared to the more peripheral clustered pro-acinar "tip" cells, even though both progenitor cell types express several common pancreatic markers such as *PDX1*, *SOX9* and *NKX6-1* (Figure 2). A similar separation of acinar cells was also observed in the mouse (Esni et al., 2004, Solar et al., 2009, Schaffer et al., 2010), although *Sox9* is lost more promptly in these peripheral tip cells in mouse (Schaffer et al., 2010) than in humans, where the loss of *SOX9* is delayed to between 10 and 14 weeks post-conception (wpc) (Jennings et al., 2013).

The commencement of endocrine specification is marked by the transient expression of transcription factor neurogenin 3 (*NGN3*) (Figure 2). In the mouse, *Ngn3* is transiently expressed to enable progenitor cells within the central duct-like structures to commit into the endocrine lineage (Gradwohl et al., 2000, Schwitzgebel et al., 2000, Gu et al., 2002). In humans, *NGN3* expression is detected at CS21 (8 wpc) around the end of the first trimester of human pregnancy upon the formation of foetal β-cells, which function as true endocrine cells by this time (Piper et al., 2004, Lyttle et al., 2008). The peak expression level of *NGN3* occurs at 10 to 14 wpc and declines at and after 18 wpc, and diminishes in human foetuses after 35 wpc (Salisbury et al., 2014). By contrast, *SOX9* is absent in cells robustly expressing *NGN3* and continues to be absent in subsequent endocrine cells, but is present in pancreatic duct cells (Jennings et al., 2013). By 10 wpc, β-cells are well vascularised and at 12-13 wpc, islets containing α-cells, β-cells, δ-cells and γ-cells are apparent (Piper et al., 2004, Jennings et al., 2013).

1.1.2. Diabetes Mellitus as a pancreatic disease

Diabetes Mellitus ('diabetes') represents a family of metabolic disease caused primarily by dysfunction in the pancreas. Diabetes is a growing health problem worldwide. The global prevalence of diabetes has nearly doubled since 1980, rising from 4.7% to 8.5% in the adult population and causing 1.5 million deaths in 2012 (World Health Organisation, 2016). It is estimated that 366 million people were diagnosed with diabetes in 2011; by 2030 this would have risen to 552 million (International Diabetes Federation, 2011). The need to understand human pancreas development is, therefore, critical because of its relevance to the different types of diabetes and therapies for this disease. There are multiple forms of diabetes, such as polygenic forms Type 1 diabetes (T1D), Type 2 diabetes (T2D), and monogenic forms maturity onset diabetes of the young (MODY), and neonatal diabetes mellitus (NDM). Polygenic forms of diabetes i.e. T1D and T2D make up an overwhelming majority (98%) of diabetic cases and its risk is related to multiple genes. Rare, monogenic forms of diabetes such as MODY and NDM result from mutations in a single gene and account for about 1 to 5% of all cases of diabetes in young people. Gene mutations in monogenic diabetes can either be inherited or spontaneous (*de novo*).

T1D, or juvenile-onset diabetes, accounts for approximately 5-10% of diabetic patients and is caused by the chronic autoimmune destruction of insulin-secreting β-cells, usually leading to complete insulin deficiency or hyperglycaemia (Anderson and Bluestone, 2005). Hyperglycaemia occurs when blood glucose levels are high, and this can lead to serious health conditions such as ketoacidosis, kidney failure, heart disease, stroke, and blindness. Despite being able to affect children and adult with normal weight, the childhood onset is most prevalent. Although the main effector mechanism of T1D is clearly an autoimmune reaction, T1D is also suggested to result from genetic susceptibility and/or environmental triggers (reviewed in (Atkinson and Eisenbarth, 2001, Van Belle et al., 2011)). T1D can be fatal if not treated with exogenous insulin to compensate for the lack of insulin production by the body.

T2D, or adult-onset diabetes, on the other hand, accounts for approximately 90% of diabetic patients and is usually associated with obesity or older age. T2D is characterised by insulin resistance, where insulin-sensitive target tissues such as the pancreatic β-cells, liver, muscle or adipose cells do not respond adequately to normal levels of insulin produced by intact β-cells. Consequently, this leads to disruption of the pancreatic β-cell function and decreased β-cell mass. Although T2D is most prevalent in adults, there has been increasing evidence of this form of diabetes affecting younger individuals (Fajans et al., 2001). T2D is a polygenic disease influenced by many environmental and behavioural risk factors. Thus, it has not only been challenging to identify the underlying genetic causes for this disease, but also to devise universal therapeutic strategies. No cure has yet been found for the disease. Several forms of treatment that currently exist, especially for obese patients; include lifestyle modifications, treatment of obesity, oral hypoglycaemic drugs, and insulin sensitizers such as metformin that reduces insulin resistance.

MODY was first recognised by Tattersall (Tattersall, 1974, Tattersall and Fajans, 1975) and is characterised by autosomal dominant inheritance, onset typically before 25 years of age and β-cell dysfunction leading to hyperglycaemia. The prevalence of MODY is higher than NDM, accounting for about 1% of people with diabetes in Europe (Frayling et al., 2001, Ledermann, 1995). Many MODY genes are pancreatic development transcription factors with the exception of glucokinase (*GCK*), acinar cell carboxyl ester lipase (*CEL*) and insulin (*INS*). Common MODY genes include MODY1 (Hepatic Nuclear Factor 4 Alpha; *HNF4A*-MODY), MODY2 (*GCK*-MODY), MODY3 (Hepatic Nuclear Factor 1 Alpha; *HNF1A*-MODY) which account for 70% of MODY cases, and MODY5 (Hepatic Nuclear Factor 1 Beta; *HNF1B*-MODY).

NDM is a rare, genetically heterogeneous monogenic form of diabetes occurring in approximately 1 in 200,000 live births (Stanik et al., 2007, Kanakatti Shankar et al., 2013). It presents in the days and weeks after birth, and almost always before 6 months of age (Iafusco et al., 2002, Edghill et al., 2006). NDM may be transient or permanent. In transient NDM (TNDM), the condition remits during infancy but may reappear later in life whereas in permanent NDM (PNDM),

hyperglycaemia persists during life. Approximately 50% of NDM patients have heterozygous activating mutations in the potassium inwardly-rectifying channel, subfamily J, member 11 (*KCNJ11*) and ATP-binding cassette, sub-family C, member 8 (*ABCC8*) genes encoding the adenosine triphosphate-sensitive potassium channel subunits (De Franco et al., 2015). Failure of the potassium channels to close appropriately in response to rising glucose, thus inhibiting the release of insulin from β-cells, leads to diabetes. Sulfonylurea therapy permits insulin secretion through closure of the channel (Pearson et al., 2006, Rafiq et al., 2008).

Other rare gene mutations leading to monogenic NDM include eukaryotic translation initiation factor 2 alpha kinase 3 (*EIF2AK3*), forkhead box P3 (*FOXP3*), *GATA6*, GLIS family zinc finger 3 (*GLIS3*), neuronal differentiation 1 (*NEUROD1*), *NGN3*, *PDX1*, pancreas specific transcription factor 1a (*PTF1A*), regulatory factor X6 (*RFX6*) and methylation defects at chromosome 6q24.

Studies of rare monogenic diseases provide an invaluable opportunity to learn about underlying molecular mechanisms, thereby contributing significantly to our understanding of the molecular genetic basis of common, complex diseases (Antonarakis and Beckmann, 2006).

1.1.3. Pancreatic agenesis

Congenital pancreatic agenesis is an extremely rare cause of NDM with a prevalence of less than 1/1 000 000 and around 50 cases being reported in the literature so far. It is caused by an impaired formation of the pancreas during embryonic development. Morphologically, the pancreas can either be totally absent or extremely reduced in size (pancreatic hypoplasia).

Clinically, pancreatic agenesis is defined as insulin dependent neonatal diabetes diagnosed before 6 months of age and pancreatic exocrine insufficiency requiring enzyme replacement therapy. Patients affected by pancreatic agenesis usually present with intrauterine growth retardation (IUGR) as a result of reduced insulin secretion *in utero* and are diagnosed with hyperglycaemia in the first days of life. Patients with pancreatic agenesis usually require insulin treatment.

Diagnosis of pancreatic agenesis can be made by imaging (MRI or ultrasound) showing reduction or absence of pancreatic tissue, measurement of fecal elastase which is often undetectable in patients with pancreatic agenesis as a result of exocrine dysfunction, or clinically by the presence of insulin-dependent neonatal diabetes and exocrine insufficiency requiring enzyme replacement therapy.

A genetic diagnosis is also possible for over 80% of patients with pancreatic agenesis and transcription factor *GATA6* has recently been identified to be a major cause of pancreatic agenesis (Lango Allen et al., 2012). In these patients, pancreatic agenesis is commonly associated with other extrapancreatic malformations such as cardiac malformation, neurocognitive defects, hypothyroidism, gut abnormalities and gallbladder agenesis/biliary atresia (De Franco et al., 2013). These defects affect organs of endodermal origin, suggesting a defect in early embryonic differentiation.

1.2. Human pluripotent stem cells as an *in vitro* system to model the development of the human pancreas

Stem cells are cells with unique properties such as the capacity to self-renew indefinitely and the ability to differentiate into many diverse cell types. Being pluripotent, these cells are able to differentiate into all derivatives of the three primary germ layers, namely endoderm, mesoderm and ectoderm (Evans and Kaufman, 1981). This is in contrast to adult stem cells which are multipotent and more restrictive in their differentiation to various cell types (Suda et al., 1987, Zwaka and Thomson, 2005). With advances in pluripotent stem cell (PSC) technology, a large number of stem cells can now be expanded and maintained *in vitro* whilst retaining their unique properties (Suda et al., 1987, Solter, 2006). This allows for studies that were once difficult using primary tissues or biopsies to progress. In the clinical setting, human PSCs (hPSCs) bring vast potential in providing opportunities for treating and curing diseases.

hPSCs can be broadly categorised into two categories; embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs). ESCs were first derived from the inner cell mass (ICM) of the mouse embryo at the early post-implantation blastocyst stage by Evans and Kaufman, and Martin in 1981 (Evans and Kaufman, 1981, Martin, 1981). The ICM in the blastocyst is a transient pluripotent pool of cells that rapidly differentiates during gastrulation into the primary germ layers. They can be maintained indefinitely *in vitro* in their pluripotent state in the presence of cytokine leukaemia inhibitory factor (LIF) on a layer of mitotically inactivated mouse embryonic fibroblasts (MEF) feeder cells (Smith et al., 1988, Williams et al., 1988).

There was a significant lag before the first human ESC (hESC) line was successfully isolated from human blastocysts in 1998 (Thomson et al., 1998). This was largely due to the fact that human embryos were much more difficult to obtain than mouse embryos and the ethical dilemmas that accompanied it. Prior to this, the first primate ESC line from the blastocyst of a rhesus monkey was also isolated and successfully derived (Thomson et al., 1995). The derivation of hESCs paved the way to an accelerated expansion on stem cell research.

Approximately a decade later, pioneering studies describing human iPSCs (hiPSCs) emerged (Takahashi and Yamanaka, 2006, Takahashi et al., 2007, Gurdon and Melton, 2008). These studies showed that by introducing a cocktail of four specific transcription factors (SOX2, KLF4, c-MYC and OCT4) *ex-vivo*, differentiated fibroblasts could be converted to a pluripotent state resembling ESCs derived from the blastocyst ICM. Like ESCs, iPSCs also had the ability to form teratomas in mice (Wernig et al., 2007). iPSC technology has the advantage over ESCs in that it was able to circumvent the ethical issues associated with human embryos. Furthermore, since iPSCs can be derived from patients' cells, it brings with it the potential application of excluding immunosuppression treatments that are required in conventional cell replacement or transplantation therapies to prevent tissue rejection when cells are transferred between genetically different individuals. iPSC is therefore a robust and ethical way of re-programming differentiated cells to a pluripotent state. Similarly to ESCs, the iPSCs can then be directed, by growth factors important and specific for development and differentiation, to form functional differentiated cells of a variety of lineages. It has been suggested that hiPSCs and hESCs are functionally equivalent since they both utilise similar signalling pathways to maintain pluripotency and drive differentiation, and the mechanisms controlling the early cell fate decision of these pluripotent stem cells are similar (Vallier et al., 2009a).

Since their discovery, both hESCs and hiPSCs have proven to be powerful tools in biomedical research, overcoming ethical limitations in human embryonic development studies where access to human embryos is scarce. hPSCs have the potential to be used in disease models for studying the molecular basis of diseases, including genetically inherited human diseases (Yusa et al., 2011). It brings tremendous potential not only in disease modelling, but also in regenerative medicine, cell replacement therapy, drug testing and targeted gene-repair strategies, such as homologous recombination to repair genetic defects (Figure 3). Thus, they serve as ideal model systems for human developmental scientific studies. This dissertation focuses on using hPSCs to model the development of the human pancreas by specifying the cells down the pancreatic lineage.

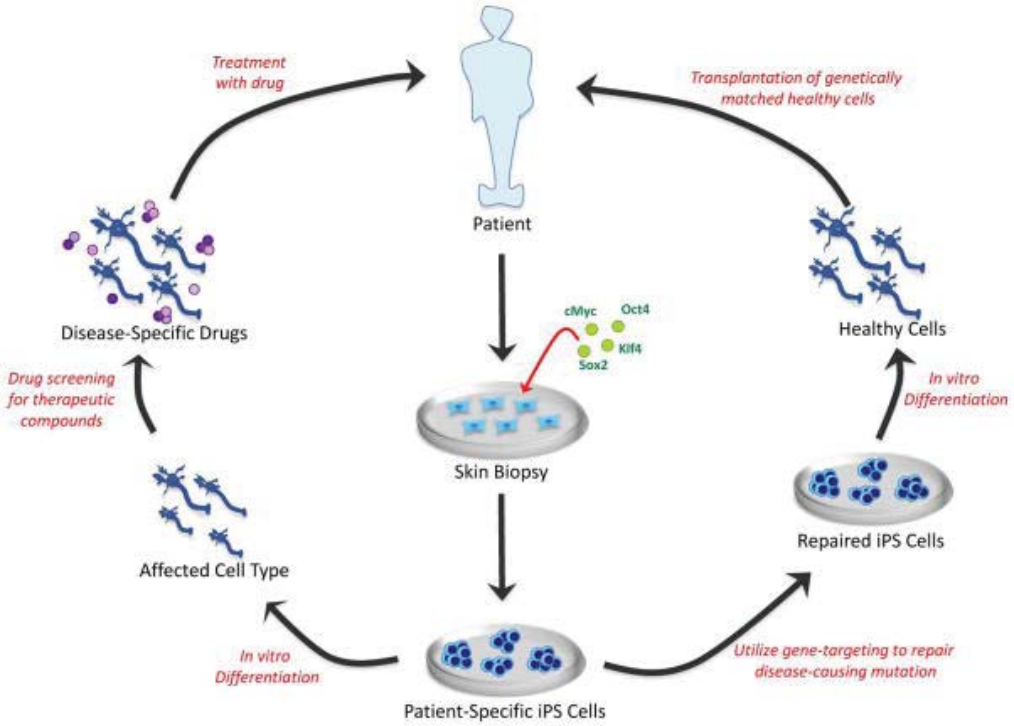


Figure 3. Applications of hiPSCs. hiPSCs have the potential to be used to model and treat human disease, in this case, via drug testing or targeted gene-repair strategies. Patient-specific hiPSCs are derived from co-transfection of pluripotency-inducing transcription factors in cells isolated from a skin biopsy. The hiPSCs are differentiated *in vitro* into the affected cell type where they could be used in a drug screen to test for suitable drugs for treatment of the disease. Alternatively, the disease-causing mutation is corrected and repaired hiPSCs are differentiated *in vitro* into healthy cells of the affected cell type, and the genetically matched cells are subsequently transplanted into the patient. Adapted from (Robinton and Daley, 2012).

1.2.1 Pancreatic specification protocols to date

Pancreatic differentiation was first published by Soria *et al.* in a landmark study reporting the successful differentiation of ESCs into insulin-producing cells (Soria et al., 2000). In this study, mouse ESCs constructed to contain a neomycin resistance gene under the control of the human insulin promoter were able to correct hyperglycaemia a week upon implantation into streptozotocin-induced diabetic mice. A subsequent study described the specification of mouse ESCs into the definitive endoderm (DE) in the presence of activin A and absence of fetal bovine serum, establishing the first stepping stone for directed differentiation to many organ systems (Kubo et al., 2004), namely the digestive and respiratory tracts, thyroid, liver, and pancreas. The high levels of activin A mimics the action of Nodal, a ligand for transforming growth factor-β (TGF-β) superfamily, which activates downstream signalling pathways and transcriptional networks that regulates the formation of the DE (Lowe et al., 2001, Champeris Tsaniras and Jones, 2010). Later, hESCs were also efficiently differentiated into DE using elevated concentration of activin A (D'Amour et al., 2005).

Shortly after, a comprehensive stepwise pancreatic specification protocol was introduced, describing the differentiation of hESCs to endocrine cells capable of synthesising pancreatic hormones such as insulin, glucagon, somatostatin, pancreatic polypeptide and ghrelin with the use of specific growth factors and chemical compounds (D'Amour et al., 2006). With this *in vitro* differentiation protocol, the cells mimic *in vivo* pancreas organogenesis by being directed through stages resembling DE, gut-tube endoderm, pancreatic endoderm and endocrine precursor, thus recapitulating the major stages of normal pancreatic endocrine development. Each stage is recognised by the expression of specific markers. One striking difference between *in vitro* differentiation protocols and *in vivo* pancreas organogenesis is the duration, which are 2-3 weeks and 12-13 weeks respectively. The cells produced from this protocol have an insulin content almost mimicking that of adult islets, and released C-peptide in response to various secretagogues, but only minimally to glucose. The presence of immature polyhormonal cells e.g. insulin and

glucagon, or insulin and somatostatin double-positive cells, however, suggested a lack of precision in the endocrine specification, as mature β-cells solely secrete insulin.

The publication of an improved serum-free protocol by *Jiang et al.* which involved activin A, retinoic acid (RA), basic fibroblast growth factor (bFGF) and nicotinamide to promote pancreatic differentiation resulted in islet-like structures with distinct insulin-, glucagon-, and somatostatin-positive mono-hormonal cells (Jiang et al., 2007b). The protocol, composed of four stages (definitive endoderm induction, pancreatic endoderm formation, endocrine induction, islet-like cluster maturation) generated about 24% *PDX1*-positive cells and 4% C-peptide-positive cells. These cells were insulin-producing and responsive to fluctuations in glucose levels in a suspension cell culture system but showed low levels of response when cultured in adherence.

These studies were succeeded by numerous modified variations of pancreatic differentiation protocols (Jiang et al., 2007a, Kroon et al., 2008, Zhang et al., 2009, Cai et al., 2010, Nostro et al., 2011, Loh et al., 2014, Pagliuca et al., 2014, Rezania et al., 2014, Russ et al., 2015, Cho et al., 2012). With the introduction of such a wide variety of different protocols coupled with the ability to generate hPSC lines from healthy individuals or patients with different genetic backgrounds, it is not unexpected that reports on variations in differentiation efficiencies due to the use of different protocols cell lines have arisen (Osafune et al., 2008, Chin et al., 2009). A recent study that closely compared protocol variations at both the DE (Table 2) and pancreatic progenitor (Table 3) stages, and variation in lineage propensity among hPSC lines reported varying differentiation efficiencies between the tested hiPSC and hESC (H9) lines (Rostovskaya et al., 2015). Interestingly, the different protocols specifying pancreatic progenitors yielded no significant difference between the hPSC lines. Furthermore, it was also reported that certain protocols displayed higher endodermal and pancreatic differentiation efficiencies than others (Rostovskaya et al., 2015); two protocols that were the most recently published fared the best for

pancreatic specification, generating over 90% *PDX1*-positive cells (Rezania et al., 2014, Pagliuca et al., 2014).

Although there have been substantial improvements in pancreatic differentiation protocols over the years, several challenges hindering the complete generation of functional β-cells that fully mimics those *in vivo* still remains. Transplantation of these hPSC-derived pancreatic progenitors into immunocompromised mice often resulted in the formation of teratomas, indicating the presence of pluripotent cells and incomplete differentiation. Efficient and consistently reproducible generation of pure pancreatic lineages derivatives is, therefore, key toward driving research in human β-cell biology, drug testing, disease modelling, development of cell replacement therapy and other applications. Understanding the molecular mechanisms promoting β-cell specification, including the studying the transcriptional regulatory networks of β-cell specification, will greatly contribute to the advancement of this field of research.

protocol	references	stage 1		
DE-1	Loh *et al.*	100 ng ml^{-1} Activin A 100 nM PI103 3 μM Chiron 10 ng ml^{-1} FGF2 3 ng ml^{-1} BMP4 −1 day	100 ng ml^{-1} Activin A 100 nM PI103 20 ng ml^{-1} FGF2 250 nM DM3189 −2 days	
DE-2	Touboul *et al.*	100 nM PI103 100 ng ml^{-1} Activin A 20 ng ml^{-1} FGF2 10 ng ml^{-1} BMP4 −3 days		
DE-3	Rezania *et al.*	100 ng ml^{-1} GDF8 3 μM Chiron −1 day	100 ng ml^{-1} GDF8 0.3 μM Chiron −1 day	100 ng ml^{-1} GDF8 −1 day
DE-4	D'Amour *et al.*	100 ng ml^{-1} Activin A 25 ng ml^{-1} Wnt3a −1 day	0.2% FBS 100 ng ml^{-1} Activin A −2 days	
DE-5	Cheng *et al.*	100 ng ml^{-1} Activin A 40 ng ml^{-1} Wnt3a −1 day	0.5 ng ml^{-1} BMP4 10 ng ml^{-1} bFGF 100 ng ml^{-1} Activin A 10 ng ml^{-1} VEGF −4 days	

Table 2. Comparison of definitive endoderm differentiation protocols for hPSCs. Table adapted from Rostovskaya et al., 2015 showing the different conditions of various protocols that were developed and published by independent groups.

protocol	references	stage 2 primitive gut endoderm	stage 3 PDX1+, presumptive pancreatic endoderm	stage 4[a]	validation of differentiation potential of the resulted cells
P-1	Kroon *et al.*	2% FCS 50 ng ml^{-1} FGF7–3 days	2 µM ATRA 250 nM SANT-1 250 nM DM3189 −3 days	—	formation of polyhormonal cells *in vitro*; maturation *in vivo* to functional beta cells
P-2	Nostro *et al.*	3 ng ml^{-1} Wnt3a 50 ng ml^{-1} FGF10 250 nm DM3189 −3 days	2 µM ATRA 250 nM SANT-1 250 nM DM3189 50 ng ml^{-1} FGF10 −3 days	—	formation of polyhormonal cells *in vitro*
P-3	Loh *et al.*	250 nM DM3189 4 µM IWP2 500 nM PD0325901 2 µM ATRA −1 day	2 µM ATRA 250 nM SANT-1 250 nM DM3189 500 nM PD0325901 −3 days	—	not reported
P-4	Rezania *et al.* /Pagliuca *et al.*	250 µM ascorbic acid 50 ng ml^{-1} FGF7 −2 days	250 µM ascorbic acid 50 ng ml^{-1} FGF7 250 nM SANT-1 1 µM ATRA 100 nM DM3189 200 nM TPB −3 days	—	differentiation to monohormonal insulin+ cells *in vitro*; maturation *in vivo* to functional beta cells
P-5	Rezania *et al.*			250 µM ascorbic acid 2 ng ml^{-1} FGF7 250 nM SANT-1 100 nM ATRA 200 nM DM3189 100 nM TPB −3 days	
P-6	Pagliuca *et al.*			250 µM ascorbic acid 50 ng ml^{-1} FGF7 250 nM SANT-1 100 nM ATRA −5 days	

[a]Stage 4 conditions were applied only after protocol 4, constituting protocols 5 and 6.

Table 3. Comparison of pancreatic progenitor differentiation protocols for hPSCs. Table adapted from Rostovskaya et al., 2015 showing the different conditions of various pancreatic progenitor differentiation protocols that were developed and published by independent groups.

1.2.2 Transcription factors associated with pancreas development

Most biological processes are regulated on a transcriptional level. In mammalian cells, the transcription of genes is regulated by several regulatory proteins known as transcription factors (TFs). TFs recognise specific DNA sequences near the target gene, often in regulatory promoter regions that are located upstream of the transcriptional start site (TSS), and can either activate or repress these promoter regions. During islet cell development, TFs play an integral role in directing cell fates by regulating the transcriptional network controlling pancreatic specification and ultimately mature function (Figure 4). Some of the TFs that play a vital role in promoting β-cell function and identity are described below.

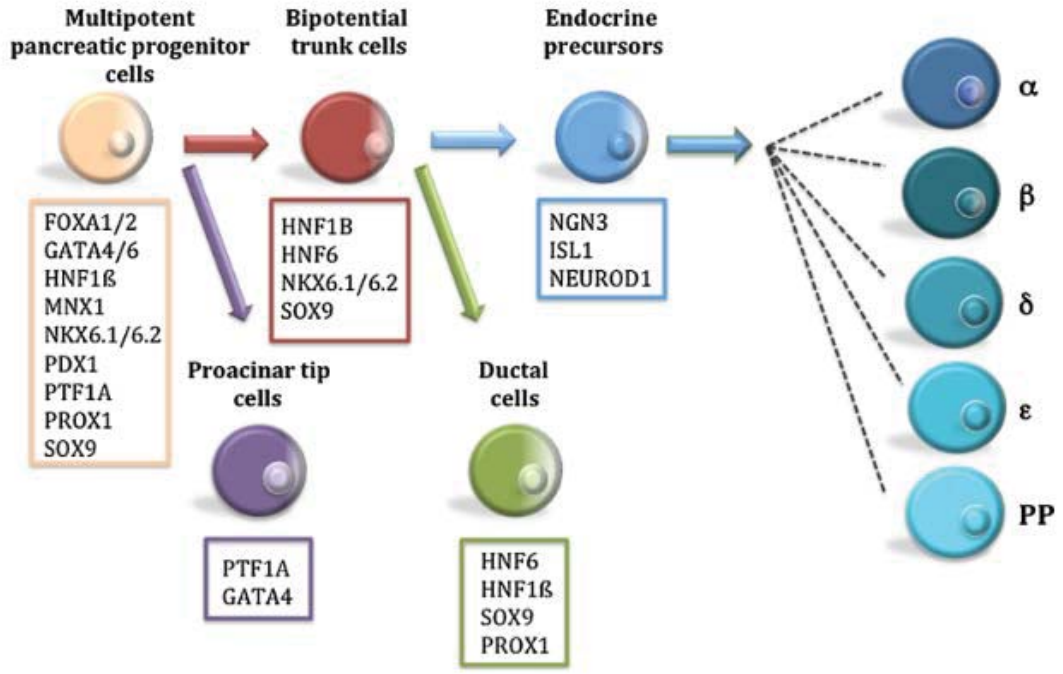


Figure 4. Key transcription factors controlling lineage specification during pancreatic development. Highlighted in this diagram are the key transcription factors known to have an integral role at each stage of pancreatic development. Adapted from (Cano et al., 2014).

SRY (sex determining region Y)-box 17 (*SOX17*)

The Sox family of TFs has a conserved DNA-binding HMG (high mobility group) domain (Bowles et al., 2000), and its early expression is necessary for endoderm formation. In *Xenopus*, *Xsox17* plays important roles in endoderm formation (Hudson et al., 1997, Clements and Woodland, 2000). It has also been shown to be a direct regulator of *FoxA1* and *FoxA2* (Sinner et al., 2004). β-catenin, an intracellular signal transducer in the Wnt signalling pathway, physically interacts with *Sox17* and promotes transcriptional activation of its target genes (Sinner et al., 2004).

Sox17/*SOX17* expression in the mouse and human is similar; it is expressed as an early endoderm marker immediately before 4 weeks post conception in human and e6.0 in mouse, then excluded from pancreatic cells 1 week or 2.5 days later in human and mouse respectively (Kanai-Azuma et al., 2002, Piper et al., 2004, Jennings et al., 2013). *Sox17* is required for the induction of *Pdx1* expression and *Sox17*$^{-/-}$ mice are deficient in the formation of the DE, although they form some foregut but not midgut or hindgut (Kanai-Azuma et al., 2002). Furthermore, constitutive expression of *SOX17* in hESCs produced stable definitive endoderm progenitors, while absence of *SOX17* blocked endoderm differentiation (Seguin et al., 2008).

Forkhead Box A2 (*FOXA2*)

Winged helix/forkhead transcription factors such as *FoxA2* have been reported to be necessary for DE formation (Dufort et al., 1998). In early human development, *FOXA2* is consistently expressed from week 4 (Lyttle et al., 2008, Jeon et al., 2009, Jennings et al., 2013). This is similar in the mouse where *FoxA2* is expressed throughout pancreatic development, and remains expressed in all mature pancreatic cell types of both mice and humans (reviewed in (Pan and Wright, 2011, Cano et al., 2014)). *FoxA1* and *FoxA2* both regulate the expression of key pancreatic gene *Pdx1* by co-occupying multiple regulatory domains in the *PDX1* gene, although this has not been verified in human (Gao et al., 2008, Pan and Wright, 2011). Compound conditional ablation of both *FoxA1* and *FoxA2* in mice resulted in

complete loss of *Pdx1* expression and severe pancreatic hypoplasia (Gao et al., 2008). Forkhead genes of the FOXA class have also been proposed to interact with GATA factors during DE formation (Bossard and Zaret, 1998, Cirillo et al., 2002). *FoxA2* is also strongly expressed in DE derivatives such as the liver (Ang et al., 1993). In contrast to *Sox17*$^{-/-}$ mice, *FoxA2*$^{-/-}$ mice can form the hindgut but not the foregut and midgut (Dufort et al., 1998).

Hepatocyte nuclear factor 1 homeobox beta (*HNF1B*)

HNF1B is highly expressed in humans around 7 weeks post conception and remains expressed throughout pancreatic development (Jeon et al., 2009). In the mouse, *Hnf1B* is expressed in the foregut endoderm prior to the onset of *Pdx1* expression (e8), and later restricted to the epithelial trunk domain and exocrine ducts (Coffinier et al., 1999, Maestro et al., 2003, Haumaitre et al., 2005). *Hnf1B*$^{-/-}$ mice die before gastrulation due to defective visceral endoderm formation, but when the embryonic lethality was rescued by tetraploid aggregation, *Hnf1B*$^{-/-}$ mice displayed absence of the ventral pancreatic bud and an extremely reduced and transient dorsal bud that leads to pancreas agenesis by e13.5 (Haumaitre et al., 2005). In humans, heterozygous mutations in the *HNF1B* gene are associated with MODY5 (Nishigori et al., 1998, Lindner et al., 1999, Bingham et al., 2000, Horikawa et al., 1997). This is in contrast to the mouse where only homozygous mutations produced diabetes in mice.

PDX1

PDX1 is expressed in all pancreatic precursor cells and has been shown to be critically important for early pancreatic development (Bernardo et al., 2009). In humans, *PDX1* is broadly expressed at around 4 weeks post conception, peaking at a later stage when its expression becomes restricted to β-cells (Lyttle et al., 2008, Jennings et al., 2013). In mice, *Pdx1* is first expressed in the primitive gut tube at e8.5, marking the pre-pancreatic endoderm. *Pdx1* is not expressed exclusively in the pancreas and by e10.5, its expression has been reported in parts of the posterior foregut including the stomach, duodenum and bile duct (Guz et al., 1995, Offield et

al., 1996). *Pdx1* high expression then becomes restricted mostly to endocrine cells in the pancreas just before birth (Guz et al., 1995, Offield et al., 1996, Stoffers et al., 1999). Growth of the pancreatic epithelium in *Pdx1*$^{-/-}$ mice is arrested around e10.5 despite the presence of initial budding (Jonsson et al., 1994, Offield et al., 1996, Ahlgren et al., 1996). Microarray analyses performed on e10.5 *Pdx1*$^{-/-}$ mutant mice embryos found downregulation of several TFs including *Nkx6-1* and *Ptf1A*, supporting its critical role in the pancreatic transcriptional network (Svensson et al., 2007). However, the direct regulation of *Pdx1* to *Nkx6-1* and *Ptf1A* has yet to be established, though this has been shown for other TFs such as *Gata4*, *FoxA2* and *Hnf1B* (Rojas et al., 2009, Oliver-Krasinski et al., 2009). In humans, homozygous inactivating mutations of the *PDX1* gene result in pancreatic agenesis, known as MODY4 (Stoffers et al., 1997).

PTF1A

PTF1A expression in the human foetal pancreas is only detectable by quantitative real time polymerase chain reaction (qRT-PCR) around mid-gestation, when acinar cells are formed (Jeon et al., 2009). From studies in mice, the separation process of pro-acinar tip cells and trunk cells is established by an antagonistic relationship between TFs *Nkx6-1* and *Ptf1A* (Schaffer et al., 2010). Thus, *Nkx6-1* and *Ptf1A* have important roles in specifying progenitors toward an endocrine or acinar fate, respectively. From being broadly expressed in the dorsal and ventral pancreatic buds, *PTF1A* is progressively restricted to the pro-acinar tip cells, while *NKX6-1* and other TFs such as *SOX9*, and *HNF1B* are localised to the trunk (Obata et al., 2001, Jeon et al., 2009, Schaffer et al., 2010). In mice, *Ptf1A* is first expressed in the pancreatic epithelium at e9.5. By e13.5, its expression is restricted to acinar precursor cells as the tip and trunk domains become segregated. In contrast to *Pdx1*, *Ptf1A* expression is expressed only in the pancreas during development (Kawaguchi et al., 2002). *Ptf1A*$^{-/-}$ mice died shortly after birth and displayed a complete absence of exocrine pancreatic tissue (Krapp et al., 1998). In humans, mutations in the *PTF1A* enhancer may lead to pancreatic agenesis (Sellick et al., 2004, Weedon et al., 2014).

GATA factors

The GATA family identified in vertebrates is composed of six zinc-finger TFs, each playing important roles in the specification and differentiation of multiple cell types (Arceci et al., 1993, Molkentin, 2000, Patient and McGhee, 2002). All members of the GATA family contains a highly conserved DNA binding domain consisting of two zinc fingers that recognise and bind to the motif WGATAR, in which W indicates A/T and R indicates A/G, in the regulatory sequences of target genes (Ko and Engel, 1993). The six GATA members are known as *Gata1* (Evans and Felsenfeld, 1989, Tsai et al., 1989), *Gata2* (Yamamoto et al., 1990, Lee et al., 1991, Dorfman et al., 1992), *Gata3* (Yamamoto et al., 1990, Ho et al., 1991, Joulin et al., 1991, Ko et al., 1991), *Gata4* (Arceci et al., 1993, Kelley et al., 1993), *Gata5* (Laverriere et al., 1994), and *Gata6* (Laverriere et al., 1994). Based on their expression patterns in restricted tissues, the GATA members have been divided into two subfamilies: *GATA1-3* and *GATA4-6* (Molkentin, 2000). *Gata1-3* are prominently expressed in hematopoietic stem cells where they regulate lineage-specific gene expression in T-lymphocytes, erythroid cells, and megakaryocytes (reviewed in (Orkin, 1998)). *Gata4 -6* are expressed in various mesoderm- and endoderm-derived tissues such as the heart, liver, lung, pancreas and gut where they play critical roles in regulating tissue-specific gene expression (Arceci et al., 1993, Kelley et al., 1993, Laverriere et al., 1994, Morrisey et al., 1996a, Suzuki et al., 1996, Morrisey et al., 1997). Of these GATA family members, only *Gata4* and *Gata6* have been shown to be expressed in the pancreas and have a role in pancreatic development (Decker et al., 2006, Carrasco et al., 2012). In the mouse embryo, *Gata4* and *Gata6* overlap in the foregut endoderm at e9.5, including the pre-pancreatic endoderm (Molkentin, 2000). As embryonic development proceeds, *Gata4* and *Gata6* expression diverges to be expressed in acinar cells at e16.5, and endocrine islets at e14.5 respectively (Ketola et al., 2004).

Gata4 null mice display severe developmental abnormalities, resulting in embryonic lethality between e7.0 and e9.5 (Molkentin et al., 1997). Tetraploid embryo complementation experiments were able to rescue these defects, enabling the generation of clonal embryonic e9.5 *Gata4* $^{-/-}$ embryos directly from embryonic

stem cells (Watt et al., 2004). Similarly, *Gata6* null mice die after implantation because of defects in visceral endoderm function and extraembryonic development (Morrisey et al., 1998). This early embryonic deficiency associated with *Gata6* null mice could also be rescued with tetraploid embryo complementation by providing *Gata6*-null embryos with a wild-type extraembryonic endoderm (Zhao et al., 2005). Thus, although the early embryonic lethality associated with the *Gata4*$^{-/-}$ and *Gata6*$^{-/-}$ mice has precluded loss-of-function analyses in the pancreas, *in vivo* mouse studies using tetraploid complementation and a transgenic Gata-engrailed fusion protein have suggested that Gata4 and/or Gata6 contribute to the regulation of pancreas development (Decker et al., 2006, Watt et al., 2007).

As mentioned earlier, *GATA4* is expressed during early human pancreatic budding between 4 to 5 weeks post conception, then becomes reduced in pancreatic progenitors, remaining mainly in mature acinar cells, an expression pattern similar to that of mice. In humans, the precise expression pattern of *GATA6* during pancreatic development from has not been closely studied. Recently, studies established a critical regulatory role for *GATA4* and *GATA6* in human pancreas formation, and reported that heterozygous mutations in *GATA4* or *GATA6* can lead to pancreatic agenesis (Lango Allen et al., 2012, Shaw-Smith et al., 2014, Bonnefond et al., 2012). Heterozygous mutations in *GATA4* and *GATA6* have also been associated with congenital heart defects (Garg et al., 2003, Lango Allen et al., 2012, Bonnefond et al., 2012). Strikingly, this is not the case in mice. Heterozygous or homozygous inactivation of either *Gata4* or *Gata6* does not impair pancreas formation, but simultaneously inactivation of both three or four *Gata4* and *Gata6* alleles in the pancreatic progenitor domain leads to pancreatic agenesis and loss of *Pdx1* expression, indicating a functional redundancy for these TFs during pancreas development in mice (Carrasco et al., 2012, Xuan et al., 2012).

SOX9

SOX9, a member of the SRY/HMG box family, is found in *PDX1*-positive cells in early human pancreas by about 4 weeks post conception and in the mouse is

expressed in the *Pdx1* domain from e9.5 (Seymour et al., 2007, Lynn et al., 2007, McDonald et al., 2012, Jennings et al., 2013). Although *SOX9* expression is absent in subsequent endocrine cells and restricted to pancreatic duct cells (Jennings et al., 2013), it plays integral roles in maintaining the pancreatic progenitor pool, supporting endocrine cell differentiation, and co-localising with and regulating the expression of other important TFs such as *FOXA2* and *NGN3* (Seymour et al., 2007, McDonald et al., 2012). In mice, conditional inactivation of *Sox9* in the *Pdx1* domain results in severe pancreatic hypoplasia (Seymour et al., 2007). In addition, the $Sox9^{+/-}$ mice display a similar phenotype to *SOX9* haploinsufficiency in humans, where failed maintenance of endocrine progenitors result in islet hypoplasia (Sosa-Pineda et al., 1997, Piper et al., 2002, Seymour et al., 2008).

NKX6-1

As mentioned earlier, *NKX6-1* expression in humans is detected after 4 weeks post conception, once *SOX17* is excluded from the pancreatic buds. Its expression then becomes restricted to β-cells by 14-16 weeks (Brissova et al., 2005, Jennings et al., 2013). Similarly in rodent, *Nkx6.1* expression is broadly expressed in the early stages of pancreatic development, then gradually becomes restricted to β-cells (Sosa-Pineda et al., 1997). $NKX6\text{-}1^{-/-}$ mice exhibit a severe reduction in β-cells, and failure of conditional *Nkx6.1* mutants to express *Pdx1* reveal its role in specifying endocrine precursors toward β-cell lineage (Sander et al., 2000, Henseleit et al., 2005, Schaffer et al., 2013). In human T2D islets, there is a reduced expression of *NKX6-1* (Guo et al., 2013).

NGN3

The expression pattern of *NGN3* in human has been described in an earlier section (1.1.1 Development of the human pancreas). $NGN3^{-/-}$ mice fail to generate pancreatic endocrine cells and die postnatally from severe hyperglycemia (Gradwohl et al., 2000). In humans, it has been reported that a rare biallelic *NGN3* null mutation resulted in PNDM with no histologically detectable islets, but detectable C-peptide levels suggest the presence of some β-cells (Rubio-Cabezas et al., 2011).

1.2.3 Transcription factor *GATA6*

The human (Suzuki et al., 1996) and mouse (Narita et al., 1996, Morrisey et al., 1996a) *GATA6*/*Gata6* gene was first described in 1996. In humans, the *GATA6* gene is located on human chromosome 18 q11.1–q11.2 (Suzuki et al., 1996). Initially, the *GATA6* cDNA was reported to have an open reading frame (ORF) from nucleotide residues 348 to 1697 extending from an initiator methionine codon at 716 bp, encoding a predicted protein size of 45.3 kDa composed of 449 amino acids (MYQ-ORF, Figure 5) (Suzuki et al., 1996, Huggon et al., 1997). It was subsequently discovered that translation of the *GATA6* gene can initiate from two alternative initiator methionine codons, giving rise to two protein isoforms (Brewer et al., 1999). In this study, a longer potential ORF encoding a protein of 595 amino acids, which commences at a more upstream, "in-frame" putative initiator methionine codon at 278 bp was revealed (MALT-ORF, Figure 5) (Brewer et al., 1999). Both methionine codons are within a theoretically favourable context for translation initiation (Kozak, 1981, Cavener and Ray, 1991, Kozak, 1997) and are located within exon 2, out of the 7 exons of the *GATA6* gene (Figure 5). Both isoforms possess an N-terminal transactivation domain and two zinc finger domains, both of which are essential for activity (Takeda et al., 2004). It has been reported that the two isoforms differ in their transactivation potential; full length *GATA6* which expresses both isoforms and an altered *GATA6* which only produces the longer isoform had the highest transactivation potentials (Brewer et al., 1999). However, deletion of the extended N-terminal 146 amino acids reduced transactivation potential by approximately 50%, and deletion of the region proximal to the zinc finger domains resulted in very little transactivation activity (Brewer et al., 1999).

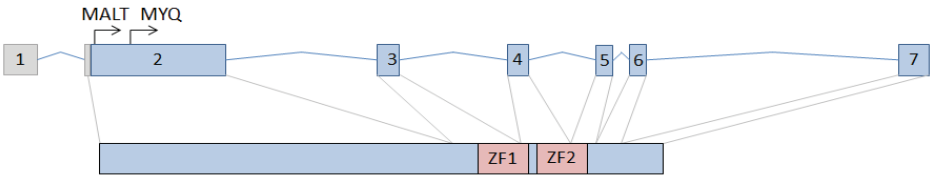


Figure 5. Corresponding cDNA transcript and protein product of *GATA6*. Top row represents cDNA, noncoding (grey) and coding (blue regions) of exons 1-7. Two initiator methionine codons (MALT and MYQ) are indicated in exon 2. Bottom row represents the full length isoform of the *GATA6* protein, including locations of the two zinc finger DNA binding domains (red boxes, ZF1 and ZF2).

The distribution of *GATA6* transcripts in embryonic tissue appeared to be high in the heart and lungs, and absent in brain, liver, or kidney (Suzuki et al., 1996). In adult tissues, *GATA6* transcripts were expressed at high levels in the heart, ovary, lung, and pancreas, low levels in the liver and spleen, and absent in the brain, placenta, skeletal muscle, thymus, prostate, testes, small intestine, colon, or leukocytes. Distribution of *GATA4* differed slightly as was not detected in either adult or embryonic lung or in adult spleen, but present in testes. Of note, *GATA6* and *GATA4* expression overlapped in the adult pancreas and heart (Suzuki et al., 1996).

In mice, the *Gata6* gene maps to a region of chromosome 18 that shows homology to human chromosome 18 (Narita et al., 1996). The *Gata6* cDNA was initially reported to include a 1332 bp ORF encoding a 444 amino acid polypeptide with a predicted protein size of 45 kDa (Morrisey et al., 1996a). Similarly to human, the mouse *Gata6* gene encodes a longer polypeptide in addition to the one described earlier, with both methionine initiator codons located within exon 2, resulting in two protein isoforms (Brewer et al., 1999). The extended N-terminal sequence comprises 147 amino acid residues. *Gata6* distribution overlaps with *Gata4* in the adult mouse, and is abundant in the heart, lung, stomach, small intestine, large intestine and ovary, has lower levels in the liver and is absent in the brain, kidney, or skeletal muscle (Narita et al., 1996). Importantly, during mouse development, *Gata4* and *Gata6*, but not *Gata5*, are expressed in overlapping domains within the primitive and foregut endoderm, including the regions that give rise to liver and pancreas (Bossard and Zaret, 1998, Decker et al., 2006).

The recent genome sequencing of 27 neonatal diabetic patients with pancreatic agenesis or severe pancreas hypoplasia that revealed 56% of the patients had spontaneous heterozygous mutations in the *GATA6* gene sparked a potential interest in the *GATA6* gene (Lango Allen et al., 2012). This study associated *GATA6*, on top of previously identified genes such as *PDX1* and *PTF1A* whose inactivation causes pancreatic agenesis in humans, to a potential role in pancreas morphogenesis.

1.3. Disease modelling of pancreatic agenesis

hPSC technology is a ground-breaking step toward modelling human disease in a controlled laboratory setting. hiPSCs can be derived from healthy volunteers or patients, thus hiPSC technology allows cellular models of disease to be formed from differentiated human pluripotent stem cells. Although animal models have proven invaluable in uncovering fundamental biology, inherent differences between human and rodent biology lead to limitations in the ability of animal systems to recapitulate human disease. Several such examples are human islets comprise a lower proportion of β-cells, and a higher proportion of α- and δ- cells compared to mouse islets (Brissova et al., 2005), and a different islet architecture between the two species, with human β-cells being dispersed among α- and δ- cells, while mouse β-cell maintaining a core surrounded by the four other endocrine cell types (Figure 6) (Cabrera et al., 2006). In rodents, insulin is encoded by two genes (*INS1* and *INS2*), whereas in humans, insulin is only encoded by one gene (*INS*) (Melloul et al., 2002). Indeed, discordant phenotypes in the mouse compared to human caused by *GATA6* mutations have been noted as discussed in an earlier section (1.2.2). As such, this precludes the use of animal models in modelling the disease.

The advancement of hPSC technology coupled with the availability of genome editing tools (discussed in the next section) provide a valuable opportunity to accurately model pancreatic agenesis as a disease and investigate the role of TF *GATA6* in pancreatic development and elucidate how *GATA6* mutations can impair the formation of the human pancreas.

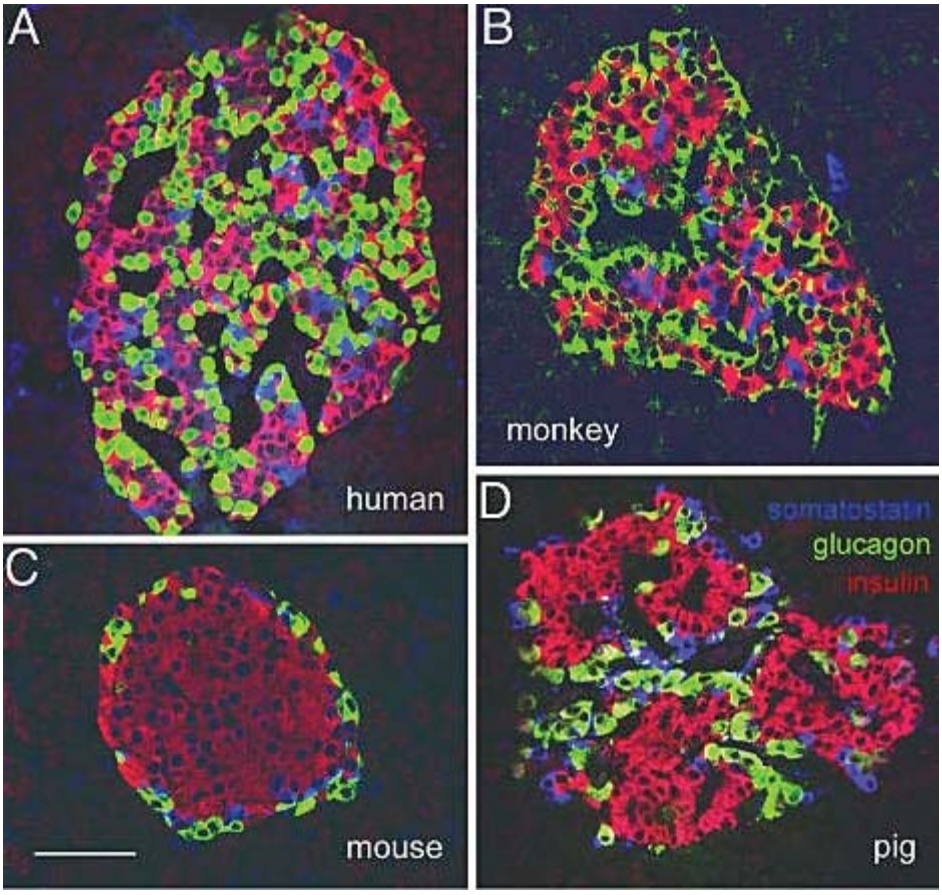


Figure 6. Confocal microscopy images of α-, β- and δ- cells within Islets of Langerhans show striking interspecies differences. (A, B) In human and monkey islets, insulin-immunoreactive β-cells (red), glucagon-immunoreactive α-cells (green), and somatostatin-immunoreactive δ-cells (blue) cells were all found randomly distributed. **(C)** In mouse islets, insulin-containing cells were located in the core, and glucagon- and somatostatin-containing cells in the circumference. **(D)** In pig islets, α-, β- and δ- cell distributions are similar to that of the mouse but appear to be formed in smaller units. Scale bar, 50 µm. Adapted from (Cabrera et al., 2006).

1.3.1 Genome editing tools

The concept of genetic engineering was first introduced in 1972 by Paul Berg's lab in the form of recombinant DNA, where scientists successfully combined genes from different species (Jackson et al., 1972). Over the years, there has been tremendous progress in the development of methods not only to manipulate DNA, but also to generate vector systems and optimise their delivery into cells. The success of the Human Genome Project opened up many doors toward a deeper understanding of how the nucleotide sequence of human nuclear DNA relates to pathology of hereditary as well as multifactorial diseases. It also enabled the study of functional elements within the human genome, such as transcription factors. In order to establish relationships between gene function and disease, two strategies are often used: repression by switching off a gene i.e. knockdown or knockout and activation by overexpressing a gene.

In 1996, a study reporting precise genome editing using a zinc finger protein domain coupled with the FokI endonuclease domain was published (Kim et al., 1996). For the first time, it was possible to perform site-specific nuclease cutting of DNA at strictly defined sites *in vitro*. Zinc finger nuclease (ZFN) thus became the basis for editing cultured cells including pluripotent stem cells, plants and animals (Bibikova et al., 2002, Townsend et al., 2009, Zhang et al., 2010, Lombardo et al., 2011, Provasi et al., 2012, Torikai et al., 2012). However, the technology brought several disadvantages such as the high complexity and cost of assembling the DNA-binding protein domains, low efficiency and potential off-target effects. This drove the discovery of two more genome editing tool which succeeded the ZFN: transcription activator-like effector nuclease (TALEN) from *Xanthomonas* bacteria (Boch et al., 2009, Moscou and Bogdanove, 2009, Christian et al., 2010, Miller et al., 2011) and RNA-guided DNA endonuclease Cas9 from the type II bacterial adaptive immune system clustered regulatory interspaced short palindromic repeats (CRISPR/Cas) (Cong et al., 2013, Mali et al., 2013). As CRISPR technology was only just emerging when the project started, TALEN was the genome editing tool that was used in this project, and will be described in detail below.

1.3.2 Transcription activator-like effector nuclease (TALEN)

The TALEN system was developed based on the study of the bacteria *Xanthomonas* genus, which act as pathogens to crop plants by secreting naturally occurring effector proteins (transcription activator-like effectors, TALEs) that support bacterial virulence, proliferation, and dissemination (Boch and Bonas, 2010). The TALE proteins bind to genomic DNA to alter transcription in host cells, thereby facilitating pathogenic bacterial colonisation.

TALEN proteins contain a DNA-binding TALE repeat domains composed of a series of 33 to 35 amino acid repeat domains each recognising a single nucleotide base (Figure 7d), flanked by an amino (N)-terminal domain and a carboxy (C)-terminal domain that is fused to a *Fok*I restriction endonuclease domain (Figure 7a). In order for the TALEN to recognise a specific sequence on the double-stranded DNA, two TALEN proteins are required, commonly called the left and right TALEN arms, and they each recognise a particular sequence on the forward or reverse strand of the DNA (Figure 7b).

Each TALE repeat domain has an almost identical amino acid sequence, except for two hypervariable residues typically found at positions 12 and 13 of the repeat domain, which determines which nucleotide base the individual TALE repeat domain will recognise (Figure 7c) (Boch et al., 2009). These hypervariable residues are also known as the repeat variable di-residues (RVDs). The RVDs NN, NI, HD and NG code for the recognition of nucleotide base guanine, adenine, cytosine and thymine, respectively (Figure 7d). The last repeat that binds a nucleotide at the 3'-end of the recognition site consists only of 20 amino acid residues, and is therefore called a half-repeat. Subsequent improvements were made to increase binding and specificity of these residues to their respective nucleotide. For example, the RVDs NK was reported to be more specific for guanine than NN, although it also recognises adenine (Moscou and Bogdanove, 2009, Miller et al., 2011) and displayed less activity as compared to NN (Streubel et al., 2012). The RVDs NH were suggested to be more specific than the NN repeat but with lower activity (Streubel et al., 2012,

Cong et al., 2012). It was also reported that RVDs HD and NH bind their preferred nucleotides strongly, while NI and NG bind to their preferred nucleotides relatively weaker (Streubel et al., 2012). Furthermore, a second generation scaffold greatly increased *in vivo* modification efficacy (Bedell et al., 2012).

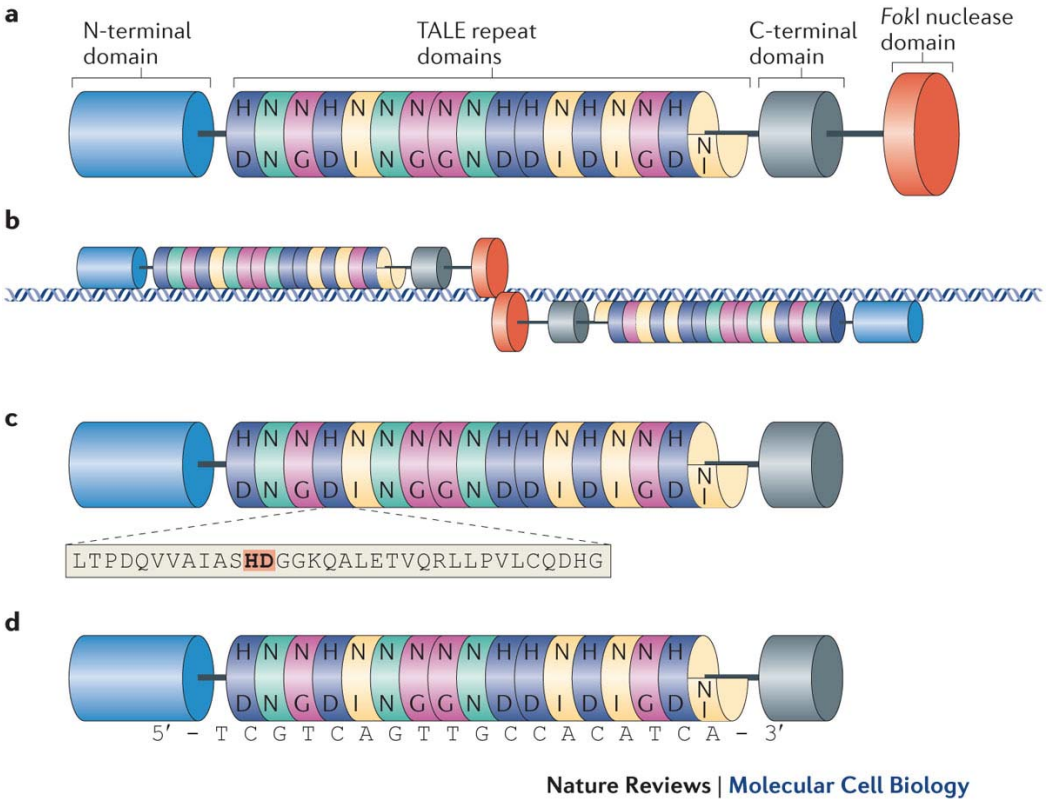


Figure 7. Overview of TALEN proteins. (a) Schematic of one arm of a fully constructed TALEN protein comprising of an amino (N)-terminal and carboxy (C)-terminal domain that are required for DNA-binding, a non-specific *Fok*I endonuclease domain and TALE repeat domains forming an array ending with a truncated half repeat. **(b)** A pair of TALEN arms, namely the left and right TALEN arm, binds and cleaves as dimers on a specific target site of the double-stranded DNA, resulting in a double-stranded break of the DNA. Cleavage by the *Fok*I nuclease domains occurs within a spacer region that lies between binding sites of the left and right TALEN arms. **(c)** The amino acid sequence of a single TALE repeat domain is highly conserved and is similar in all the various domains except for the 12th and 13th amino acid known as the hypervariable residues or repeat variable di-residues (RVDs) (highlighted in orange and bold text). **(d)** Each RVD confers specificity to a single nucleotide and is arranged in the order of its target sequence during construction of the TALEN protein. The TALE array is responsible for binding to DNA at a specific site. Preceding the first base bound by a TALE repeat at the 5’ end is a thymine. Adapted from (Joung and Sander, 2013).

TALEN construction can be challenging due to the nature of the nearly identical repeat sequences of the TALE repeat domains, and the assembly of numerous domains. Many groups have devised platforms for TALEN assembly to facilitate a simple and efficient construction process and these can be broadly grouped into three categories: standard restriction enzyme and ligation-based cloning (Sander et al., 2011, Huang et al., 2011), 'Golden Gate' cloning (Engler et al., 2008, Engler et al., 2009), and solid-phase assembly (Reyon et al., 2012). In this project, the standard restriction enzyme and ligation-based cloning platform was used. This method utilises an archive of plasmids encoding single or multiple TALE repeat domains and join them in a parallel hierarchical fashion via restriction digestion and ligation reactions. The 'Golden Gate' cloning platform is a multi-fragment ligation strategy and allows for 3 to 10 TALE repeat domains to be simultaneously ligated in a particular linear order into a plasmid vector (Weber et al., 2011). Solid-phase assemblies such as Fast Ligation-based Automatable Solid-phase High-throughput (FLASH) assembly, Iterative Capped Assembly (ICA) are automated, high-throughput methods for assembling numerous TALE repeat arrays (Reyon et al., 2012, Briggs et al., 2012, Wang et al., 2012).

1.3.3 Nuclease-mediated mutations

Simultaneous introduction of the left and right TALEN arms into cells often lead to site-specific DNA double-stranded breaks (DSB). The *Fok*I endonuclease domain is crucial for the successful cleavage of the double-stranded DNA by recognising a non-palindromic DNA sequence and making a double-stranded cut outside of that sequence, commonly within a region known as a spacer, resulting in a 5' overhang (Hiroyuki and Susumu, 1981). The spacer must be an appropriate length of around 16 nucleotides to permit dimer formation. In order to cleave the DNA, each of the *Fok*I domain within the left or right TALEN arm must dimerise during adjacent and independent binding events of each arm onto the site-specific DNA in the correct orientation (Vanamee et al., 2001). The need for two DNA binding events to occur and for the *Fok*I domains to form a heterodimer pair prior to DNA cleavage improves specificity and reduces off-target effects via the elimination of unwanted homodimers (Miller et al., 2007, Szczepek et al., 2007).

Numerous studies have established that normal cellular repair of DSB occur through two pathways; non-homologous end-joining (NHEJ) or homologous recombination (HR) (Rudin et al., 1989, Plessis et al., 1992, Rouet et al., 1994, Choulika et al., 1995, Lieber, 2008, Jackson and Bartek, 2009). It was subsequently realised that these two highly conserved cell repair pathways can be exploited to introduce targeted mutations in a wide variety of cell types and species after nuclease-induced DSBs have taken place (Figure 8) (Bibikova et al., 2001, Bibikova et al., 2002, Bibikova et al., 2003). The NHEJ repair pathway is error-prone and often lead to mutations containing insertions and/or deletions (indel) of variable length originating from the site of the DSB, thus resulting in frameshift mutations that can lead to the knockout of gene function (Bibikova et al., 2002). Alternatively, if a double-stranded DNA 'donor template' is supplied in combination with the pair of TALEN arms, HR of a nuclease-induced DSB can be used to introduce precise nucleotide substitutions or insertions by repairing the DSB with the information encoded on this template (Moehle et al., 2007).

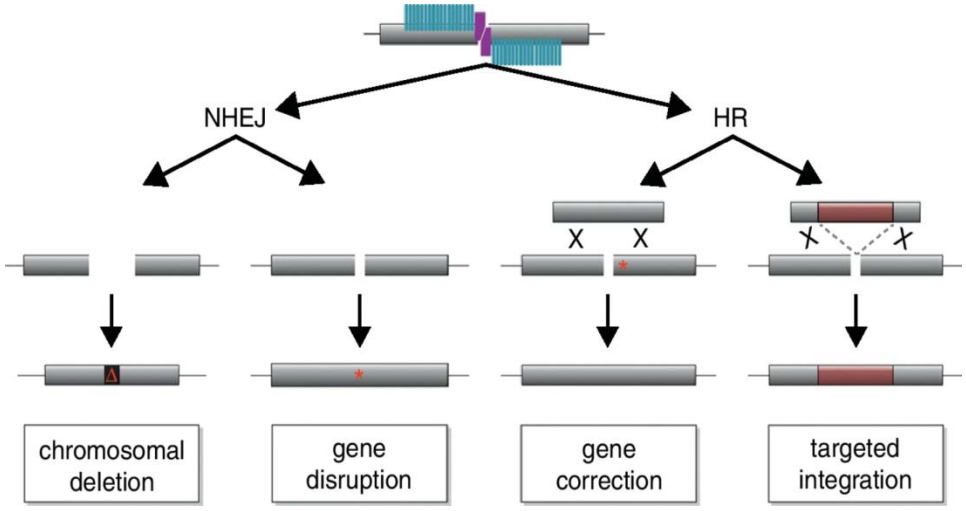


Figure 8. TALEN-mediated genome editing. After a TALEN-induced double-stranded break in the DNA occurs, the cell may undergo one of two highly conserved repair pathways, NHEJ or HR, to repair the cleaved DNA. In the absence of a donor template, the cell undergoes the error-prone NHEJ pathway by ligating the DNA. Nucleotide insertions and/or deletions (asterisk) will commonly be introduced, disrupting the open reading frame and possibly resulting in a premature stop codon. In the presence of a donor template, the cell undergoes the HR pathway which can be used to either correct a mutation (asterisk) in the genome or to target integration of a transgene into a specific site. Adapted from (Mussolino and Cathomen, 2012).

Genome editing tools such as the TALEN technology provide tremendous potential for experimental, biotechnological and therapeutic purposes. Such applications include gene disruptions in model organisms, cell-based disease modelling such as in hPSCs, and gene corrections (allele editing). In cell-based disease modelling, the impact of a specific gene disruption and of specific sequence variants on gene function can be studied closely and be directly associated with diseases. This enables the generation of isogenic cell lines so any possible effects on the disease phenotype under investigation that may be caused by genetic background variations can be excluded (Figure 9).

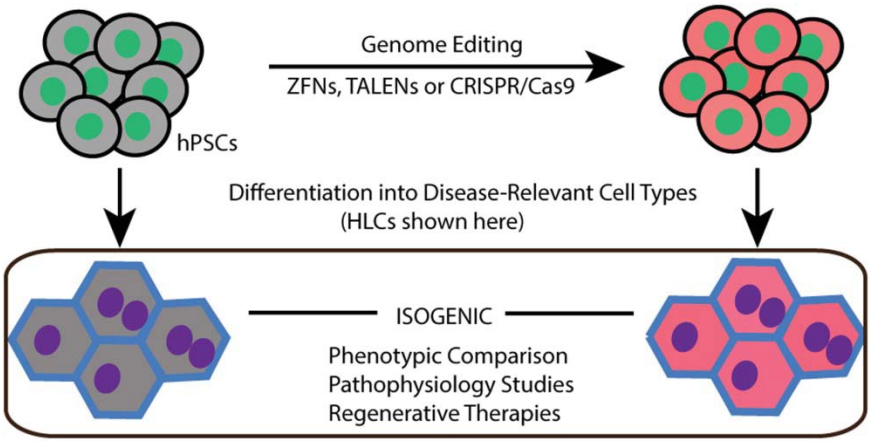


Figure 9. hPSC-based disease modelling. Wild-type hPSCs are targeted by genome editing tools such as TALEN to generate isogenic cell lines. Wild-type and mutated hPSCs are then differentiated into disease-relevant cell types followed by phenotypic comparison and pathophysiological studies to determine the direct association of the gene to the disease phenotype. Adapted from (Yu and Cowan, 2016).

1.4. Objectives of the project

Advancement in genome sequencing technologies in the recent years has provided a major unexpected discovery in this field. The recent study published by Allen et al. where 15/27 (56%) patients with pancreatic agenesis and exocrine pancreatic insufficiency requiring enzyme replacement therapy, born to non-diabetic parents harboured *de novo* heterozygous inactivating mutations in *GATA6* established a key role for the transcription factor *GATA6* in human pancreatic development (Lango Allen et al., 2012). This study is the basis for my project. Human genetics, therefore, has established that *GATA6* is an essential regulator of human pancreas development, but it does not shed light on the underlying molecular mechanism, nor does it define the precise cell types or developmental stages in which the essential role takes place. The role of *GATA6* in the development of the pancreas has been well studied in the mouse, but this is not the case in humans. From mouse studies, it is known that *GATA6* is expressed in the developing pancreas and is an important regulator of pancreas development. Thus, the overall objective of my project is to elucidate the role *GATA6* in the development of the human pancreas. Knowledge gained from this project could potentially contribute to therapies for neonatal diabetes.

The first objective of this study is to perform directed differentiation of hPSCs into the pancreatic lineage using a fully defined culture system. The second objective is to obtain *GATA6* patient lines, reprogram them to obtain patient-derived *GATA6* mutant lines, and perform directed differentiation into the pancreatic lineage to assay the effect of *GATA6* mutations on the development of the pancreas. The third objective is to perform disease modelling of pancreatic agenesis by generating *GATA6* heterozygous and homozygous mutations in hPSC lines via TALEN as a genome editing tool. The fourth objective is to perform phenotypic comparisons between these TALEN-generated *GATA6* mutant hPSCs and their respective isogenic control hPSC lines. The final objective is to define the molecular mechanisms of *GATA6* by investigating the transcriptional networks controlled by *GATA6* through RNA-sequencing and identifying its interacting partners through ChIP-sequencing.

CHAPTER 2 MATERIALS AND METHODS

2.1. Tissue culture

2.1.1. Human pluripotent stem cell lines

H9 hESCs were purchased from WiCell (WA09). FSPS13.B hiPSCs were derived by a colleague from peripheral blood mononuclear cells (PBMCs) using Sendai virus as described previously (Yusa et al., 2011) by the Human induced pluripotent stem cell initiative (HIPSCI). The resulting cells are available for distribution through European Collection of Authenticated Cell Cultures (ECACC). Further information is available at http://www.hipsci.org/lines/#/lines/HPSI0813i-fpdm_2.

The generation of *GATA6* patient-derived hiPSCs was approved by the Great Ormond Street Hospital and Institute of Child Health Research Ethics Committee (ethics reference number: 08/H0713/82), and informed consent was obtained from all patients. Skin fibroblasts from *GATA6* patients were obtained with the help of Dr Ranna El Khairi. Skin punch biopsy samples were collected from volunteering patients and all hiPSC lines used were derived and validated by the Cambridge Biomedical Research Centre hiPSC Core Facility.

Reprogramming of the *GATA6* patient fibroblasts to derive *GATA6*-patient iPSCs was done with the help of the hiPSC core facility at the Anne McLaren Laboratory for Regenerative Medicine using sendai virus reprogramming as described previously (Yusa et al., 2011).

2.1.2. Growth conditions

hPSCs were routinely cultured in feeder-free conditions on vitronectin-coated (Stemcell Technologies #07180) sterile plates (Corning, Costar) with Essential 8 (E8) media (Life Technologies #A1517001). Subculture was performed every 4-6 days by aspirating the cell culture media, washing the cells once with D-PBS, incubating them in 0.5 mM EDTA (Invitrogen, #15575020) for 5 min at room temperature, aspirating the EDTA and adding appropriate volumes of fresh E8 media (Table 4) to the cells. hPSCs were further physically detached from the plates by gentle pipetting, collected in a 15 ml Falcon tube (BD Falcon) and allowed to settle for 10 min. The supernatant was aspirated and hPSCs were re-suspended in fresh E8 media, and then dissociated into smaller clumps by gentle pipetting. The cells were split and plated onto fresh vitronectin-coated plates (Table 5) containing appropriate volumes of media (Table 4). Cells were maintained by refreshing E8 media on a daily basis.

Prior to differentiation into any lineages, culture plates were pre-coated with 0.1% Gelatin (Millipore, #ES-006-B) for at least 20 min at room temperature. The gelatin solution was then aspirated and replaced with MEF media (Table 6 and Table 7). The MEF-coated plates were allowed to incubate at 37°C overnight, up to 1 week, and used when required.

Table 4. Volume of E8 used in splitting and maintenance of cells

Culture vessel	During splitting prior to physical detachment	Maintenance
10 cm dish	3 ml	7 ml
1 well of 6wp	1 ml per well	1.5 ml
1 well of 12wp	0.5 ml per well	1 ml
1 well of 24wp	0.3 ml per well	0.5 ml

Table 5. Splitting ratio of cells

Initial growth area	Final growth area
1 x 10 cm dish	4 x 10 cm dishes
1 well of 6wp	1 full 6wp, 12wp or 24wp

Table 6. MEF media formulation

Component	Volume
Advanced DMEM/F-12 (Gibco, #12634028)	450 ml
FBS (Biosera, #S04253S181S)	50 ml
L-Glutamine (Gibco, #25030024)	5 ml
+/- Penicillin-Streptomycin (Gibco, #15140122)	5 ml
β-Mercaptoethanol (Sigma-Aldrich, #M6250)	3.5 µl

Table 7. Volume of MEF media used for coating

Culture vessel	Volume
10 cm dish	8 ml
1 well of 6wp	2 ml
1 well of 12wp	1.5 ml
1 well of 24wp	1 ml

2.1.3. Definitive endoderm (DE) differentiation

To plate hPSCs for differentiation, hPSCs were grown to about 80% confluency and then passaged as described in the previous section. The cells were broken into clumps smaller than routine passaging by doubling the number of times they pass through the pipette tip. They were then filtered through a 70 µm cell strainer (BD Falcon, #352350) to obtain a uniformly-sized cell suspension, re-suspended in fresh E8 media containing ROCK inhibitor Y-27632 (Sigma-Aldrich, #Y0503) and plated onto the appropriate MEF-coated culture plates. Prior to plating, the MEF media was aspirated and plates were washed once with D-PBS (Life Technologies, #14190094). Cells were plated at a density such that they were about 60 - 70% confluent at the start of DE differentiation. 24 hours after plating, the culture media was replaced with fresh E8 media without ROCK inhibitor Y-27632. DE differentiation commenced 48 hours after plating.

Using the lab's established DE differentiation protocol, the E8 media was aspirated and cells were washed once with D-PBS to remove any residual E8 media. Chemically defined medium-poly vinyl alcohol (CDM-PVA; Table 8) media of appropriate volume (Table 9) supplemented with 100 ng/ml Activin A (produced in-house), 20 ng/ml basic FGF (bFGF/FGF2; produced in-house), 10 ng/ml BMP4 (R&D Systems, #314-BP-010), 10 µM LY294002 (Promega, #V1201) and 3 µM CHIR99201 (Stratech Scientific, #CT99021) was added to the cells on Day 1 (Table 10). On Day 2, media from Day 1 was aspirated and CDM-PVA media containing 100 ng/ml Activin A, 20 ng/ml FGF2, 10 ng/ml BMP4, and 10 µM LY294002 was added to the cells. On Day 3, media from Day 2 was aspirated and RPMI-B27 media (Table 11) supplemented with 100 ng/ml Activin A and 20 ng/ml FGF2 was added to the cells. Cells were harvested on Day 4 for downstream analyses of the DE.

For DE differentiation using a commercial kit PSC Definitive Endoderm Induction Kit (Gibco, #A27654SA), cells were initially passaged and plated similarly to the lab's established protocol. Volumes of media used are indicated in Table 9. On Day 1, Medium A of the kit was added and on Day 2, Medium A was replaced with Medium B. Cells were harvested on Day 3 for downstream analyses of the DE.

Table 8. CDM-PVA formulation

Component	Volume
Ham's F-12+ GlutaMAX-1 (Gibco, #31765068)	250 ml
Iscove's Modified Dulbecco's Medium (Gibco, #21980065)	250 ml
Chemically Defined Lipid Concentrate (Gibco, #11905031)	5 ml
1-Thioglycerol (Sigma-Aldrich, #M6145)	20 µl
Transferrin (30 mg/ml; Roche #652202)	250 µl
Insulin (10 mg/ml; Roche #1376497)	350 µl
Poly vinly alcohol (Sigma-Aldrich, #P8136)	0.5 g
+/- Penicillin-Streptomycin (Gibco, #15140122)	5 ml

Table 9. Volume of media used for DE differentiation

Culture vessel	Volume
10 cm dish	7 ml
1 well of 12wp	0.5 ml
1 well of 24wp	0.3 ml

Table 10. Lab DE differentiation protocol

Time point	Media	Supplements
Day -1 (plate)	E8	ROCK inhibitor
Day 0	E8	None
Day 1	CDM-PVA	100 ng/ml Activin A + 20 ng/ml FGF2 + 10 ng/ml BMP4 + 10 µM LY294002 + 3 µM CHIR99201
Day 2	CDM-PVA	100 ng/ml Activin A + 20 ng/ml FGF2 + 10 ng/ml BMP4 + 10 µM LY294002
Day 3	RPMI-B27	100 ng/ml Activin A + 20 ng/ml FGF2

Table 11. RPMI-B27 formulation

Component	Volume
RPMI 1640 + GlutaMAX (Gibco, #61870-10)	500 ml
B-27 supplement (Gibco, #17504-044)	10 ml
MEM Non-Essential Amino Acids Solution (Gibco, #11140-050)	5 ml
+/- Penicillin-Streptomycin (Gibco, #15140122)	5 ml

2.1.4. Pancreatic differentiation

Following on from DE differentiation, cells were further specified toward the pancreatic lineage by using a protocol that was previously established in the lab. On Day 4, Day 3 media was aspirated and replaced with Advanced DMEM/F-12 containing bovine serum albumin (ADV-BSA; Table 12) supplemented with 2 µM retinoic acid (RA; Sigma-Aldrich, #R2625), 50 ng/ml FGF10 (Autogen Bioclear, #ABC144), 150 ng/ml Noggin (R&D, #33-44NG/CF) and 10 µM SB-431542 (R&D, #1614/10) (Table 13) using appropriate volumes indicated in Table 14. On Days 5 and 6, media was refreshed with the same supplements as Day 4. Cells were harvested on Day 7 for downstream analyses of the primitive gut tube.

On Day 7, Day 6 media was aspirated and replaced with ADV-BSA supplemented with 2 µM RA, 50 ng/ml FGF10, 150 ng/ml Noggin and 1 mg/ml KAAD-cyclopamine (Toronto Research Chemicals Incorporated, #K171000). On Days 8 and 9, media was refreshed with the same supplements as Day 7. Cells were harvested on Day 10 for downstream analyses of the posterior foregut.

On Day 10, Day 9 media was aspirated and replaced with ADV-BSA supplemented with 2 µM RA, 50 ng/ml FGF10 and 0.25 µg/ml KAAD-cyclopamine (Toronto Research Chemicals Incorporated, #K171000). On Days 11 and 12, media was refreshed with the same supplements as Day 10. Cells were harvested on Day 13 for downstream analyses of the pancreatic progenitors.

On Day 13, Day 12 media was aspirated and replaced with ADV-BSA supplemented with 2 µM RA, 1% vol/vol B27 supplement, 1 µM DAPT (Sigma-Aldrich, #D5942) and 0.1 mM 6-Bnz-cAMP sodium salt (BNZ; Sigma-Aldrich, #B4560). Media was left unchanged on Days 14 and 15. Cells were harvested on Day 16 for downstream analyses of the endocrine progenitors.

On Days 16, 18 and 22, cells were then allowed to mature into immature β-cells by aspirating media from Days 15, 17 and 21 respectively and adding ADV-BSA supplemented with 2 µM RA, 1% B27 supplement and 0.25 µg/ml KAAD-cyclopamine. Cells were harvested on Days 19 and 25 for downstream analyses of immature endocrine cells and β-cells respectively.

Table 12. ADV-BSA formulation

Component	Volume
Advanced DMEM/F-12 (Gibco, #21634010)	500 ml
Bovine Serum Albumin (Europa Bioproducts, Lot #1260)	2.5 g
L-Glutamine (Gibco, #25030024)	5 ml
Penicillin-Streptomycin (Gibco, #15140122)	5 ml

Table 13. Pancreatic differentiation protocol

Time point	Media	Supplements
Day 4	ADV-BSA	2 µM RA + 50 ng/ml FGF10 + 150 ng/ml Noggin + 10 µM SB-431542
Day 5	ADV-BSA	2 µM RA + 50 ng/ml FGF10 + 150 ng/ml Noggin + 10 µM SB-431542
Day 6	ADV-BSA	2 µM RA + 50 ng/ml FGF10 + 150 ng/ml Noggin + 10 µM SB-431542
Day 7	ADV-BSA	2 µM RA + 50 ng/ml FGF10 + 150 ng/ml Noggin + 0.25 µg/ml KAAD-cyclopamine
Day 8	ADV-BSA	2 µM RA + 50 ng/ml FGF10 + 150 ng/ml Noggin + 0.25 µg/ml KAAD-cyclopamine
Day 9	ADV-BSA	2 µM RA + 50 ng/ml FGF10 + 150 ng/ml Noggin + 0.25 µg/ml KAAD-cyclopamine
Day 10	ADV-BSA	2 µM RA + 50 ng/ml FGF10 + 0.25 µg/ml KAAD-cyclopamine
Day 11	ADV-BSA	2 µM RA + 50 ng/ml FGF10 + 0.25 µg/ml KAAD-cyclopamine
Day 12	ADV-BSA	2 µM RA + 50 ng/ml FGF10 + 0.25 µg/ml KAAD-cyclopamine
Days 13-15	ADV-BSA	2 µM RA + 1% B27 + 1 µM DAPT + 0.1 mM BNZ
Days 16-18	ADV-BSA	2 µM RA + 1% B27 + 0.25 µg/ml KAAD-cyclopamine
Days 18-21	ADV-BSA	2 µM RA + 1% B27 + 0.25 µg/ml KAAD-cyclopamine
Days 22-24	ADV-BSA	2 µM RA + 1% B27 + 0.25 µg/ml KAAD-cyclopamine

Table 14. Volume of media used for pancreatic differentiation

Culture vessel	Volume
10 cm dish	7 ml each day from Days 4 to 9, 14 ml on Days 13, 16, 18 and 22
1 well of 12wp	0.5 ml each day from Days 4 to 9, 1.5 ml on Days 13, 16, 18 and 22
1 well of 24wp	0.3 ml each day from Days 4 to 9, 1 ml on Days 13, 16, 18 and 22

2.1.5. Pancreatic differentiation using STEMdiff pancreatic progenitor kit

Alternatively, cells were also differentiated into pancreatic progenitors using the STEMdiff pancreatic progenitor kit (Stem Cell Technologies, #05120). They were cultured in E8 media as described in section 2.1.1 and plated for differentiation as described in section 2.1.3. Cells were differentiated according to the manufacturer's protocol and were harvested on Days 4 and 14 for downstream FACS analyses of the DE and pancreatic progenitors respectively, and Days 4, 7, 10 and 14 for downstream qRT-PCR analyses of the DE, primitive gut tube, posterior foregut and pancreatic progenitors respectively.

2.1.6. Glucose response assay and Enzyme linked immunosorbent assay (ELISA) for C-peptide

hPSCs differentiated into the endocrine progenitors after Day 24 are cultured in differentiation medium without insulin for 24 hr prior to ELISA. Cells were washed thrice in D-PBS and pre-incubated in DMEM supplemented with 2.5 mM glucose (Invitrogen) for 60 min at 37°C. To perform glucose-induced insulin secretion, the buffer is replaced with 22.5 mM glucose (high glucose) alternatively with 5.5 mM glucose (low glucose) twice at 30 min intervals. The supernatant is collected after each round of incubation and stored at -80°C for determination of C-peptide release. For control, cells were incubated in DMEM supplemented with 2.5 mM glucose.

C-peptide ELISA was measured using the Mercodia Ultrasensitive C-peptide ELSIA kit (Mercodia) following the manufacturer's instructions. Absorbance was read at 450 nm using Tecan Infinite 200 Pro plate reader.

2.1.7. Hepatic differentiation

Following on from DE differentiation, cells were further specified toward the hepatic lineage by using a protocol that was previously established in the lab. From Day 4 to Day 9, cells were cultured in 0.5 ml per well of a 12 well plate of RPMI-B27 (Table 11) and supplemented with 50 ng/ml Activin A (Table 15). Media was refreshed daily.

From Days 10 to 26, cells were cultured in 1 ml per well of a 12 well plate of Hepatozyme (Table 16) and supplemented with 10ng/ml Oncostatin M (OSM; R&D, #295-OM) and 50ng/ml Hepatocyte growth factor (HGF; Peprotech, #100-39). Media was refreshed every other day.

Table 15. Hepatic differentiation protocol

Time point	Media	Supplements	Volume
Day 4	RPMI-B27	50 ng/ml Activin A	0.5 ml
Day 5	RPMI-B27	50 ng/ml Activin A	0.5 ml
Day 6	RPMI-B27	50 ng/ml Activin A	0.5 ml
Day 7	RPMI-B27	50 ng/ml Activin A	0.5 ml
Day 8	RPMI-B27	50 ng/ml Activin A	0.5 ml
Day 9	RPMI-B27	50 ng/ml Activin A	0.5 ml
Day 10-11	Hepatozyme	10 ng/ml OSM + 50 ng/ml HGF	1 ml
Day 12-13	Hepatozyme	10 ng/ml OSM + 50 ng/ml HGF	1 ml
Day 14-15	Hepatozyme	10 ng/ml OSM + 50 ng/ml HGF	1 ml
Days 16-17	Hepatozyme	10 ng/ml OSM + 50 ng/ml HGF	1 ml
Days 18-19	Hepatozyme	10 ng/ml OSM + 50 ng/ml HGF	1 ml
Days 20-21	Hepatozyme	10 ng/ml OSM + 50 ng/ml HGF	1 ml
Days 22-23	Hepatozyme	10 ng/ml OSM + 50 ng/ml HGF	1 ml
Days 24-26	Hepatozyme	10 ng/ml OSM + 50 ng/ml HGF	1 ml

Table 16. Hepatozyme formulation

Component	Volume
Hepatozyme (Gibco, #17705-021)	500 ml
MEM Non-Essential Amino Acids Solution (Gibco, #11140-050)	10.6 ml
Chemically Defined Lipid Concentrate (Gibco, #11905031)	10.6 ml
L-Glutamine (Gibco, #25030024)	5.3 ml
Transferrin (30 mg/ml; Roche #652202)	530 µl
Insulin (10 mg/ml; Roche #1376497)	742 µl
+/- Penicillin-Streptomycin (Gibco, #15140122)	5.3 ml

2.2. Cloning of plasmid DNA constructs

2.2.1. Transformation of plasmids into *Escherichia coli* cells

XL-10 Gold Ultracompetent cells (Agilent Technologies, #200314) were thawed on ice. The competent cells were gently mixed by gentle tapping of the tube then 50 µl of competent cells were aliquoted into tubes containing DNA plasmids or ligation reactions while keeping all tubes on ice. The tubes were gently tapped to mix. The cells were incubated on ice for 20 min, then heat-shocked for 45 s in a 42°C water bath. They were then placed back on ice for a further 2 min. 500 µl of room temperature S.O.C. Medium (Invitrogen) was then added to the tubes and they were incubated shaking at 225 rpm for 1 hr at 37°C. 200-500 µl of the reaction mix was then spread on LB plates containing the appropriate antibiotics and incubated overnight at 37°C.

2.2.2. Small scale DNA plasmid purification and colony selection

Plasmid DNA constructs were prepared using QIAprep Spin Miniprep kit (Qiagen, #27106). They were analysed by restriction digest to confirm presence of the insert in the correct orientation.

2.2.3. Genotyping via Sanger sequencing

Plasmid DNA constructs were sent to AITBiotech, Singapore and GATC Biotech, UK for genotype verification.

2.3. Constructs

For DNA plasmid constructs used in cell culture, they were prepared using EndoFree Plasmid Maxi Kit (Qiagen, #12362). For all other applications, DNA plasmid constructs were prepared using QIAprep Spin Miniprep kit (Qiagen, #27106).

2.3.1. Transcription activator-like effector nuclease (TALEN) vectors

Suitable TALEN target sites on the *GATA6* gene were generated using an online TALEN targeter software tool (Cermak et al., 2011, Doyle et al., 2012). Two TALEN targets within exon 2 of the *GATA6* gene were selected based on higher numbers of RVDs HD and NH and the presence of a restriction site in the spacer region; one targeting a site that is 6 base pairs downstream of the first start codon while the other targeting a site that is 149 base pairs downstream of the second start codon. They are termed TALEN 1 and TALEN 2 respectively. The sequences of the two selected TALEN target pairs are listed in Table 17.

Table 17. Two selected TALEN target sites for *GATA6*

TALEN pair	Name	Sequence (5' → 3')
TALEN 1 (T1)	Left arm	GACTGACGGCGGCTGGT
	Right arm	CCGCACCCGCGGCCCCG
TALEN 2 (T2)	Left arm	GCTGCCCGGCCTACCGT
	Right arm	GGCTGGCCCACTGCCC

TALEN vectors were assembled using the Joung Lab REAL Assembly TALEN kit (Addgene, #1000000017). Modifications of the pTAL scaffold into the second generation GoldyTALEN scaffold and NN RVDs into NH were performed by Dr Norihiro Tsuneyoshi. Plasmids obtained from this kit used for the TALEN assemblies and their corresponding RVDs are listed in Table 18. The selected target sequences were entered into a ZiFiT targeter software via the REAL assembly method prior to TALEN assembly (Sander et al., 2007, Sander et al., 2010). Assembly of the TALE repeat arrays were performed in three polymerase chain reaction (PCR) steps (Table 19 for T1 and Table 20 for T2) with the primers listed in Table 21. Primers used in the final PCR step included the Kozak sequence. PCR was performed using PrimeSTAR Max DNA polymerase (Takara, # R045B) according to the manufacturer's protocol.

Table 18. Plasmids used and their corresponding RVDs

Plasmid	RVD
TAL006	NI
TAL007	HD
TAL009	NN to NH
TAL010	NG
TAL011	NI
TAL012	HD
TAL014	NN to NH
TAL015	NG
TAL016	NI
TAL017	HD
TAL019	NN to NH
TAL020	NG
TAL021	NI
TAL022	HD
TAL024	NN to NH
TAL025	NG

Table 19. TALEN 1 repeat array assembly via three step PCR method

PCR step	TALEN name	Primer name	RVD number	Plasmid
1	T1 left arm	TALEN-RVDs 1 Fwd TALEN-RVDs 1 Rev	1	TAL009
			5	TAL009
			9	TAL009
			13	TAL007
		TALEN-RVDs 2 Fwd TALEN-RVDs 2 Rev	2	TAL011
			6	TAL011
			10	TAL012
			14	TAL015
		TALEN-RVDs 3 Fwd TALEN-RVDs 3 Rev	3	TAL017
			7	TAL017
			11	TAL019
			15	TAL019
		TALEN-RVDs 4 Fwd TALEN-RVDs 4 Rev	4	TAL025
			8	TAL024
			12	TAL024
			16	TAL024
	T1 right arm	TALEN-RVDs 1 Fwd TALEN-RVDs 1 Rev	1	TAL007
			5	TAL006
			9	TAL009
			13	TAL007

		TALEN-RVDs 2 Fwd TALEN-RVDs 2 Rev	2 6 10 14	TAL012 TAL012 TAL012 TAL012
		TALEN-RVDs 3 Fwd TALEN-RVDs 3 Rev	3 7 11 15	TAL019 TAL017 TAL019 TAL017
		TALEN-RVDs 4 Fwd TALEN-RVDs 4 Rev	4 8 12 16	TAL022 TAL022 TAL024 TAL022
2	T1 left arm	TALEN-RVDs 1 Fwd TALEN-RVDs 2 Rev	1 2	TAL009, TAL011
		TALEN-RVDs 3 Fwd TALEN-RVDs 4 Rev	3 4	TAL017, TAL025
		TALEN-RVDs 1 Fwd TALEN-RVDs 2 Rev	5 6	TAL009, TAL011
		TALEN-RVDs 3 Fwd TALEN-RVDs 4 Rev	7 8	TAL017, TAL024
		TALEN-RVDs 1 Fwd TALEN-RVDs 2 Rev	9 10	TAL009, TAL012
		TALEN-RVDs 3 Fwd TALEN-RVDs 4 Rev	11 12	TAL019, TAL024
		TALEN-RVDs 1 Fwd TALEN-RVDs 2 Rev	13 14	TAL007, TAL015
		TALEN-RVDs 3 Fwd TALEN-RVDs 4 Rev	15 16	TAL019, TAL024
	T1 right arm	TALEN-RVDs 1 Fwd TALEN-RVDs 2 Rev	1 2	TAL007, TAL012
		TALEN-RVDs 3 Fwd TALEN-RVDs 4 Rev	3 4	TAL019, TAL022
		TALEN-RVDs 1 Fwd TALEN-RVDs 2 Rev	5 6	TAL006, TAL012
		TALEN-RVDs 3 Fwd TALEN-RVDs 4 Rev	7 8	TAL017, TAL022
		TALEN-RVDs 1 Fwd TALEN-RVDs 2 Rev	9 10	TAL009, TAL012
		TALEN-RVDs 3 Fwd TALEN-RVDs 4 Rev	11 12	TAL019, TAL024
		TALEN-RVDs 1 Fwd TALEN-RVDs 2 Rev	13 14	TAL007, TAL012
		TALEN-RVDs 3 Fwd TALEN-RVDs 4 Rev	15 16	TAL017, TAL022

3	T1 left arm	TALEN-RVDs 1-4 Fwd TALEN-RVDs 1-4 Rev	1 2 3 4	TAL009, TAL011, TAL017, TAL025
		TALEN-RVDs 5-8 Fwd TALEN-RVDs 5-8 Rev	5 6 7 8	TAL009, TAL011, TAL017, TAL024
		TALEN-RVDs 9-12 Fwd TALEN-RVDs 9-12 Rev	9 10 11 12	TAL009, TAL012, TAL019, TAL024
		TALEN-RVDs 13-16 Fwd TALEN-RVDs 13-16 Rev	13 14 15 16	TAL007, TAL015, TAL019, TAL024
	T1 right arm	TALEN-RVDs 1-4 Fwd TALEN-RVDs 1-4 Rev	1 2 3 4	TAL007, TAL012, TAL019, TAL022
		TALEN-RVDs 5-8 Fwd TALEN-RVDs 5-8 Rev	5 6 7 8	TAL006, TAL012, TAL017, TAL022
		TALEN-RVDs 9-12 Fwd TALEN-RVDs 9-12 Rev	9 10 11 12	TAL009, TAL012, TAL019, TAL024
		TALEN-RVDs 13-16 Fwd TALEN-RVDs 13-16 Rev	13 14 15 16	TAL007, TAL012, TAL017, TAL022

Table 20. TALEN 2 repeat array assembly via three step PCR method

PCR step	TALEN name	Primer name	RVD number	Plasmid
1	T2 left arm	TALEN-RVDs 1 Fwd TALEN-RVDs 1 Rev	1 5 9 13	TAL009 TAL007 TAL009 TAL006
		TALEN-RVDs 2 Fwd TALEN-RVDs 2 Rev	2 6 10 14	TAL012 TAL012 TAL012 TAL012

		TALEN-RVDs 3 Fwd TALEN-RVDs 3 Rev	3	TAL020
			7	TAL017
			11	TAL017
			15	TAL017
		TALEN-RVDs 4 Fwd TALEN-RVDs 4 Rev	4	TAL024
			8	TAL024
			12	TAL025
			16	TAL024
	T2 right arm	TALEN-RVDs 1 Fwd TALEN-RVDs 1 Rev	1	TAL009
			5	TAL009
			9	TAL007
			13	TAL009
		TALEN-RVDs 2 Fwd TALEN-RVDs 2 Rev	2	TAL014
			6	TAL014
			10	TAL011
			14	TAL012
		TALEN-RVDs 3 Fwd TALEN-RVDs 3 Rev	3	TAL017
			7	TAL017
			11	TAL017
			15	TAL017
		TALEN-RVDs 4 Fwd TALEN-RVDs 4 Rev	4	TAL025
			8	TAL022
			12	TAL025
2	T2 left arm	TALEN-RVDs 1 Fwd TALEN-RVDs 2 Rev	1 2	TAL009, TAL012
		TALEN-RVDs 3 Fwd TALEN-RVDs 4 Rev	3 4	TAL020, TAL024
		TALEN-RVDs 1 Fwd TALEN-RVDs 2 Rev	5 6	TAL007, TAL012
		TALEN-RVDs 3 Fwd TALEN-RVDs 4 Rev	7 8	TAL017, TAL024
		TALEN-RVDs 1 Fwd TALEN-RVDs 2 Rev	9 10	TAL009, TAL012
		TALEN-RVDs 3 Fwd TALEN-RVDs 4 Rev	11 12	TAL017, TAL025
		TALEN-RVDs 1 Fwd TALEN-RVDs 2 Rev	13 14	TAL006, TAL012
		TALEN-RVDs 3 Fwd TALEN-RVDs 4 Rev	15 16	TAL017, TAL024
	T2 right arm	TALEN-RVDs 1 Fwd TALEN-RVDs 2 Rev	1 2	TAL009, TAL014
		TALEN-RVDs 3 Fwd TALEN-RVDs 4 Rev	3 4	TAL017, TAL025

		TALEN-RVDs 1 Fwd	5	TAL009,
		TALEN-RVDs 2 Rev	6	TAL014
		TALEN-RVDs 3 Fwd	7	TAL017,
		TALEN-RVDs 4 Rev	8	TAL022
		TALEN-RVDs 1 Fwd	9	TAL007,
		TALEN-RVDs 2 Rev	10	TAL011
		TALEN-RVDs 3 Fwd	11	TAL017,
		TALEN-RVDs 4 Rev	12	TAL025
		TALEN-RVDs 1 Fwd	13	TAL009,
		TALEN-RVDs 2 Rev	14	TAL012
		TALEN-RVDs 3 Fwd	15	TAL017
		TALEN-RVDs 4 Rev		
3	T2 left arm	TALEN-RVDs 1-4 Fwd	1	TAL009,
		TALEN-RVDs 1-4 Rev	2	TAL012,
			3	TAL020,
			4	TAL024
		TALEN-RVDs 5-8 Fwd	5	TAL007,
		TALEN-RVDs 5-8 Rev	6	TAL012,
			7	TAL017,
			8	TAL024
		TALEN-RVDs 9-12 Fwd	9	TAL009,
		TALEN-RVDs 9-12 Rev	10	TAL012,
			11	TAL017,
			12	TAL025
		TALEN-RVDs 13-16 Fwd	13	TAL006,
		TALEN-RVDs 13-16 Rev	14	TAL012,
			15	TAL017,
			16	TAL024
	T2 right arm	TALEN-RVDs 1-4 Fwd	1	TAL009,
		TALEN-RVDs 1-4 Rev	2	TAL014,
			3	TAL017,
			4	TAL025
		TALEN-RVDs 5-8 Fwd	5	TAL009,
		TALEN-RVDs 5-8 Rev	6	TAL014,
			7	TAL017,
			8	TAL022
		TALEN-RVDs 9-12 Fwd	9	TAL007,
		TALEN-RVDs 9-12 Rev	10	TAL011,
			11	TAL017,
			12	TAL025
		TALEN-RVDs 13-16 Fwd	13	TAL009,
		TALEN-RVDs 13-16 Rev	14	TAL012,
			15	TAL017

Table 21. Primers used to assemble TALEN repeat arrays

Primer name	Primer sequence (5' → 3')
TALEN-RVDs 1 Fwd	CTGACCCCAGACCAGGTAGTCGCA
TALEN-RVDs 1 Rev	CACGACTTGATCCGGTGTAAGGCCGTGGTCTTGACAAAGG
TALEN-RVDs 2 Fwd	CCTTTGTCAAGACCACGGCCTTACACCGGATCAAGTCGTG
TALEN-RVDs 2 Rev	TACAACTTGATCGGGAGTCAGCCCGTGGTCTTGACAGAGA
TALEN-RVDs 3 Fwd	TCTCTGTCAAGACCACGGGCTGACTCCCGATCAAGTTGTA
TALEN-RVDs 3 Rev	GACCACTTGGTCAGGCGTCAAACCGTGATCTTGACACAAC
TALEN-RVDs 4 Fwd	GTTGTGTCAAGATCACGGTTTGACGCCTGACCAAGTGGTC
TALEN-RVDs 4 Rev	TCCATGATCCTGGCACAGTACAGG
TALEN-RVDs 1-4 Fwd	tcagGGTCTCAGAACCTGACCCCAGACCAGGTAGTC
TALEN-RVDs 1-4 Rev	tcagGGTCTCTAGTCCATGATCCTGGCACAGT
TALEN-RVDs 5-8 Fwd	tcagGGTCTCAGACTGACCCCAGACCAGGTAGTC
TALEN-RVDs 5-8 Rev	tcagGGTCTCTGTCAGTCCATGATCCTGGCACAGT
TALEN-RVDs 9-12 Fwd	tcagGGTCTCATGACCCCAGACCAGGTAGTC
TALEN-RVDs 9-12 Rev	tcagGGTCTCTCAGTCCATGATCCTGGCACAGT
TALEN-RVDs 13-16 Fwd	tcagGGTCTCAACTGACCCCAGACCAGGTAGTC
TALEN-RVDs 13-16 Rev	tcagGGTCTCTTCAGTCCATGATCCTGGCACAGT

Plasmids derived from PCR steps 1 and 2 were purified by gel extraction using QIAquick Gel Extraction Kit (Qiagen, #28706) and the final TALE repeat array plasmids derived from PCR step 3 were purified using QIAquick PCR Purification Kit (Qiagen, #28106). The TALE repeat arrays were then digested with restriction enzyme BsaI-HF (New England Biolabs, #R3535S) for 2 hours at 37°C then purified by gel extraction, while the respective vector backbones for each TALE repeat array (Table 22) were digested with restriction enzyme BsmBI (New England Biolabs, #R0580S) for 3 hours at 55°C then purified by gel extraction. The TALE repeat array were next ligated with their respective vector backbones by using T4 DNA ligase (New England Biolabs, #M0202S) for 2 hours at room temperature. The success of the TALEN assembly was verified by Sanger sequencing (Table 23).

Table 22. TALE repeat arrays and their corresponding vector backbone

TALEN name	Vector backbone
T1 left arm	JDS78
T1 right arm	JDS74
T2 left arm	JDS78
T2 right arm	JDS71

Table 23. Primers used for sequencing TALEN assembly

Primer name	Primer sequence (5' → 3')
TALEN-006seq Fwd	TCGCAATCGCGTCGAACATTG
TALEN-006seq Rev	GCTTGCTTTCCCCCAATGTTC
TALEN-007seq Fwd	TCGCAATCGCGTCACATGAC
TALEN-009seq Fwd	TCGCAATCGCGTCAAATCAT
TALEN-009seq Rev	CTTGCTTTCCCCCATGATTT
TALEN-010seq Fwd	TCGGCAATCGCGTCAAACGGA
TALEN-010seq Rev	CTTGCTTTCCCCCTCCGTTT
TALEN-011seq Fwd	GTGGCCATTGCAAGCAACATC
TALEN-011seq Rev	GAGCCTGTTTGCCACCGATG
TALEN-012seq Fwd	TGGCCATTGCATCCCACGAC
TALEN-012seq Rev	CTGTTTGCCACCGTCGTGG
TALEN-020seq Rev	AATGCTTGTTTCCCTCCACCG
TALEN-022seq Rev	CTGCTTACCGCCATCATGG
TALEN-024NHseq Rev	AGCGCCTGCTTACCGCCATG
TALEN-025seq Fwd	TCGCCATCGCCTCGAATGGC
TALEN-025seq Rev	CTGCTTACCGCCGCCATTC

Next, the assembled TALEN arms were cloned into vectors containing a CAG promoter and an antibiotic resistance gene (Table 24) that were already available in the lab. To do this, the TALEN arms were PCR amplified using PrimeSTAR Max DNA polymerase using primers listed in Table 25 and PCR conditions listed in Table 26. They were then purified using QIAquick PCR Purification Kit, digested by NheI restriction enzyme (New England Biolabs, #R0131S), blunted using Quick Blunting Kit (New England Biolabs, #E1201S) for 5 min at 72°C, purified using QIAquick PCR Purification Kit once more, digested by XhoI restriction enzyme (New England Biolabs, #R0146S), then gel purified by gel extraction. For the vectors, they were first digested by EcoRI-HF restriction enzyme (New England Biolabs, # R3101S), blunted as described earlier, purified using QIAquick PCR Purification Kit, digested by XhoI restriction enzyme, then gel purified by gel extraction. The TALEN arms were next ligated with their respective vector containing the appropriate antibiotic resistance gene by using T4 DNA ligase for 2 hours at room temperature. The final TALEN constructs were then sequenced to confirm that the TALEN arms were cloned in the correct orientation (Table 27).

Table 24. Antibiotic resistance gene specific to each TALEN arm

TALEN construct name	Antibiotic resistance gene
T1 left arm	Puromycin
T1 left arm	Blasticidin
T1 right arm	Zeocin
T2 left arm	Puromycin
T2 right arm	Zeocin

Table 25. Primers used to amplify TALEN arms

Primer name	Primer sequence (5' → 3')
T7 fwd	AATACGACTCACTATAG
TALEN-pCAG-IRES Rev	AACTTTTAAACCGGTCTCGAGCTGA

Table 26. Parameters for PCR cycling reaction to amplify TALEN arms

Step	Temperature	Time	Number of cycles
Denaturation Annealing Extension	98°C 55°C 72°C	10 s 15 s 5 s/1kb	35
Soak	4°C	Indefinite	1

Table 27. Primers used for sequencing TALEN constructs

Primer name	Primer sequence (5' → 3')
pre-KpnI TALEN-Core-Fwd	GTTGTAAAACGACGGCCAGTG
025seq-Fwd	TCGCCATCGCCTCGAATGGC
post-BamHI TALEN-Core-Rev	TACGCCAAGCTTGCATGCAGG
Fok1-Fwd	GTGAACTGGAGGAGAAGAAATCTG
pCAG-IP +1760 Rev	GGGCGGAATTTACGTAGCGG
007seq-Fwd	TCGCAATCGCGTCACATGAC
006seq-Fwd	TCGCAATCGCGTCGAACATTG
011seq-Rev	GAGCCTGTTTGCCACCGATG

A donor plasmid aimed at terminating transcription of *GATA6* prematurely by inserting a 'donor template' through HR was also constructed. Within the donor plasmid is a cassette which contains 5' and 3' homology arms each 1kb in length recognising the flanking regions of the TALEN 1 target site, an EmGFP gene, a puromycin antibiotic resistant gene and a polyA tail. PCR amplification of genomic DNA from H9 hESC line and a vector already available in the lab containing the EmGFP gene, puromycin gene and polyA tail was performed to obtain PCR products of the 5' and 3' homology arms and the other components of the donor plasmid respectively (Table 28). PCR was performed using PrimeSTAR Max DNA polymerase and conditions listed in Table 26. The PCR products were then ligated and the final construct was sequenced to confirm that the donor plasmid was cloned successfully (Table 29).

Table 28. Primers used to construct the donor plasmid

Primer name	Primer sequence (5' → 3')
5' Arm-KpnI-GATA6 Fwd	tcagGGTACCTTTGGGGTCGCCTCGGCTCTGG
5' Arm-GATA6 Rev	CTTGCTCACCATGGTGGCCACGGTCCGGCGCCGCTCCAA
5' Arm-GATA6-emGFP Fwd	CGCCGGACCGTGGCCACCATGGTGAGCAAGGGCGAGGAGC
3' Arm-XbaI-TALEN1 Fwd	tcagTCTAGAAAGCGCTTCGGGGCCGCGGGTG
3' Arm-SacI-TALEN1 Rev	tcagGAGCTCTGGCGCCCCCACGTAGGGCGAG

Table 29. Primers used for sequencing donor plasmid

Primer name	Primer sequence (5' → 3')
EmGFP3'-Fwd	TCACATGGTCCTGCTGGAGTTC
BGHpA-mid-Rev	TTAGGAAAGGACAGTGGGAGTG
EmGFP5'-Rev	CGCTGAACTTGTGGCCGTTTAC
EmGFP-mid-Rev	GACCTTGTGGCTGTTGTAGTTG
mPGKpA-Fwd	AAGAAGGGTGAGAACAGAGTACC
M13-Rev (-24)	GGAAACAGCTATGACCATG
M13-Fwd (-20)	GTAAAACGACGGCCAGT
pCAGGS pre-SA Fwd	CTGCTAACCATGTTCATGCCTTC

2.3.2. Introducing constructs into hPSCs

Electroporation using the Amaxa Nucleofector Technology (Lonza, Human Stem Cell Nucleofector Kit 1) was used to introduce the TALEN constructs and donor plasmid into the hPSCs. Cells were grown in 6 well plates to a confluency of 70-80%, then washed once with D-PBS and treated with 0.5 ml of StemPro Accutase Cell Dissociation Reagent (Gibco, #A1110501) per well for 5 min at 37°C. 1 ml of 5% fetal bovine serum (FBS; Gibco, #10082147) diluted in D-PBS was added per well and the cells were detached from the wells by gentle pipetting. Cells were collected in a 15 ml tube and an aliquot was taken for cell counting. Cell counting was performed using an automated cell counter (Biorad, TC20). 8×10^5 cells were used per electroporation and the appropriate volume that contained 8×10^5 cells was taken and divided into individual 15 ml tubes. The tubes were centrifuged at 1,200 rpm to pellet the cells. Cells were electroporated with a pre-determined amount of DNA for each construct. Electroporation was performed according to the manufacturer's protocol using Nucleofector programme B-016. The electroporated cells were next plated as single cells onto 10 cm dishes containing 10 ml of E8 with ROCK inhibitor.

2.4. Generation of *GATA6* mutant lines

2.4.1. Non-homologous end joining (NHEJ) pathway

TALENs were introduced into H9 and FSPS13.B hPSC lines via electroporation as described in section 2.3.2. For electroporation, 2.5 µg of DNA for each of the corresponding TALEN arm was used (Table 30). 24 hr after electroporation, simultaneous antibiotic selection of 1 µg/ml puromycin (Sigma-Aldrich, #P8833) and 2.5 µg/ml zeocin (Gibco, #R250-01) was done for 24 hr, after which the cells were allowed to recover and form colonies. Colonies formed approximately 7 days after the antibiotic selection, and they were individually picked using a pipette tip, re-plated and expanded. During picking of each colony, half of one colony was re-plated whereas the other half was pipetted into 8-tube PCR strips for subsequent screening.

Table 30. Electroporation of TALENs into hPSCs

TALEN target site	TALEN construct name	Antibiotic resistance gene	Amount of DNA electroporated
1	T1 left arm	Puromycin	2.5 µg
	T1 right arm	Zeocin	2.5 µg
2	T2 left arm	Puromycin	2.5 µg
	T2 right arm	Zeocin	2.5 µg

Screening of the colonies was performed by first extracting the genomic DNA of each colony using a direct PCR approach. 150-200 µl of D-PBS was added to each tube on the PCR strips. The strips were next centrifuged for 10 s to pellet the cells. The supernatant was removed carefully and 10 µl of Proteinase K reaction mix was added to each tube containing cells from one colony (Table 31 and Table 32). The strips were then loaded onto a PCR machine and subjected to conditions listed in Table 33 for genomic extraction. PCR of the genomic region flanking the TALEN 1 and 2 target sites was performed using PrimeSTAR GXL DNA polymerase (Takara, #R050A) and the reagents were set up as shown in Table 34. PCR cycling conditions are shown in Table 35 using primers listed in Table 36.

Table 31. 50X detergent mix formulation

Component	Volume
Tween-20	10 µl
Igepal CA-630	10 µl
dH_2O	180 µl

Table 32. Proteinase K reaction mix formulation

Component	Volume
50x detergents mix	0.2 µl
5x PrimeSTAR GXL PCR Buffer	2 µl
20 mg/mL Proteinase K	0.2 µl
dH_2O	7.6 µl

Table 33. Parameters for PCR cycling conditions for genomic DNA extraction

Temperature	Time
50°C	30 min
95°C	5 min
4°C	Indefinitely

Table 34. Volume of reagents in genomic PCR reaction

Component	Volume
dH_2O	13.4 µl
5x PrimeSTAR GXL PCR Buffer	4 µl
2.5 mM dNTP	1.6 µl
100 µM Primer Fwd	0.05 µl
100 µM Primer Rev	0.05 µl
Genomic DNA	0.5 µl
PrimeSTAR GXL DNA polymerase	0.4 µl

Table 35. Parameters for PCR to amplify TALEN 1 and 2 genomic regions

TALEN target site	Step	Temperature	Time	Number of cycles
1	Denaturation Annealing Extension	98°C 55°C 72°C	10 s 5 s 5 s	30
	Soak	4°C	∞	1
2	Denaturation Annealing Extension	98°C 62°C 72°C	10 s 5 s 5 s	30
	Soak	4°C	∞	1

Table 36. Primers used to amplify TALEN 1 and 2 genomic regions

TALEN target site		Primer sequence (5' → 3')
1	F	CTTTGAGAAGTCAGATCCCATTTGA
	R	CGCCTCCGCTGCCGTATGGAGGGCT
2	F	CGCCAGCAAGCTGCTGTGGTCCAGC
	R	TCCGCGCACCCGGACGAGAAAGTCC

The PCR products were next treated with ExoSAP-IT (Affymetrix, #78250) then digested with restriction enzymes AfeI (New England Biolabs, #R0652S) and PstI (New England Biolabs, #R0140S) for TALEN 1 and TALEN2 PCR products respectively for 1 hr at 37°C. The digested PCR products were analysed by agarose gel electrophoresis as a first pass screen for any mutations. PCR products of the colonies that showed a potential for the occurrence of mutations were next sent for sequencing as a second pass screen and confirmation for mutations using forward primers listed in Table 36.

2.4.2. Homologous recombination (HR) pathway

TALENs were introduced into H9 and FSPS13.B hPSC lines via electroporation as described in section 2.3.2. For electroporation, 2 μg of DNA for each TALEN arm targeting the TALEN 1 site and 1 μg of DNA for the donor plasmid was used (Table 37). 24 hr after electroporation, simultaneous antibiotic selection of 3.5 μg/ml blasticidin and 1 μg/ml puromycin was done for 24 hr. After which, simultaneous antibiotic selection of 3.5 μg/ml blasticidin and 2.5 μg/ml zeocin was done for the next 24 hours. Colonies formed approximately 7 days after the antibiotic selection, and they were individually picked using a pipette tip, re-plated and expanded. Genomic DNA was performed the same way as described in section 2.4.1. Successful HR was determined by PCR using PrimeSTAR GXL DNA polymerase (Takara, #R050A) with reaction mix listed in Table 38. The PCR cycling conditions are shown in Table 39 using forward and reverse primers for the TALEN 1 target site listed in Table 36. Colonies that showed positive HR via PCR were confirmed by sequencing.

Table 37. Electroporation of TALENs with donor plasmid into hPSCs

TALEN target site	TALEN construct name	Antibiotic resistance gene	Amount of DNA electroporated
1	T1 left arm	Blasticidin	2 μg
	T1 right arm	Zeocin	2 μg
	Donor plasmid	Puromycin	1 μg

Table 38. Volume of reagents in PCR reaction to check for HR

Component	Volume
dH_2O	13.4 μl
5x PrimeSTAR GXL PCR Buffer	4 μl
2.5 mM dNTP	1.6 μl
100 μM Primer Fwd	0.05 μl
100 μM Primer Rev	0.05 μl
Genomic DNA	50 ng
PrimeSTAR GXL DNA polymerase	0.4 μl

Table 39. Parameters for PCR cycling reaction to check for HR

TALEN target site	Step	Temperature	Time	Number of cycles
1	Initial denaturation	94°C	3 min	1
	Denaturation Annealing Extension	98°C 62°C 68°C	10 s 15 s 1 min	30
	Final extension	68°C	2 min	1
	Soak	4°C	∞	1

2.5. Western blot

2.5.1. Cell lysate preparation and normalisation

hPSCs or their differentiated progenitors were washed once in D-PBS and incubated in 0.5 ml of Accutase per well of a 6 well plate for 5 min at 37°C. The Accutase was neutralised by adding 1 ml of 5% FBS diluted in D-PBS per well and the cells were dissociated by gentle pipetting. The cells were washed twice with D-PBS and pelleted by centrifuging at 1,200 rpm. The pelleted cells were re-suspended in 50-200 µl of Lysis Buffer (Table 40) containing freshly added inhibitors cOmplete Protease Inhibitor Cocktail (Roche, #11697498001), Sodium Fluoride (NaF; New England Biolabs, #P0759), Sodium Vanadate (Na_3VO_4; New England Biolabs, #P0758) according to Table 41. One tablet of the cOmplete Protease Inhibitor Cocktail was dissolved in 1 ml of Lysis Buffer to make up a 25x stock solution. The cell lysates were kept on ice for at least 15 min, vortexed at maximum speed for 15 s then centrifuged for 30 min at 15,000 g at 4°C. The supernatants were collected and protein concentrations were determined by Bradford assay (Protein Assay Dye Reagent Concentrate, Bio-Rad) according to the manufacturer's protocol. The protein concentrations of the cell lysates were normalised to 10 µg of protein for probing with *GATA6* and *GATA4* and 1 µg for probing with alpha-tubulin. The normalised cell lysates were heat denatured at 98°C in the presence of Laemmli Sample Buffer (Bio-Rad) and β-mercaptoethanol for 5 min, then subjected to SDS-PAGE electrophoresis.

Table 40. Lysis Buffer formulation

Component	Working concentration
1M Tris-Cl pH 7.5 (Cambridge Bioscience, #600201)	50mM
5M NaCl (Ambion, #AM9759)	150mM
Triton X-100 (Sigma-Aldrich, #T8787)	1%
Glycerol (Sigma-Aldrich, #49781)	10%
Deocycholate (Sigma-Aldrich, #D6750)	0.1%
β-glycerophosphate (Sigma-Aldrich, #G5422)	25mM

Table 41. Lysis Buffer with protease inhibitors formulation

Component	Volume
25x Complete solution	40 µl
Lysis Buffer	930 µl
50 mM NaF	20 µl
100 mM Na_3VO_4	10 µl

2.5.2. SDS-page, blotting and blocking

Electrophoresis of NuPAGE Novex 4-12% Bis-Tris Protein Gels was performed using the XCell SureLock Mini-Cell (Invitrogen). 1x SDS Running Buffer was prepared by adding 50 ml of 20x NuPAGE MOPS SDS Running Buffer to 950 ml deionised water. Precision Plus protein kaleidoscope standards (Bio-Rad, #161-0375) protein ladder was used. Electrophoresis was carried out at a constant volt of 120 V until the ladders were completely separated. The separated proteins were next transferred from the gel onto Immun-Blot PVDF membrane (Bio-Rad, #162-0177) using Mini Trans-Blot Cell (Bio-Rad) at 25 V overnight at 4°C.

2.5.3. Antibody incubation and detection

The membranes were incubated shaking in 5% Blotting-Grade Blocker (Bio-Rad, #170-6404) diluted in 0.1% Triton X-100 in D-PBS (PBST) for 1 hr at room temperature. The blocking solution was removed and primary antibodies (Table 42) diluted in PBST were added to the membranes and incubated shaking for 2 hr at room temperature. Membranes were then subjected to three 10 min washes in PBST and incubated with horseradish peroxidase (HRP)-conjugated secondary antibodies in 5% Blotting-Grade Blocker in PBST for 1 hr at room temperature while shaking. Unbound antibodies were removed by three 10 min washes while shaking in PBST. Proteins were detected via chemiluminescence using SuperSignal West Femto Maximum Sensitivity Substrate (ThermoFisher Scientific, #PI34095) and finally developed using Amersham Hyperfilm ECL (GE Healthcare).

Table 42. Antibodies used in western blot experiments

Primary antibody	**Dilution ratio**	**Duration**	**Expected molecular weight**
Rabbit anti-human GATA6 (N-terminus; Cell Signaling, #5851)	1:2000	2 hr	Long isoform: 64 kDa Short isoform: 52 kDa
Rabbit anti-human GATA6 (C-terminus; Cell Signalling, #4253)	1:2000	2 hr	Short isoform: 52 kDa
Rabbit anti-human GATA4 (Cell Signalling, #36966)	1:2000	2 hr	55 kDa
Mouse anti-alpha-Tubulin (Sigma-Aldrich, #T6199)	1:5000	1 hr	50 kDa
Secondary antibody	**Dilution ratio**	**Duration**	
Anti-Rabbit IgG-Peroxidase antibody produced in goat (Sigma-Aldrich, #A6154)	1:10,000	1 hr	-
Anti-Mouse IgG-Peroxidase antibody produced in goat (Sigma-Aldrich, #A5278)	1:10,000	1 hr	-

2.6. Immunocytochemistry (ICC)

2.6.1. Fixation and blocking

Cells in 12 well plates were fixed by aspirating the culture media then immediately adding 500 µl of 4% paraformaldehyde (PFA; VWR, #43368.9M) solution diluted in D-PBS per well and incubating for 20 min at 4°C. They were then washed thrice in D-PBS. Cells were next incubated in 500 µl of PBST (0.1% Triton X-100 in D-PBS) containing 10% donkey serum (AbD Serotec, #C06SB) per well for 20 min at room temperature for blocking.

2.6.2. Antibody incubation and detection

Cells in 12 well plates were then incubated overnight at 4°C with 300 µl of primary antibodies (Table 43) diluted in PBST containing 1% donkey serum. Cells were then washed thrice with PBST to remove unbound primary antibodies and thereafter incubated with 300 µl of fluorescence-dye-conjugated secondary antibodies (Table 43) diluted in PBST containing 1% donkey serum in for 1 hr at room temperature. Unbound antibodies were removed by three 5 min washes in D-PBS. 4’,6-Diamidino-2-phenylindole dihydrochloride (DAPI; Sigma-Aldrich, #D-8417) at a dilution of 1:1000 was added to the first wash.

Table 43. Antibodies used in immunocytochemistry experiments

Primary antibody	**Dilution ratio**	**Duration**
Goat anti-human Nanog (R&D, #AF1997)	1:100	Overnight
Goat anti-human Sox2 (R&D, #AF2018)	1:100	Overnight
Goat anti-human Oct4 (Santa Cruz, #sc-8628)	1:100	Overnight
Rabbit anti-human Eomes (Abcam, #Ab23345)	1:100	Overnight
Rabbit anti-human GATA6 (Cell Signaling, #5851)	1:200	Overnight
Mouse anti-human GATA4 (Santa Cruz, #sc25310)	1:100	Overnight
Goat anti-human Sox17 (R&D, #AF1924)	1:200	Overnight
Goat anti-human FoxA2 (R&D, #AF2400)	1:100	Overnight
Mouse anti-human Hex (Abcam, #Ab117864)	1:100	Overnight
Mouse anti-human CDX2 (CDX-88; Abcam, #Ab86949)	1:100	Overnight
Goat anti-human HNF1B C-20 (Santa Cruz, #sc7411)	1:100	Overnight
Rabbit anti-human HNF4A H-171 (Santa Cruz, #sc8987)	1:100	Overnight
Rabbit anti-human HNF6 H-100 (Santa Cruz, #sc13050)	1:100	Overnight
Goat anti-human PDX1 (R&D, #AF2419)	1:100	Overnight
Rabbit anti-human Sox9 H-90 (Santa Cruz, #sc20095)	1:100	Overnight
Sheep anti-human NGN3 (R&D, #AF3444)	1:100	Overnight
Mouse anti-human C-Peptide (Acris Antibodies, #BM270S)	1:100	Overnight
Goat anti-human Glucagon G-17 (Santa Cruz, #sc7780)	1:100	Overnight
Rabbit anti-human Somatostatin (Daka, #A0566)	1:200	Overnight
Secondary antibody	**Dilution ratio**	**Duration**
Alexa Fluor 568 Donkey Anti-Goat IgG (H+L) (Invitrogen, #A11057)	1:1000	1 hr
Alexa Fluor 568 Donkey Anti-Mouse IgG (H+L) (Invitrogen, #A10037)	1:1000	1 hr
Alexa Fluor 568 Donkey Anti-Rabbit IgG (H+L) (Invitrogen, #A10042)	1:1000	1 hr
Alexa Fluor 568 Donkey Anti-Sheep IgG (H+L) (Invitrogen, #A21099)	1:1000	1 hr
Alexa Fluor 488 Donkey anti-Goat IgG (H+L) (Invitrogen, #A11055)	1:1000	1 hr
Alexa Fluor 488 Donkey anti-Mouse IgG (H+L) (Invitrogen, #A21202)	1:1000	1 hr
Alexa Fluor 488 Donkey anti-Rabbit IgG (H+L) (Invitrogen, #A21206)	1:1000	1 hr
Alexa Fluor 488 Donkey anti-Sheep IgG (H+L) (Invitrogen, #A11015)	1:1000	1 hr
Alexa Fluor 647 Donkey anti-Goat IgG (H+L) (Invitrogen, #A21447)	1:1000	1 hr
Alexa Fluor 647 Donkey anti-Mouse IgG (H+L) (Invitrogen, #A31571)	1:1000	1 hr
Alexa Fluor 647 Donkey anti-Rabbit IgG (H+L) (Invitrogen, #A31573)	1:1000	1 hr

2.7. Fluorescence activated cell sorting (FACS) analysis

2.7.1. Cell preparation

Cells in 12 well plates were washed twice in D-PBS and incubated in 0.3 ml of Accutase per well for 5 min at 37°C. The Accutase was neutralised by adding 0.6 ml of 5% FBS diluted in D-PBS and the cells were dissociated by gentle pipetting. Cells were re-suspended in D-PBS at approximately 0.1-1 x 10^6 cells/ml and washed twice with D-PBS. They were then pelleted and fixed by re-suspending in 500 µl of 4% PFA solution diluted in D-PBS per well and incubating at for 20 min at 4°C, then washed twice in D-PBS.

2.7.2. Antibody incubation and detection

Next, for all primary antibodies except CXCR4, cells were permeabilised in 500 µl of D-PBS containing 1% Saponin (Sigma-Aldrich, #47036-50G-F) for 30 min at room temperature. Cells were then incubated for 2 hr at room temperature with primary antibody (Table 44) diluted in 100 µl of Staining Solution (1% Saponin and 5% FBS in D-PBS). After which, they were washed three times with 1 ml of Staining Solution per wash and incubated with secondary antibodies (Table 44) diluted in 100 µl of Staining Solution for 30 min at room temperature. Unbound antibody was then removed by three washes in 1 ml of Staining Solution per wash and cells were re-suspended in 200 µl of 2% FBS diluted in D-PBS prior to analysis.

For CXCR4 staining, cells were fixed in 4% PFA and washed as described above. Thereafter, primary antibody (Table 44) diluted in 100 µl of 5% FBS in D-PBS was added to the cells and incubated for 1 hr at room temperature. Unbound antibody was then removed by three washes of 1ml 2% FBS in D-PBS per wash. Cells were then re-suspended in 200 µl of 2% FBS in PBS prior to analysis.

Analyses were performed using a BD LRSFortessa cell analyser (BD Biosciences). All flow cytometry experiments were gated using unstained cells. On all flow cytometry plots, the undifferentiated population is shown in blue. All gates shown on scatterplots were set according to the undifferentiated population control.

Table 44. Antibodies used in FACS experiments

Primary antibody	**Dilution ratio**	**Duration**
Goat anti-human Sox17 (R&D, #AF1924)	1:20	2 hr
Rabbit anti-human GATA6 (Cell Signaling, #5851)	1:20	2 hr
Goat anti-human PDX1 (R&D, #AF2419)	1:20	2 hr
Sheep anti-human NGN3 (R&D, #AF3444)	1:20	2 hr
Mouse anti-human C-Peptide (Acris Antibodies, #BM270S)	1:100	2 hr
Goat anti-human Glucagon G-17 (Santa Cruz, #sc7780)	1:20	2 hr
Rabbit anti-human Somatostatin (Daka, #A0566)	1:200	2 hr
Secondary antibody	**Dilution ratio**	**Duration**
Alexa Fluor 568 Donkey Anti-Goat IgG (H+L) (Invitrogen, #A11057)	1:1000	30 min
Alexa Fluor 568 Donkey Anti-Mouse IgG (H+L) (Invitrogen, #A10037)	1:1000	30 min
Alexa Fluor 568 Donkey Anti-Rabbit IgG (H+L) (Invitrogen, #A10042)	1:1000	30 min
Alexa Fluor 568 Donkey Anti-Sheep IgG (H+L) (Invitrogen, #A21099)	1:1000	30 min
Alexa Fluor 647 Donkey anti-Mouse IgG (H+L) (Invitrogen, #A31571)	1:1000	30 min
Alexa Fluor 488 Donkey anti-Rabbit IgG (H+L) (Invitrogen, #A21206)	1:1000	30 min
Conjugated primary and secondary antibody	**Dilution ratio**	**Duration**
Anti-Human CD184 (CXCR4) PE (eBioscience, #12-9999-41)	1:50	1 hr

2.8. Quantitative RT-PCR (qRT-PCR)

2.8.1. Total RNA isolation

Cells were grown in 12 well plates for total RNA isolation. 3 wells were harvested per sample to obtain technical triplicates. RNeasy Mini Kit together with the Qiacube (Qiagen) was used for total RNA extraction. Cell culture media was aspirated and the cells were washed once with D-PBS. The D-PBS was completely aspirated and cells were lysed directly in the 12 well plates by adding 350 µl of Buffer RLT. Cell lysates were transferred to 2 ml tubes and were either frozen at -80°C or used immediately with the Qiacube for RNA extraction. Each sample was treated with RNase-Free DNase (Qiagen) to avoid DNA contamination. RNA was eluted in a volume of 30 µl. RNA was either frozen at -80°C or immediately taken to the next step of first strand cDNA synthesis. If RNA samples were frozen, they were thawed on ice to prevent degradation.

2.8.2. First strand cDNA synthesis

500 ng of RNA samples were made up to a total volume of 11.875 µl with nuclease free water. The following components were then added to a nuclease-free 96 well plate (Table 45).

Table 45. Reagents to denature RNA and primer

Component	Volume
500 ng of total RNA	11.875 µl
Random primer (Promega, #C1181)	0.5 µl
dNTP (Promega, #U1511)	1 µl

The plate was centrifuged briefly to ensure that the reagents were at the bottom of the tube. The plate was incubated in a PCR machine for 5 min at 65°C then quickly chilled on ice. The plate was again centrifuged briefly and the reagents listed in Table 46 were prepared as a master mix then added to the samples. The final volume of each sample was 20 µl.

Table 46. Reagents for reverse transcription of RNA

Component	Volume
5x First-strand buffer	4 µl
0.1 M DTT	2 µl
RNaseOUT™ Recombinant Ribonuclease Inhibitor (Invitrogen, #10777019)	0.5 µl
SuperScript® II Reverse Transcriptase (Invitrogen, #18064014)	0.125 µl

The plate was again centrifuged briefly to ensure that all reagents were at the bottom of the tube. The plate was then incubated in a PCR machine programmed at 10 min at 25°C for the primer annealing step, 50 min at 42°C for the extension step, and finally 15 min at 70°C for the inactivation of the enzyme. The resulting cDNA was diluted to a final volume of 600 µl with nuclease-free water prior to use for qRT-PCR.

2.8.3. qRT-PCR

qRT-PCR master mix was prepared using Sensi Mix Sybr Low Rox Kit (Bioline, #QT625-20). The reaction was prepared according to Table 47.

Table 47. Reagents for qRT-PCR

Component	Volume
cDNA template	5 µl
Forward primer (5 µM)	0.6 µl
Reverse primer (5 µM)	0.6 µl
Sensi Mix (2x)	7.5 µl
Nuclease free water	1.3 µl

qRT-PCR reactions were performed using Mx3005P system (Stratagene) with cycling conditions as listed in Table 48. Samples were run in technical triplicates and normalised to *PBGD*. Gene-specific primers are listed in Table 49. Data represents the mean of one experiment which is representative of three independent experiments and error bars indicate standard deviation of triplicates.

Table 48. Parameters for qRT-PCR cycling conditions

Stage	Cycles	Temperature	Time
1	1	95°C	10 min
2	40	95°C	30 s
		60°C	30 s
		72°C	30 s
3	1	95°C	1 min
		55°C	30 s (↑ 1°C/30 s)
		95°C	30 s

Table 49. Primers used in qRT-PCR

Gene		**Primer sequence (5' → 3')**
OCT4	F	AGTGAGAGGCAACCTGGAGA
	R	ACACTCGGACCACATCCTTC
NANOG	F	CATGAGTGTGGATCCAGCTTG
	R	CCTGAATAAGCAGATCCATGG
SOX2	F	TGGACAGTTACGCGCACAT
	R	CGAGTAGGACATGCTGTAGGT
GSC	F	GAGGAGAAAGTGGAGGTCTGGTT
	R	CTCTGATGAGGACCGCTTCTG
BRACHURY	F	TGCTTCCCTGAGACCCAGTT
	R	GATCACTTCTTTCCTTTGCATCAAG
EOMESODERMIN	F	ATCATTACGAAACAGGGCAGGC
	R	CGGGGTTGGTATTTGTGTAAGG
GATA4	F	TCCCTCTTCCCTCCTCAAAT
	R	TCAGCGTGTAAAGGCATCTG
GATA6	F	TGTGCAATGCTTGTGGACTC
	R	AGTTGGAGTCATGGGAATGG
SOX17	F	CGCACGGAATTTGAACAGTA
	R	GGATCAGGGACCTGTCACAC
CXCR4	F	CACCGCATCTGGAGAACCA
	R	GCCCATTTCCTCGGTGTAGTT
FOXA2	F	GGGAGCGGTGAAGATGGA
	R	TCATGTTGCTCACGGAGGAGTA
GCG	F	AAGCATTTACTTTGTGGCTGGATT
	R	TGATCTGGATTTCTCCTCTGTGTCT
HLXB9	F	CACCGCGGGCATGATC
	R	ACTTCCCCAGGAGGTTCGA
HNF1B	F	TCACAGATACCAGCAGCATCAGT
	R	GGGCATCACCAGGCTTGTA
HNF4A	F	CATGGCCAAGATTGACAACCT
	R	TTCCCATATGTTCCTGCATCAG

HNF6	F	GTGTTGCCTCTATCCTTCCCAT
	R	CGCTCCGCTTAGCAGCAT
INSULIN	F	GAAGCGTGGCATTGTGGAAC
	R	GCTGCGTCTAGTTGCAGTAGT
NGN3	F	GCTCATCGCTCTCTATTCTTTTGC
	R	GGTTGAGGCGTCATCCTTTCT
NKX6.1	F	GGCCTGTACCCCTCATCAAG
	R	TCCGGAAAAAGTGGGTCTCG
PDX1	F	GATTGGCGTTGTTTGTGGCT
	R	GCCGGCTTCTCTAAACAGGT
SST	F	CCCCAGACTCCGTCAGTTTC
	R	TCCGTCTGGTTGGGTTCAG
PBGD	F	GGAGCCATGTCTGGTAACGG
	R	CCACGCGAATCACTCTCATCT
SOX9	-	Quantitect primers (QT00001498)

2.9. RNA-sequencing

2.9.1. Illumina sequencing

Library preparation and deep sequencing were performed at the Wellcome Trust Sanger Institute (Hinxton, UK). RNA-sequencing was run on Illumina Hiseq v3 with read length 75bp and paired-ends, and a library fragment size of 100-1000bp using a multi-plex strategy. Samples were run in biological triplicates.

2.9.2. RNA enrichment analysis

RNA-sequencing analyses were performed with partial help from Dr Pedro Madrigal. Tophat v2 (Kim et al., 2013), provided by Ensembl release 76 annotations, was used to align short reads from each sample to the reference human genome assembly (GRCh38/hg20). Feature counts was used to summarize paired-end reads and count fragments with parameters '-p -T 8 -t exon -g gene_id' (Liao et al., 2014). DESeq2 Bioconductor package was used to search for significant differences between samples, requiring at least a 2 fold expression change and adjusted p-value less than 0.01 (Love et al., 2014). R package edgeR function 'rpkm' was used with default parameters to normalize count gene expression (Robinson et al., 2010). Raw bedGraphs were normalized per million mapped reads in the library per library size in all ChIP-seq and RNA-seq samples. Genome browser panels were generated using IGV (Thorvaldsdóttir et al., 2013).

2.9.3. Functional annotations

Gene Ontology (GO) analyses were performed using Amigo2 separately for up- and down- regulated differentially expressed genes (Carbon et al., 2009). Spearman's correlation values were calculated for FPKM values for gene expressed at more than 5 FPKM in at least one sample.

2.10. Chromatin Immunoprecipitation (ChIP)

2.10.1. Cross-linking of protein and DNA

Co-binding of DNA to DNA-binding proteins was determined by ChIP against *GATA6* (Cell Signaling, #5851) on approximately 1 x 10^7 cells per antibody or control sample. hPSCs grown on 10 cm dishes were differentiated and harvested either at the endoderm (D3) or pancreatic progenitor (D12) stage. 1 x 10 cm dish of cells harvested at D3 was used for one antibody immunoprecipitation (IP). 1 x 10 cm dish of cells harvested at D12 was used for four antibody IPs. Cells were cross-linked by adding 312.5 µl of 16% formaldehyde (ThermoFisher UK, #11586711) to 5 ml of media to make a final concentration of 1%. The cells were incubated rocking for 10 min at room temperature to allow cross-linking of protein-DNA complexes.

The reaction was quenched by adding 312.5 µl of 2 M glycine (Millipore, #357002) to make a final concentration of 0.125 M and incubated for 5 min with rocking. Thereafter, the media was aspirated and cells were washed twice with 5 ml of ice-cold PBS. The cells were detached by scraping into 3 ml of ice-cold PBS containing freshly-added protease inhibitors (10 µl/ml of 5 mg/ml phenylmethylsulfonylfluoride (PMSF; Sigma-Aldrich, #93482), 10 µl/ml of 1 M Sodium Butyrate (Sigma-Aldrich, #303410) and 1 µl/ml of 1 mg/ml Leupeptin (Roche, #11017101001)) and pooled into 15 ml tubes, each tube containing approximately 2 x 10^7 cells, or 2 IPs. The cells were then centrifuged for 5 min at 1,200 rpm at 4°C to pellet.

For all subsequent steps, the samples were kept on ice. For all subsequent buffers used, the aforementioned protease inhibitors were added freshly to the buffers before use. The pelleted cells were subsequently re-suspended in 2 ml of ice-cold Cell Lysis Buffer (10 mM Tris-Cl pH 8.0, 10 mM NaCl and 0.2% NP-40) per 15 ml tube and incubated on ice for 10 min. The cells were then centrifuged for 5 min at 1,800 rpm at 4°C. The supernatant was discarded and the pellet was gently re-suspended in 1.25 ml of ice-cold Nuclear Lysis Buffer (50 mM Tris-Cl pH 8.0, 10 mM EDTA and 1% SDS) per 15 ml tube and incubated on ice for 10 min. 0.75 ml of ice-

cold IP Dilution Buffer (20 mM Tris-Cl pH 8.0, 2 mM EDTA, 150 mM NaCl, 0.01% SDS, 1% Triton X-100) was then added per 15 ml tube.

The chromatin was then transferred into 15 ml Diagenode sonication tubes containing sonication beads (Diagenode, #C01020031) that were pre-washed twice with 10 ml D-PBS each time and once with 10 ml IP Dilution Buffer. Next, chromatin was sonicated in Diagenode Biorupter Pico for 10 cycles of 30s on/45s off. The sonicated chromatin were transferred to 1.5 ml tubes and centrifuged at 14,000 rpm for 10 min at 4°C to pellet debris. The pellet was discarded, and two 15 ml tubes, or 4 IPs, worth of sonicated chromatin was pooled into a fresh 15 ml tube. 3.5 ml of IP Dilution Buffer was added and mixed gently. The cross-linked DNA was pre-cleared by incubating with rotation 10 µg of rabbit IgG (Sigma-Aldrich, #I5006) for 1 hr at 4°C, followed by incubating with rotation 100 µl of Protein G agarose beads (50% v/v; Roche, #11243233001) pre-washed twice with D-PBS for 1 hr at 4°C. The samples were then centrifuged for 3 min at 3,000 rpm at 4°C and the supernatant was transferred to a fresh 15 ml tube. An aliquot of 300 µl for Input sample was transferred to a fresh 1.5 ml Eppendorf tube and stored at 4°C.

2.10.2. Immunoprecipitation of protein-DNA complex

The supernatant was split equally into four 15 ml tubes, each representing one sample i.e. 1 x 10^7 cells worth of material per tube. 10 µg of *GATA6* antibody or rabbit IgG control was added per tube, and samples were incubated rotating overnight at 4°C. Antibody-bound chromatin was then collected using 60 µl of Protein G agarose beads (50% v/v) pre-washed twice with D-PBS for each tube by incubating with rotation for 1 hr at 4°C. Thereafter, the tubes were centrifuged for 3 min at 3,000 rpm at 4°C. The supernatant was discarded and the pellet containing the protein-DNA complexes bound onto the protein G agarose beads were kept.

2.10.3. DNA extraction

500 µl of IP Wash Buffer 1 (20 mM Tris-Cl pH 8.0, 2 mM EDTA, 50 mM NaCl, 0.1% SDS and 1% Triton X-100) was added to each tube and the tubes were vortexed. The samples were transferred to fresh 1.5 ml Eppendorf tubes and were then centrifuged at 14,000 rpm for 1 min at 4°C. The samples were washed with IP Wash Buffer 1 once more by removing the supernatant, adding 500 µl of IP Wash Buffer 1, vortexing, and centrifuging at 14,000 rpm for 1 min at 4°C. 500 µl of IP Wash Buffer 2 (10 mM Tris-Cl pH 8.0, 1 mM EDTA, 0.25 M LiCl, 1% NP-40 and 1% Sodium deoxycholic acid) was then added to each tube after removing the supernatant and the tubes were vortexed and centrifuged at 14,000 rpm for 1 min at 4°C. The wash was repeated once more. Thereafter, the samples were washed twice with 500 µl of TE Buffer (10mM Tris-Cl pH 8.0, 1mM EDTA) according to the procedures above. After the supernatant was aspirated, the chromatin was eluted from the Protein G beads by washing twice with 150 µl of Elution Buffer (100 mM $NaHCO_3$ and 1% SDS), vortexing and centrifuging at 14,000 rpm for 1 min at 4°C. The supernatants were collected and pooled in fresh 1.5 ml Eppendorf tubes.

ChIP and Input DNA cross-links were reversed and RNA degraded by adding 1 µl of 1 mg/ml RNase A and 18 µl of 5M NaCl and incubating at 67°C in a heat block with shaking at 1,300 rpm overnight. Protein was degraded by adding 3 µl of 20 mg/ml Proteinase K and incubating for 3 hrs at 45°C in a heat block with shaking at 1,300 rpm. DNA was extracted using 300 µl of phenol/chloroform wash with vortexing then centrifuged at 14,000 rpm for 5 min at room temperature. The aqueous layer containing pulled down genomic DNA was transferred to fresh 1.5 ml Eppendorf tubes. 30 µl of 3M NaAc pH 5.2 (Ambion, #AM9740), 30 µg glycoblue (Ambion, #AM9516) and 750 µl of 100% ethanol were added to the samples, which were then vortexed. The samples were next incubated for at least 30 min at -80°C to precipitate the DNA. Precipitated DNA was pelleted by centrifuging at 14,000 rpm for 30 min at 4°C. The DNA pellet was then washed with ice-cold 70% ethanol and centrifuged at 14,000 rpm for 5 min at 4°C. The ethanol was removed and the pellet

air dried. 70 µl of deionised water was added to Input samples whereas 30 µl of deionised water was added to ChIP samples.

2.10.4. Bioanalyser

Chromatin fragments after sonication were determined by a Bioanalyser (Agilent 2100 Bioanalyzer) to ensure that fragmented chromatins are within the range of 150-300 base pairs in size.

A 10 µl aliquot was taken directly from freshly sonicated samples and transferred to a clean 1.5 ml Eppendorf tube. The tubes were centrifuged at 14,000 rpm for 10 min at 4°C to remove any insoluable material. The supernatant was transferred to a fresh 1.5 ml Eppendorf tube. Thereafter, fast reversal of the cross-links was performed by adding 74 µl of nuclease free water, 4 µl of 5M NaCl, 8 µl of 20 mg/ml Proteinase K and 4 µl of 10 mg/ml RNase A to the aliquots. The aliquots were incubated at 65°C for 2 hr in a heat block with shaking at 1,300 rpm. DNA was purified using QIAquick PCR Purification Kit (Qiagen) and DNA was eluted in 20 µl of elution buffer. DNA was diluted 1:10 in nuclease free water and 1 µl was used for analysis using High Sensitivity DNA Kit (Agilent, #5067-4626) according to the manufacturer's protocol.

2.10.5. qPCR detection

Pulldown enrichment was validated by qPCR using KAPA SYBR FAST Master Mix (2X) ROX Low qPCR Kit (KAPA Biosystems). Gene-specific primers kindly provided by our collaborator, Jorge Ferrer's group, are listed in Table 50. Results were expressed as normalised values against a negative control region and fold change compared between antibody pulldown and IgG control.

Table 50. Primers used in ChIP qPCR

Primer name		**Primer sequence (5' → 3')**
hGATA6 positive region	F	CATGGAGACAGCAACAGTCC
hGATA6 positive region	R	ACCGCCCGGTTATCTTATTG
hPDX1 positive region	F	TTTCTCGCTGCCCTTTACTC
hPDX1 positive region	R	GTGCTGTGGCTCAACTCTGA
NR0B2 positive for hGATA6 and hPDX1	F	GCTGCCCCTTATCAGATGAC
NR0B2 positive for hGATA6 and hPDX1	R	CTGGCTTAGCAAAAGCCCTA
Amy2A negative control	F	TGCTGCCAGAACCTAAGAAAA
Amy2A negative control	R	TTGAGGGCAAACTGTTTATTCA
Nanog negative control	F	AAAGCTTGCCTTGCTTTGAA
Nanog negative control	R	AGTCTCCGTGTGAGGCATCT

2.11. ChIP-sequencing

2.11.1. Illumina sequencing

Library preparation and deep sequencing were performed at the Wellcome Trust Sanger Institute (Hinxton, UK). RNA-sequencing was run on Illumina Hiseq v4 with read length 75bp and paired-ends, and a library fragment size of 100-1000bp using a multi-plex strategy. Samples were run in biological duplicates.

2.11.2. Bioinformatics analyses

ChIP-sequencing analyses were performed with partial help from Dr. Pedro Madrigal and Dr. Denil Simon Lieven Imanuel Johannes. Short-insert paired-end reads were aligned to the reference human genome assembly (GRCh38/hg18) using the Burrows-Wheeler Aligner (BWA) 0.5.10 (Li and Durbin, 2009) with -q 15 and default for the rest of parameters. Reproducibility between replicates was first assessed using the Pearson Correlation Coefficient (PCC) for the two biological replicates, using the genome-wide normalized read (extended to 300 bp) count distribution on a single nucleotide resolution. For this, the UCSC tool bigwigCorrelate provided in http://hgdownload.cse.ucsc.edu/admin/exe/. PCC is equal to 0.949326 was used.

Peak calling was performed using MACS version 2.0.10 (Zhang et al., 2008), allowing a p-value cut-off of p-value ≤ 1e-3, and default for the rest of parameters. Relaxed thresholds are suggested in order to enable the correct computation of IDR values (Landt et al., 2012). Following the recommendations for the analysis of self-consistency and reproducibility between replicates, the negative control samples (IgG and input DNA) were combined into one single control; code for IDR analysis was downloaded from https://sites.google.com/site/anshulkundaje/projects/idr (Li et al., 2011). This is also beneficial as control samples with substantially higher number of reads are recommended for peak calling (Bailey et al., 2013).

To estimate the Irreproducible Discovery Rate (IDR) between replicates, top 35k peaks for each biological replicate were submitted for IDR analysis. For IDR computation using MACS results, we used p-values rather than q-values as suggested in https://sites.google.com/site/anshulkundaje/projects/idr (Li et al., 2011). The number of peaks found passing a threshold of IDR ≤ 5% (12,107) was selected as a conservative estimated number of candidate transcription factor binding sites. Autosomal and sex chromosomes were also excluded.

Co-localization plots of the transcription factors *GATA6*, *EOMES* and *SMAD2/3* ChIP-seq, was generated with deepTools (Ramirez et al., 2014). The input data was obtained by combining my ChIP data of H9 cells at day 3 (*GATA6*) with previously published *EOMES* (Teo et al., 2011) and *SMAD2/3* ChIP data (Brown et al., 2011). To make the results more comparable, the 3 data sets were re-mapped with STAR v2.5.1a (Dobin et al., 2013) (BWA failed on short single end SMAD reads) and processed with MACS version 2.0.10 and IDR as described earlier. The resulting peak files (bed format) were used as input for deepTools. The mapped read files (bam format) were pre-processed with deepTools' "bamCompare" function (bin size = 50, assumed genome size = 2451960000 bp, ignoring chromosomes X and Y for normalization and extending single end reads by 250bp).

2.12. Statistical analyses

Unpaired two-tailed Student's t-test were used to assess statistical significance. Statistical analysis was done by GraphPad Prism software (Version 6.0 for Windows. GraphPad Software, San Diego, CA). p value < 0.05 was considered as statistically significant, *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$. Data are presented as mean ± s.d. as indicated in the figure legends. All sample numbers listed indicate the number of biological replicates employed in each experiment. For experiments showing data of one experiment that is representative of three independent experiments, this was done due to variations between different differentiation experiments; combining the experiments resulted in inaccurately large standard deviations that were not representative of the results.

CHAPTER 3 RESULTS

3.1. An *in vitro* culture system to study the development of the pancreas

3.1.1. An efficient protocol to differentiate hPSCs into endocrine progenitors

To study human pancreas development in an *in vitro* system, a chemically defined and feeder-free pancreatic specification protocol that was previously established in the lab (refer to Chapters 2.1.3 and 2.1.4) was used to differentiate hPSCs into endocrine progenitors. To confirm the efficiency of the protocol in generating a near homogenous population of pancreatic progenitors from hPSCs, I performed comprehensive analyses throughout the differentiation protocol; cells were analysed at key stages of the pancreas development and assessed for the expression of key markers to ensure successful commitment to the pancreatic lineage (Figure 10).

I first performed pancreatic differentiation on H9, a well-characterised hESC line that was routinely used in the lab and has been extensively tested for successful pancreatic differentiation. Cells were assayed by qRT-PCR and immunocytochemistry (ICC) on days 3 (DE), 6 (primitive gut tube), 9 (posterior foregut), 12 (pancreatic endoderm), 15, 18 and 24 (endocrine progenitor) of pancreatic differentiation, while FACS was performed on days 3, 12, 15 and 24 of pancreatic differentiation.

On day 0 of differentiation, the high expression levels of pluripotency factors (OCT4, SOX2 and NANOG) as shown in ICC and the absence of early germ layer markers (EOMES and SOX17) (Figure 11) indicated that the H9 cells were indeed in a fully undifferentiated and pluripotent state.

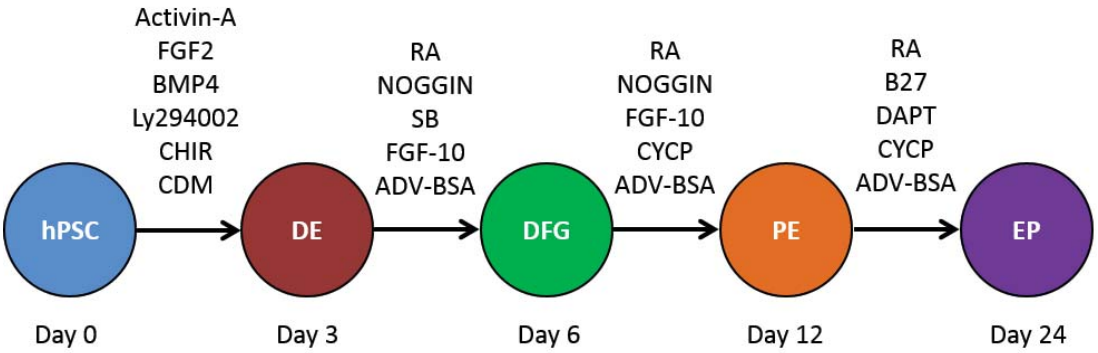


Figure 10. Schematic of the 24-day differentiation protocol. DE, definitive endoderm; DFG, dorsal foregut; PE, pancreatic endoderm; EP, endocrine progenitors. The culture medium and supplements indicated are BMP, bone morphogenetic protein 4; the PI3 kinase inhibitor Ly294002; the GSK3 inhibitor CHIR, CHIR99021; CDM, chemically defined medium; Adv-BSA, Advanced Dulbecco's Modified Eagle Medium/Ham's F-12 medium supplemented with BSA and L-glutamine; RA, retinoic acid; the ALK4/5/78 inhibitor SB-431542 (SB); FGF2, Fibroblast growth factor 2; FGF10, Fibroblast growth factor 10; the Hedgehog inhibitor, CYCP, Cyclopamine-KAAD; B27 supplement; the NOTCH inhibitor, DAPT.

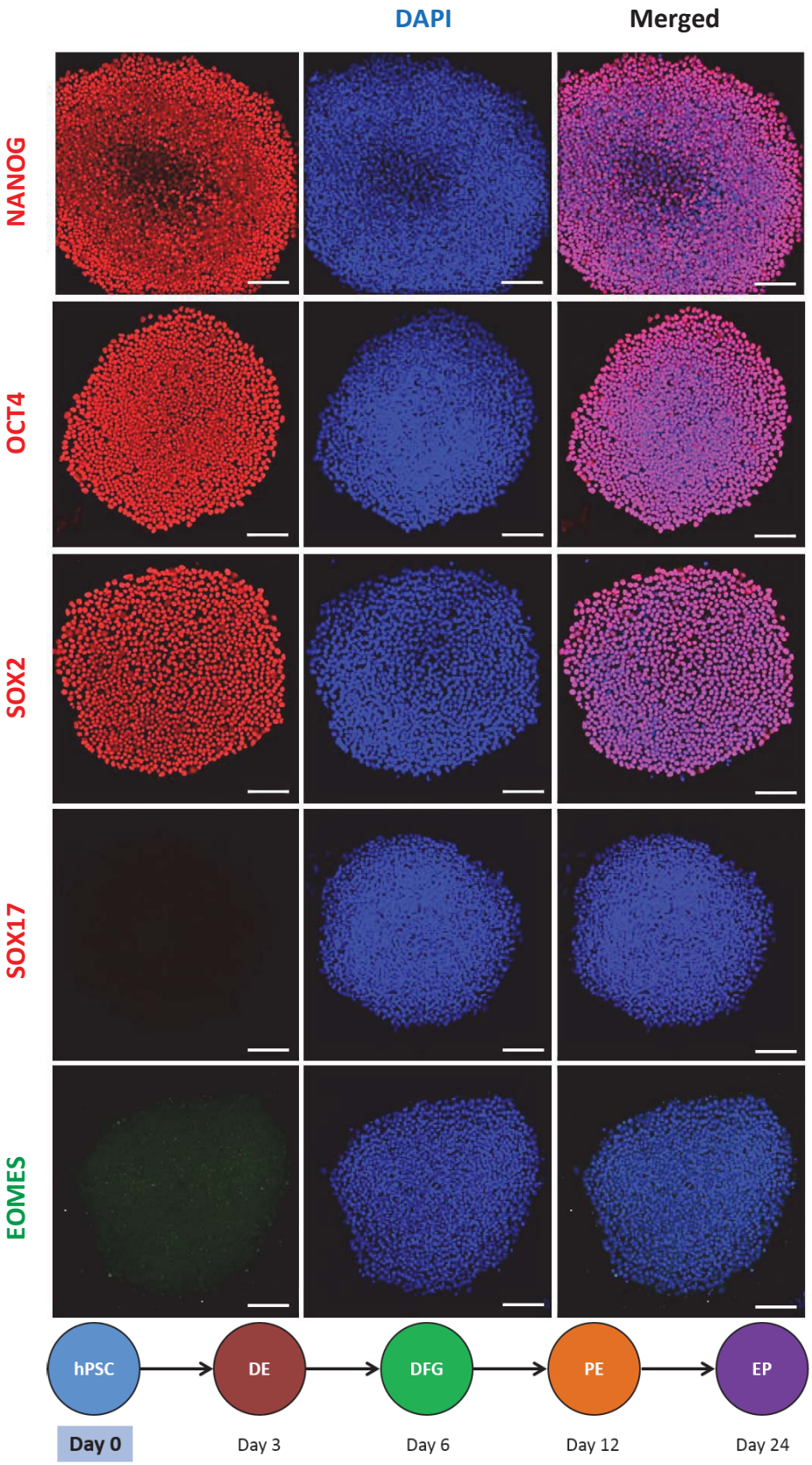


Figure 11. H9 cells are pluripotent and undifferentiated. Cells were grown in feeder-free culture conditions that maintain pluripotency and analysed via immunofluorescence. Cells were fixed on day 0 and were stained for pluripotency markers NANOG, OCT4 and SOX2 and early germ layer markers SOX17 and EOMES. Scale bar, 100 µm.

ICC analyses of DE differentiation after day 3 showed the up-regulation of endoderm marker SOX17 and early foregut marker FOXA2 accompanied by the down-regulation of pluripotency marker NANOG (Figure 12), indicating that the H9 cells successfully differentiated into the DE. FACS analyses revealed efficient DE formation with populations of 76.6% of SOX17+ cells and 67.7% of CXCR4+ cells (Figure 13). In addition, morphological analyses showed that the cells lost the compact colony morphology characteristic of stem cells and adapted a spread out "cobblestone-like" morphology indicating an epithelial-mesenchymal transition (EMT)-like phenomenon characteristic of endoderm cells (Figure 14).

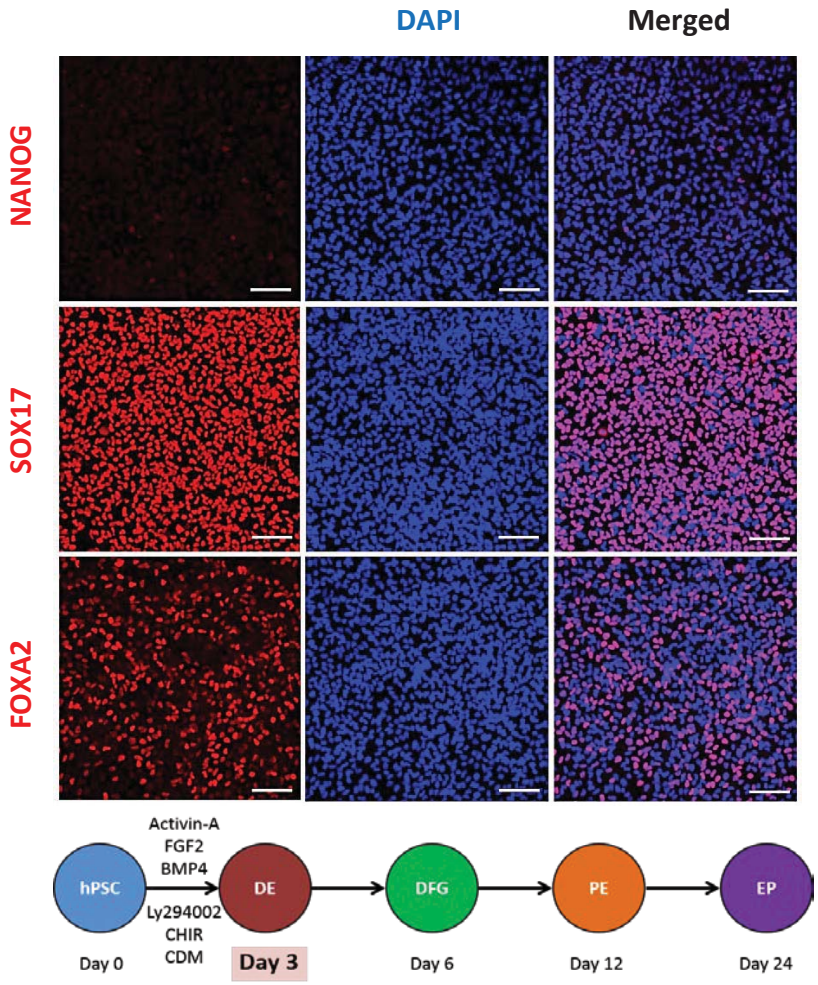


Figure 12. ICC of H9 cells differentiated into the DE. Cells were grown in culture conditions that specified them toward the DE lineage and analysed via immunofluorescence. Cells were fixed on day 3 and were stained for the pluripotency marker NANOG, as well as DE markers SOX17 and FOXA2. Scale bar, 100 µm.

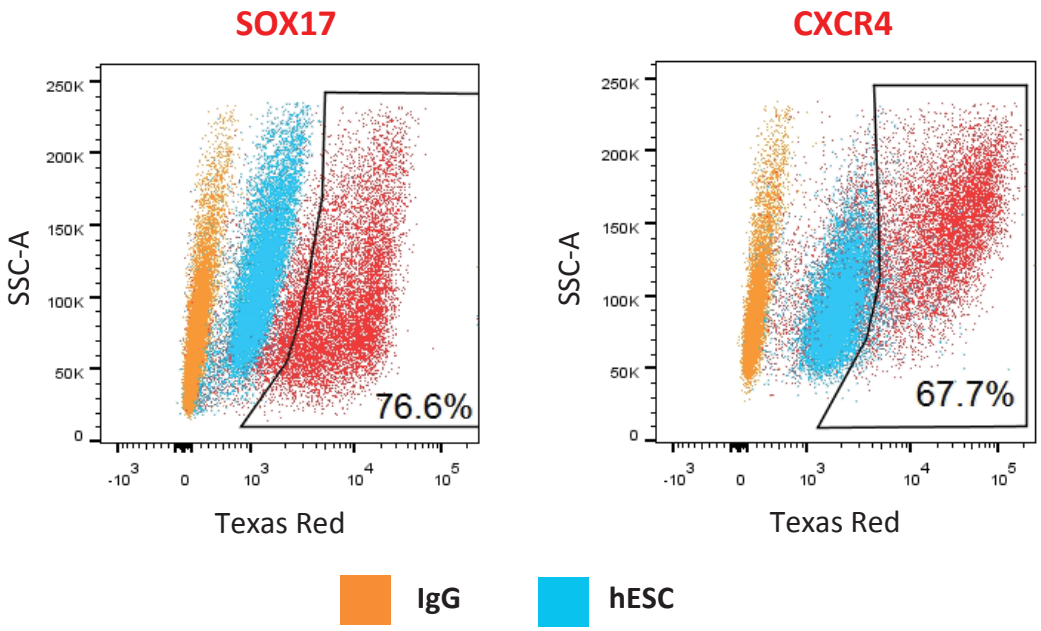


Figure 13. FACS of H9 cells differentiated into the DE. Cells were fixed on day 3 and were stained for DE markers SOX17 and CXCR4. Data show results of one experiment that is representative of at least 3 independent experiments.

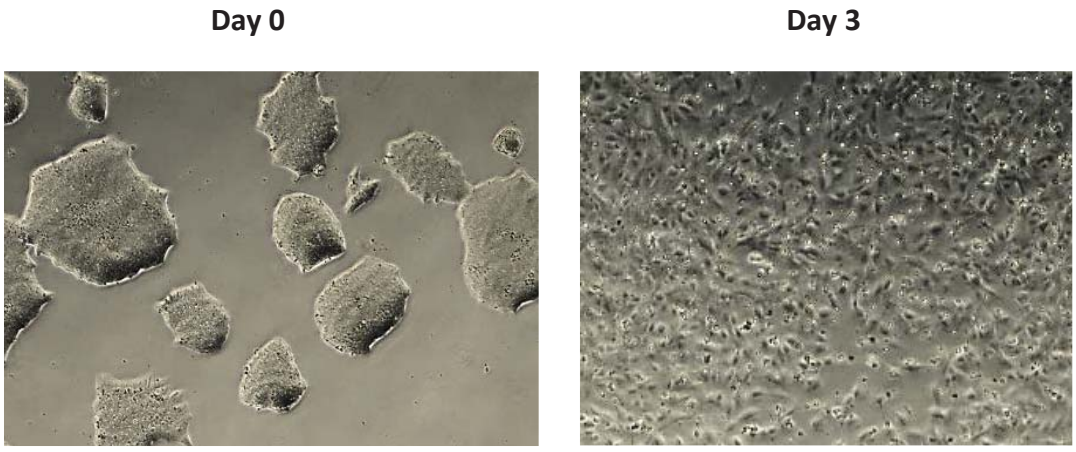


Figure 14. H9 exhibiting morphological changes upon differentiation into the DE. Bright field microscopy images showing the different morphologies between undifferentiated and DE cells.

Further differentiation toward the primitive gut tube on day 6 led to a near homogenous population of cells expressing key foregut markers such as FOXA2, HNF1B and HNF4A, as shown by ICC (Figure 15). Notably, the absence of HEX and CDX2 suggested a dorsal identity for these foregut cells (Figure 15). The cells continued to proliferate, resulting in a denser monolayer as the differentiation process progressed.

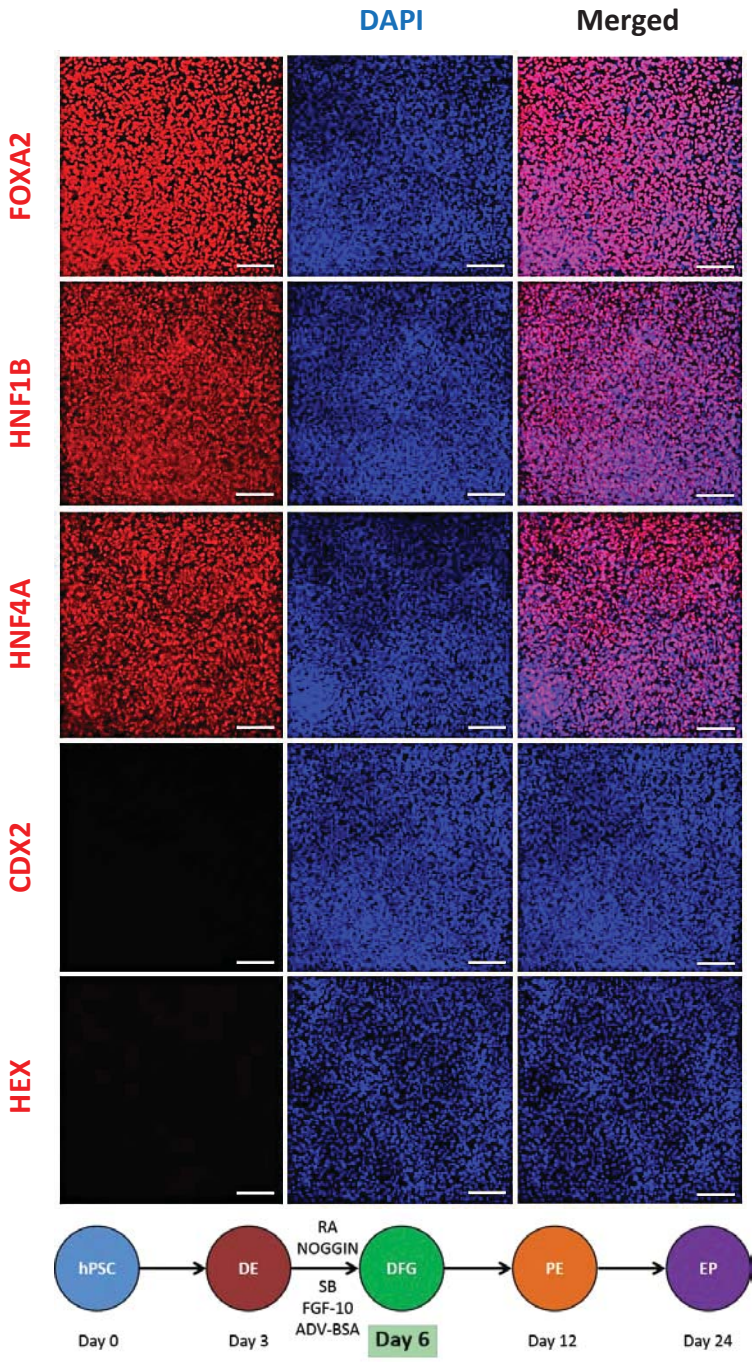


Figure 15. ICC of H9 cells differentiated into the primitive gut tube. Cells were grown in culture conditions that specified them toward the dorsal foregut and analysed via immunofluorescence. Cells were fixed on day 6 and were stained for key markers FOXA2, HNF1B and HNF4A and negative markers CDX2 and HEX. Scale bar, 100 µm.

Three days later on day 9, the cells continued expressing key foregut markers FOXA2, HNF1B, SOX2 and HNF6 (Figure 16). The key pancreatic marker PDX1 was shown to be almost homogeneously expressed by the end of day 12, indicating acquisition of pancreatic fate (Figure 17), and FACS analyses confirmed this by showing a 75.9% population of PDX1+ cells (Figure 18).

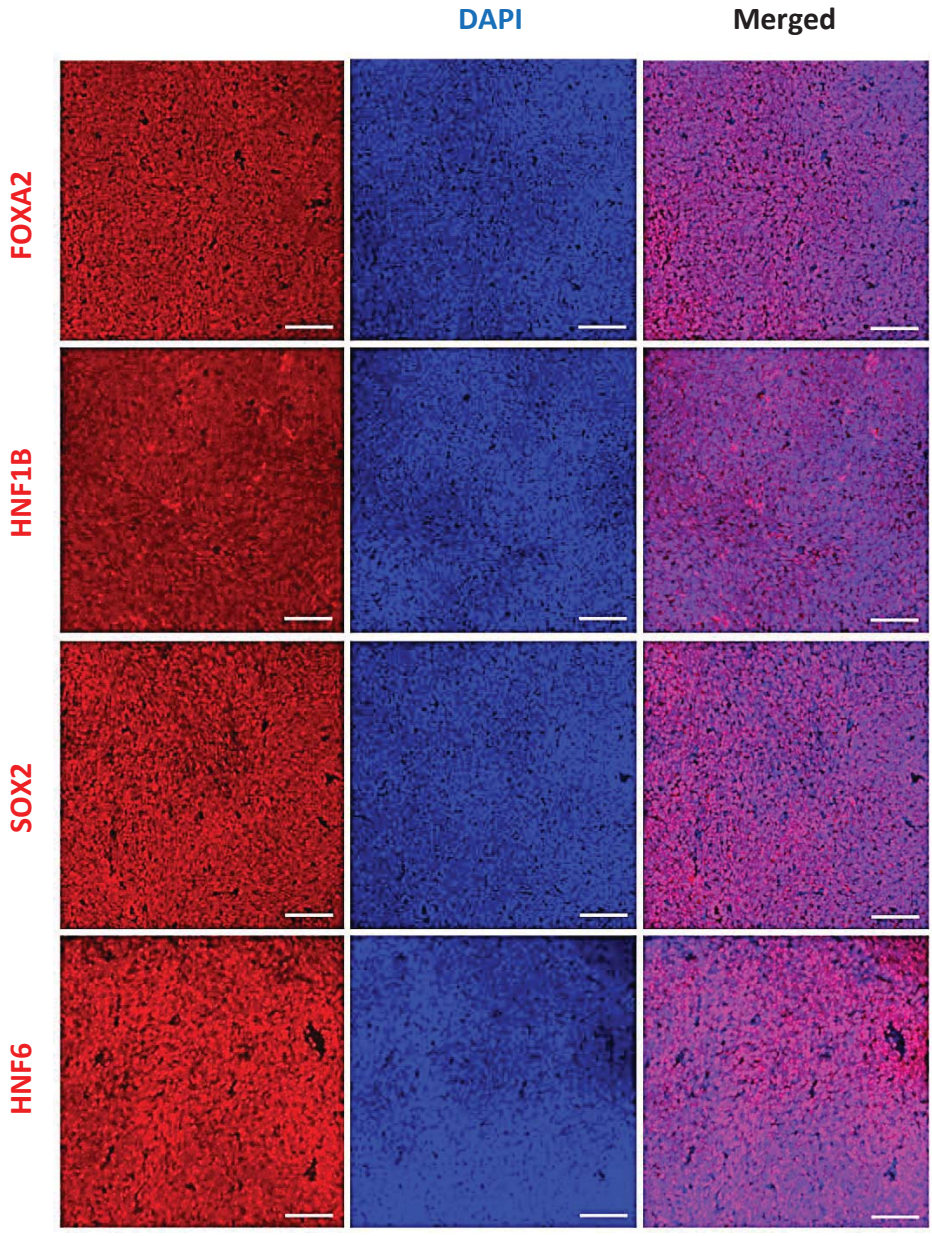


Figure 16. ICC of H9 cells differentiated into the posterior foregut. Cells were grown in culture conditions that specified them toward the foregut lineage and analysed via immunofluorescence. Cells were fixed on day 9 and were stained for key markers FOXA2, HNF1B, SOX2 and HNF6. Scale bar, 100 µm.

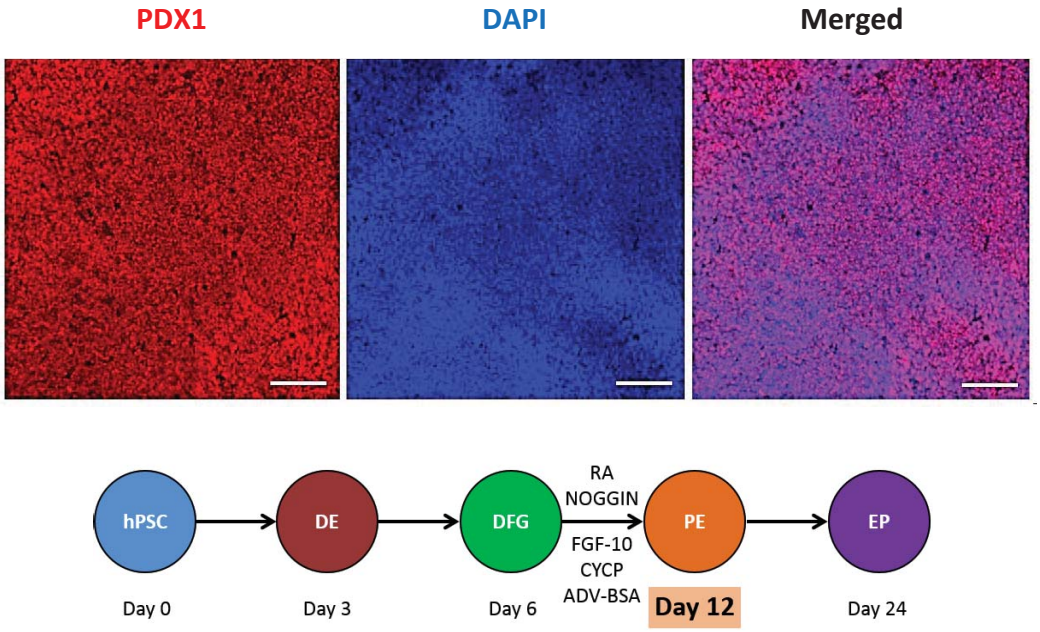


Figure 17. ICC of H9 cells differentiated into the pancreatic endoderm. Cells were grown in culture conditions that specified them toward the pancreatic endoderm lineage and analysed via immunofluorescence. Cells were fixed on day 12 and were stained for the key pancreatic marker PDX1. Scale bar, 100 µm.

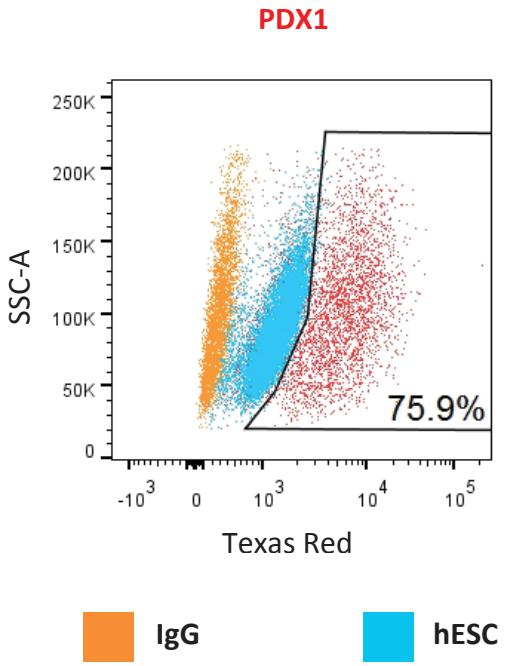


Figure 18. FACS of H9 cells differentiated into the pancreatic endoderm. Cells were fixed on day 12 and were stained for the key pancreatic marker, PDX1. Data show results of one experiment that is representative of at least 3 independent experiments.

On day 15, NGN3 was shown to be expressed by ICC (Figure 19) and FACS, at a population of 12.1% (Figure 20). Endocrine progenitors on day 18 revealed an expression of the key pancreatic marker PDX1 and endocrine progenitor markers C-PEPTIDE, SST and GCG via ICC (Figure 21). From day 18 to day 24 as the cells matured into immature β-cells, ICC showed an increased number of cells expressing C-PEPTIDE, SST and GCG (Figure 22). FACS analyses revealed populations of 10.4% C-PEPTIDE+ cells, 1.76% SST+ cells, and 3.79% GCG+ cells (Figure 23) and 7.9% of mono-hormonal C-PEPTIDE+ cells on day 24 (Figure 24).

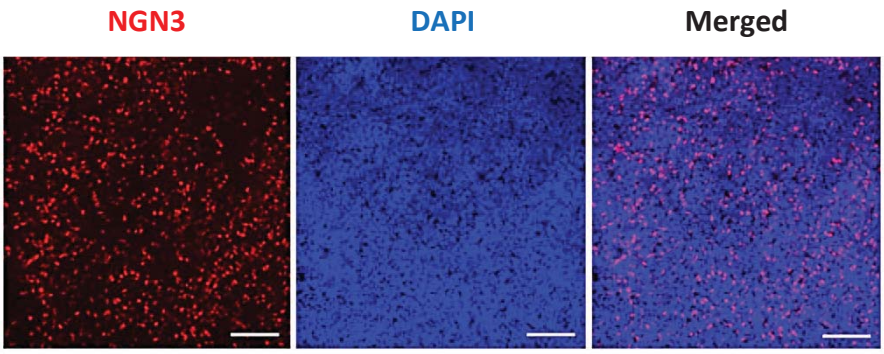


Figure 19. ICC of H9 cells on day 15. Cells were grown in culture conditions that allowed them to mature into endocrine progenitors and analysed via immunofluorescence. Cells were fixed on day 15 and were stained for NGN3. Scale bar, 100 µm.

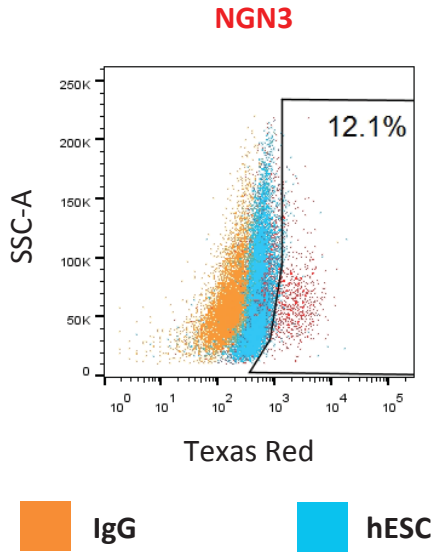


Figure 20. FACS of H9 cells on day 15. Cells were fixed on day 15 and were stained for NGN3. Data show results of one experiment that is representative of 3 independent experiments.

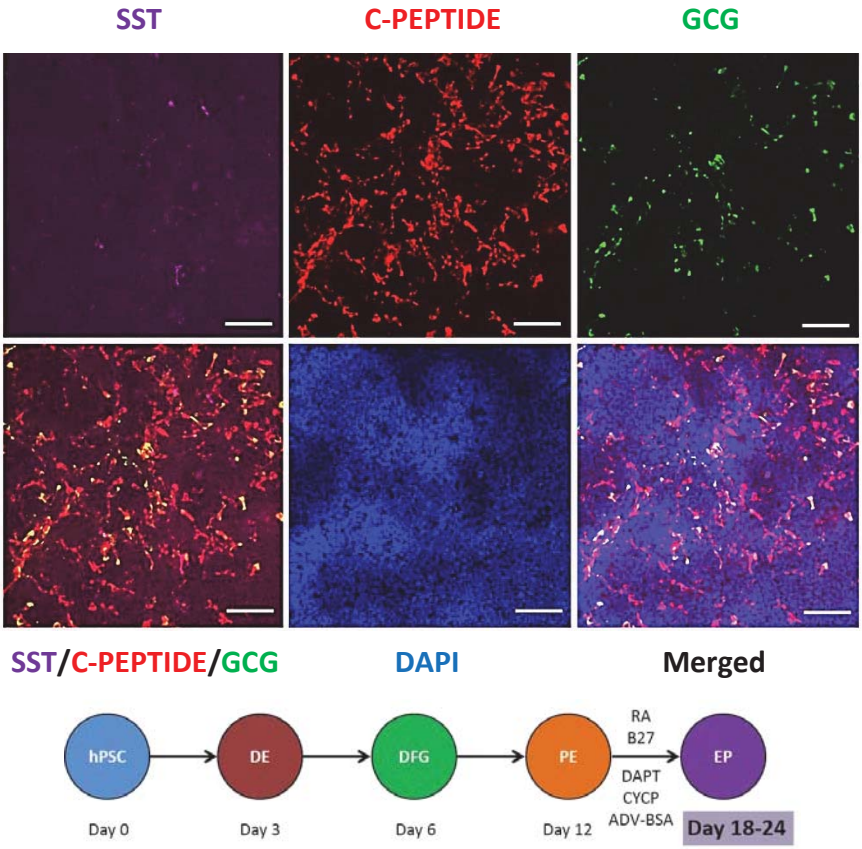


Figure 21. ICC of H9 cells on day 18. Cells were grown in culture conditions that allowed them to mature into endocrine progenitors and analysed via immunofluorescence. Cells were fixed on day 18 and were stained for key markers SST, C-PEPTIDE and GCG. Scale bar, 100 µm.

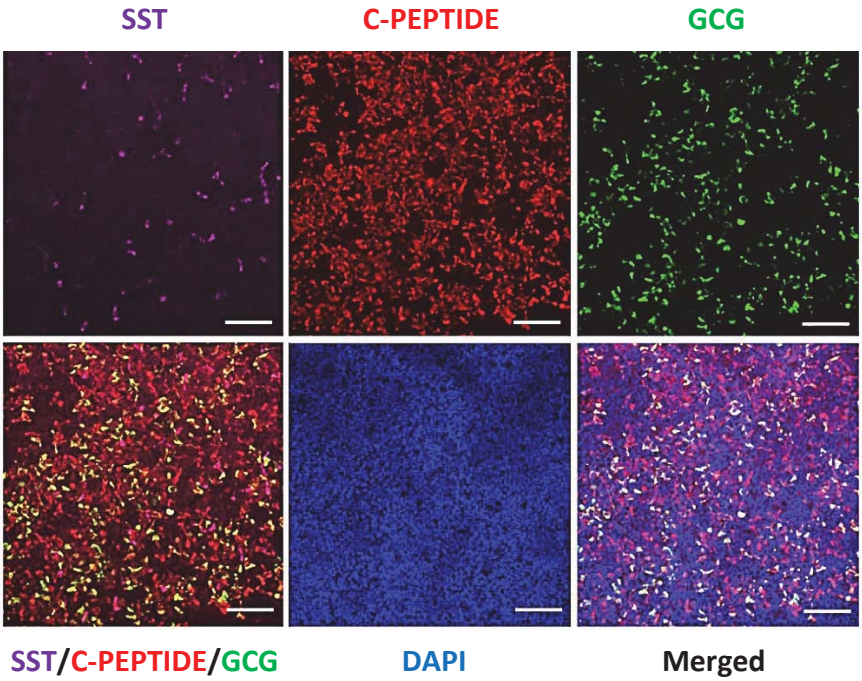


Figure 22. ICC of H9 cells differentiated into endocrine progenitors. Cells were grown in culture conditions that that allowed them to mature into endocrine progenitors and analysed via immunofluorescence. Cells were fixed on day 24 and were stained for key markers SST, C-PEPTIDE and GCG. Scale bar, 100 µm.

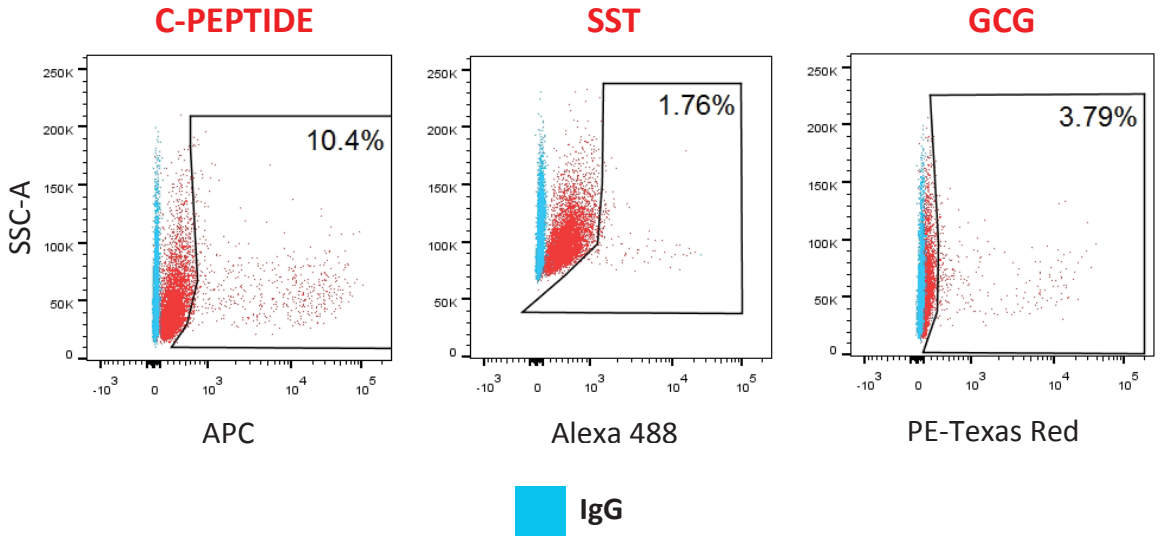


Figure 23. FACS of H9 cells differentiated into endocrine progenitors. Cells were fixed on day 24 and triple-stained for key markers, C-PEPTIDE, SST and GCG. Cell populations represent a combination of both poly- and mono-hormonal cells. Data show results of one experiment that is representative of 3 independent experiments.

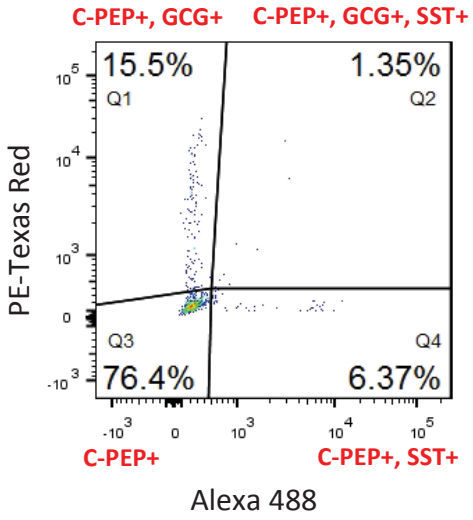


Figure 24. FACS of mono-hormonal H9 cells on day 24. Cell population consists of poly- or mono-hormonal cells which are C-PEPTIDE+ (C-PEPTIDE population of 10.4% from Figure 23). Q1 represents C-PEPTIDE+ and GCG+ poly-hormonal cells; Q2 represents C-PEPTIDE+, SST+ and GCG+ poly-hormonal cells; Q3 represents C-PEPTIDE+ mono-hormonal cells; Q4 represents C-PEPTIDE+ and SST+ poly-hormonal cells. Data show results of one experiment that is representative of 3 independent experiments.

Further qRT-PCR analyses not only validated the expression of key markers described earlier at each time point, but also provided an overview of the expression patterns of each marker over the course of pancreatic specification (Figure 25).

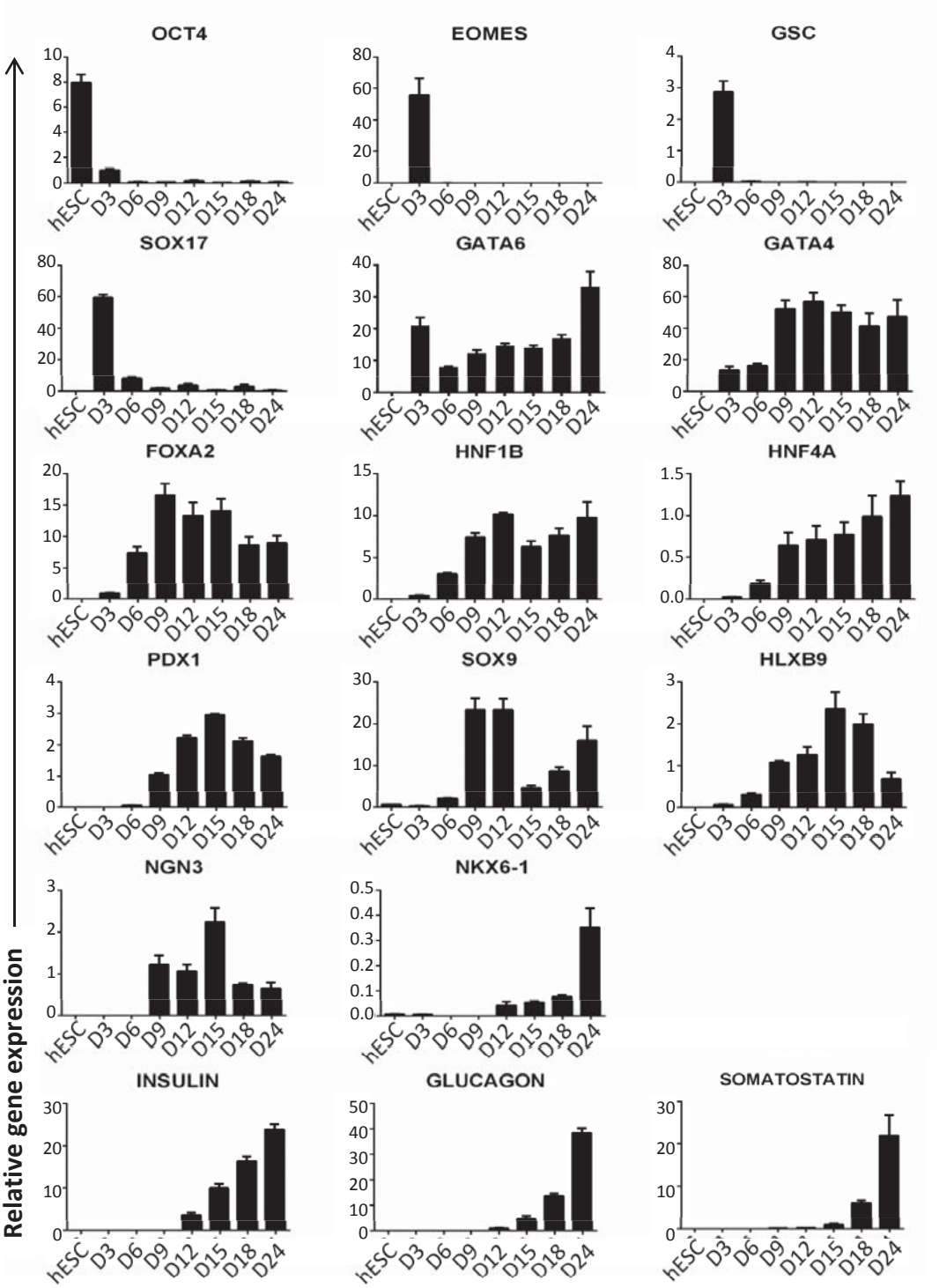


Figure 25. qRT-PCR analyses of H9 cells specified toward the pancreatic lineage. RNA was extracted at specific stages and the expression patterns of key markers were determined. Data are triplicate samples of one experiment and representative of three independent experiments. Error bars indicate standard deviation. Values are relative to the housekeeping gene *PBGD*.

To investigate the functional potential of the endocrine progenitors generated from the H9 hESCs, I performed C-peptide ELISA on cells that have undergone the 24-day differentiation protocol derived from the lab. C-peptide is often used as an alternative measurement for insulin at protein level, including ELISA, ICC and FACS. A major reason for this is because cells take up insulin present in the cell culture media, which may lead to false positive signals of insulin expression and contents in cells. The proteolytic cleavage of proinsulin prior to secretion produces the mature insulin molecule and the connecting peptide (C-peptide). As C-peptide is secreted in equimolar quantities to insulin, it serves as a substitute for the measurement of insulin. The 10% of H9-derived insulin-expressing cells seemed to elicit an inverse C-peptide releasing response upon glucose stimulation (Figure 26).

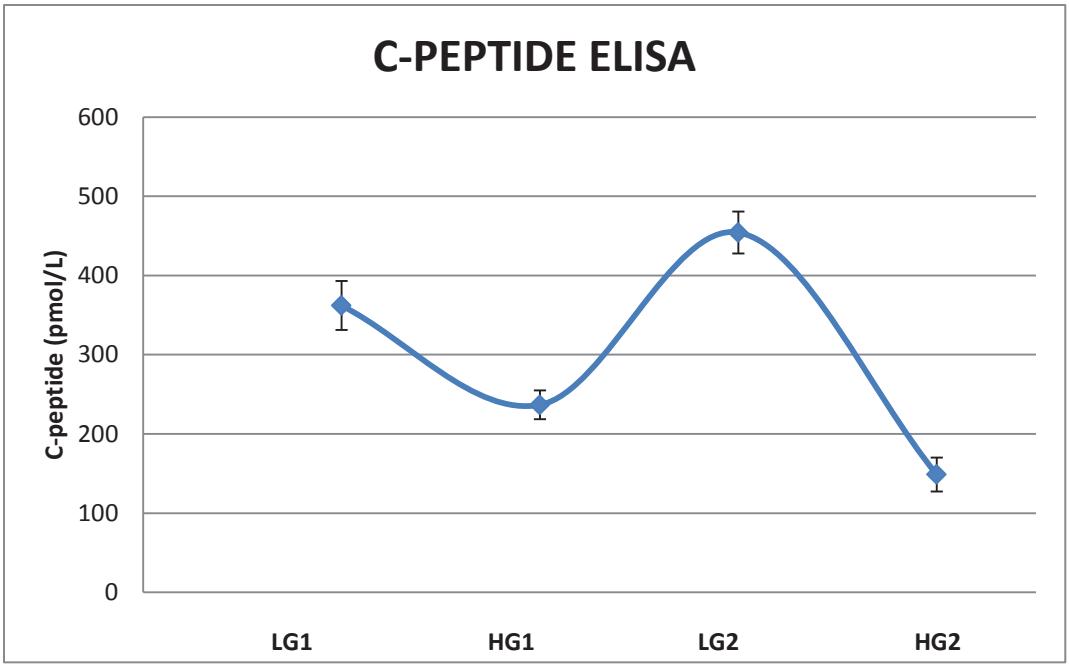


Figure 26. C-peptide secretion upon glucose stimulation on day 24. Cells were differentiated using the lab-derived protocol and assayed on day 24. Data are presented as the average of 3 biological replicates of one experiment and representative of three independent experiments. Error bars indicate standard deviation. LG1 is first incubation of low glucose; HG1 is first incubation of high glucose; LG2 is second incubation of low glucose; HG2 is second incubation of high glucose.

To enhance the reproducibility of the results, a second hPSC line was introduced into the project. This was a hIPSC line derived previously in the lab, named FSPS13.B. Pancreatic differentiation was performed on these cells and qRT-PCR and FACS analyses were done at various time points of the protocol. Expression patterns of the key genes analysed via qRT-PCR corroborated those from H9 cells (Figure 27). FACS analyses revealed 70.8% of SOX17+ cells and 83.7% of CXCR4+ cells on day 3 (Figure 28), and 86.8% of PDX1+ cells on day 12 (Figure 29), indicating similar differentiation efficiencies toward the DE and pancreatic endoderm lineages between H9 and FSPS13.B cells.

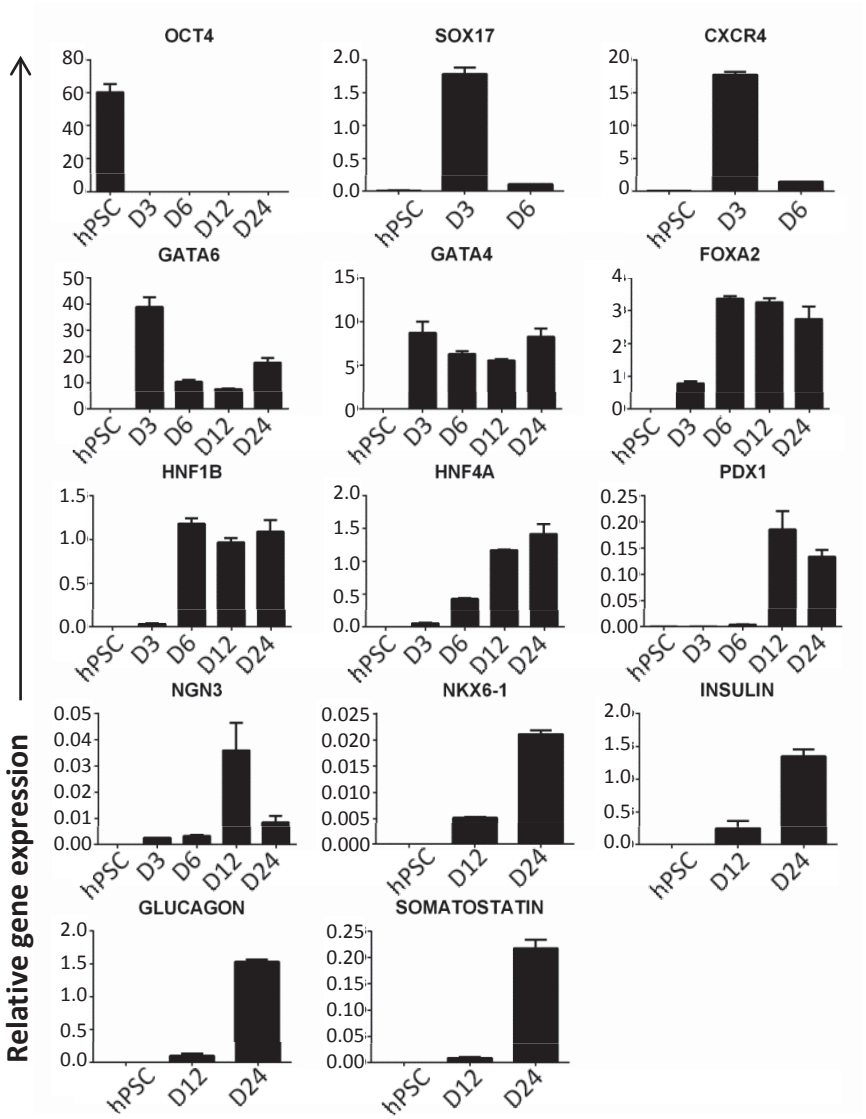


Figure 27. qRT-PCR analyses of FSPS13.B cells specified toward the pancreatic lineage. RNA was extracted at specific stages and the expression patterns of key markers were determined. Data are triplicate samples of one experiment and representative of three independent experiments. Error bars indicate standard deviation.

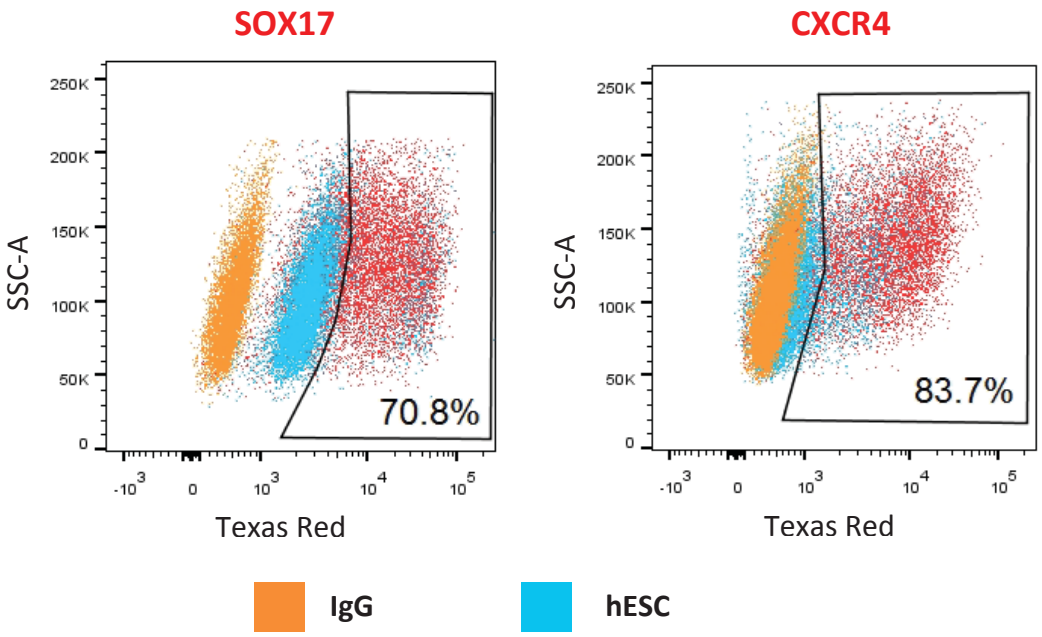


Figure 28. FACS of FSPS13.B cells differentiated into the DE. Cells were grown in culture conditions that specified them toward the DE lineage and then analysed by FACS. Cells were fixed on day 3 and were stained for the DE markers SOX17 and CXCR4. Data show results of one experiment that is representative of at least 3 independent experiments.

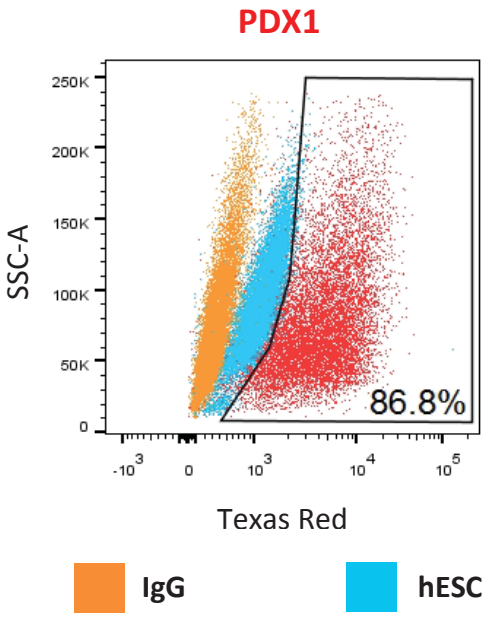


Figure 29. FACS of FPS13.B cells differentiated into the pancreatic endoderm. Cells were grown in culture conditions that specified them toward the pancreatic endoderm lineage and then analysed by FACS. Cells were fixed on day 12 and were stained for the pancreatic marker, PDX1. Data show results of one experiment that is representative of at least 3 independent experiments.

On day 24, however, FSPS13.B displayed lower endocrine progenitor differentiation efficiency as compared with H9, with FACS analysis showing 5.75% of C-PEPTIDE+ cells (Figure 30). Due to the low number of C-PEPTIDE+ cells, I could not perform triple staining of C-PEPTIDE, SST and GCG to determine the percentage of mono-hormonal and poly-hormonal cells as was done on H9 cells (Figure 24)

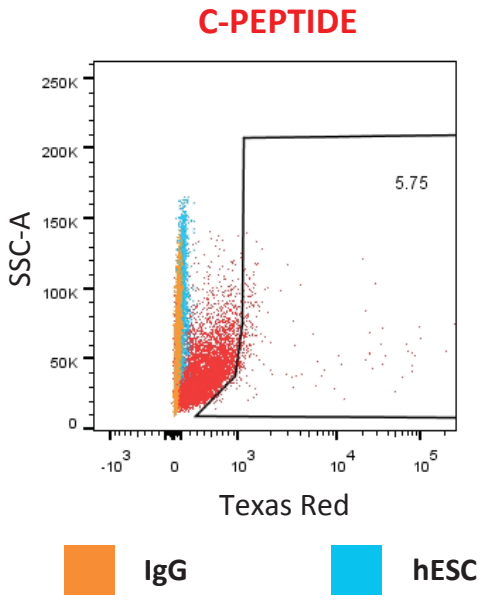


Figure 30. FACS of FPS13.B cells differentiated into the endocrine progenitors. Cells were grown in culture conditions that specified them toward the endocrine progenitor lineage and then analysed by FACS. Cells were fixed on day 24 and were stained for the endocrine marker, C-PEPTIDE. Data show results of one experiment that is representative of at least 3 independent experiments.

Together, except for the C-peptide ELISA, these results recapitulate pancreatic differentiation that was previously published (Cho et al., 2012), indicating to a large extent, the consistency of the pancreatic differentiation protocol in performing pancreatic differentiation on both hESC and hiPSC lines.

3.1.2. GATA6 is up-regulated upon definitive endoderm formation

To better understand of the role of GATA6 in the development of the human pancreas, I closely investigated the expression profile of GATA6 over the course of pancreatic specification. Using H9 cells, *in vitro* pancreatic differentiation was performed according to the protocol described in Chapters 2.1.3 and 2.1.4.

On day 0 of differentiation, GATA6 expression was demonstrated to be absent in undifferentiated and pluripotent cells, as shown by ICC (Figure 31) and qRT-PCR (Figure 33).

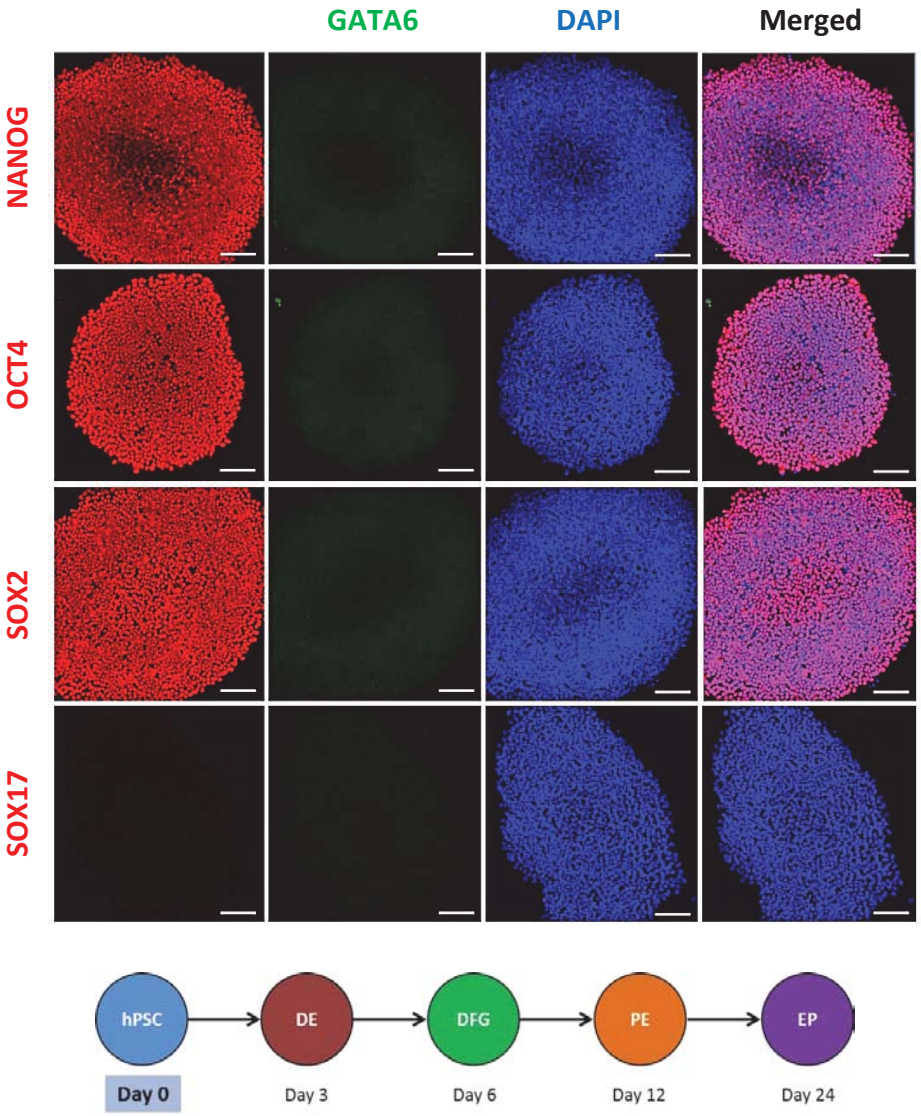


Figure 31. GATA6 expression is negligible in undifferentiated state. Cells were grown in culture conditions that maintained pluripotency and analysed via immunofluorescence. Cells were fixed on day 0 and were stained for pluripotency markers NANOG, OCT4 and SOX2 and early DE marker SOX17. Scale bar, 100 µm.

On day 3, ICC showed that GATA6 is highly expressed in the DE and is co-localised with key DE markers (Figure 32). The rapid up-regulation of *GATA6* on day 3 was confirmed by qRT-PCR (Figure 33). FACS analysis showed that 97.8% of the cells on day 3 were GATA6+ (Figure 34).

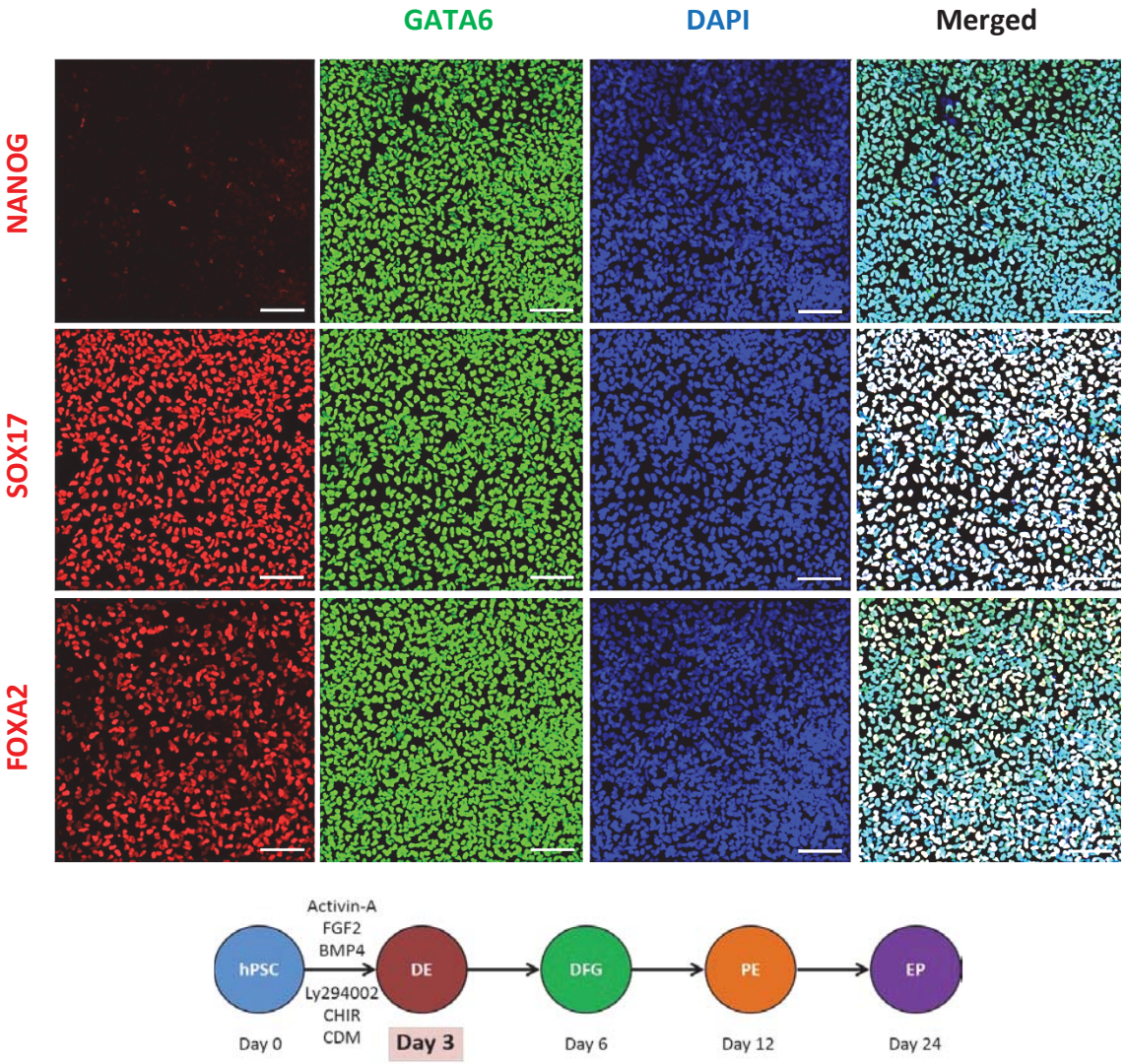


Figure 32. GATA6 is co-expressed with key DE markers. Cells were grown in culture conditions that specified them toward the DE lineage and analysed via immunofluorescence. Cells were fixed on day 3 and were stained for the pluripotency marker NANOG, as well as DE markers SOX17 and FOXA2. Scale bar, 100 µm.

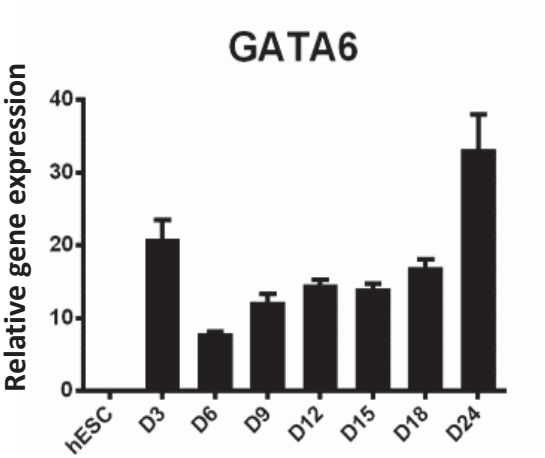


Figure 33. qRT-PCR analyses of *GATA6* expression levels in of H9 cells specified toward the pancreatic lineage. RNA was extracted at specific stages and the expression patterns of key markers were determined. Data are triplicate samples of one experiment and representative of three independent experiments. Error bars indicate standard deviation. Values are relative to the housekeeping gene *PBGD*.

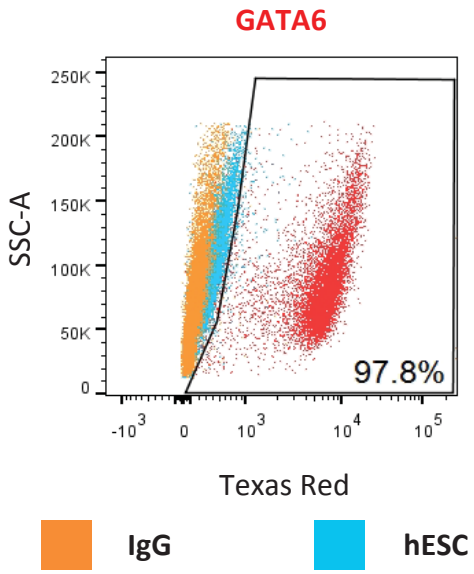


Figure 34. FACS of GATA6+ H9 cells at day 3. Cells were grown in culture conditions that specified them toward the DE lineage and then analysed by FACS. Cells were fixed on day 3 and were stained for GATA6. Data show results of one experiment that is representative of at least 3 independent experiments.

3.1.3. GATA6 is expressed throughout pancreatic development

On day 6, GATA6 continues to be expressed during pancreatic differentiation, although at lower levels compared with day 3 as indicated by the decreased staining intensity by ICC (Figure 35) and decreased relative expression level by qRT-PCR (Figure 33). On days 9 and 12, GATA6 remains expressed and co-localised with the key markers at each respective stage (Figure 36 and Figure 37). FACS analysis showed that 88.1% of cells on day 12 were GATA6+ (Figure 38).

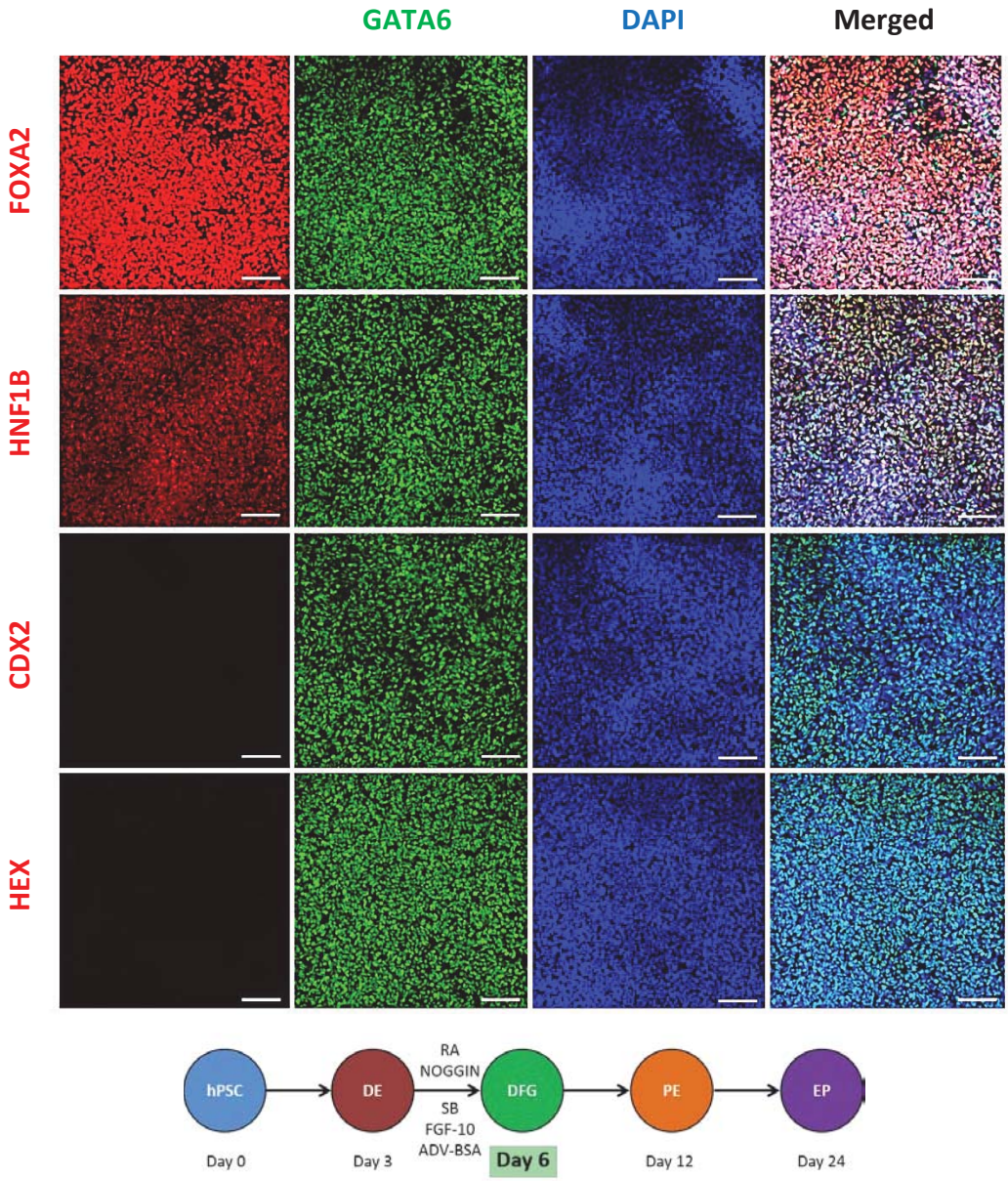


Figure 35. GATA6 is co-localised with key markers of the primitive gut tube. Cells were grown in culture conditions that specified them toward the dorsal foregut and analysed via immunofluorescence. Cells were fixed on day 6 and stained for key markers FOXA2 and HNF1B and negative markers CDX2 and HEX. Scale bar, 100 µm.

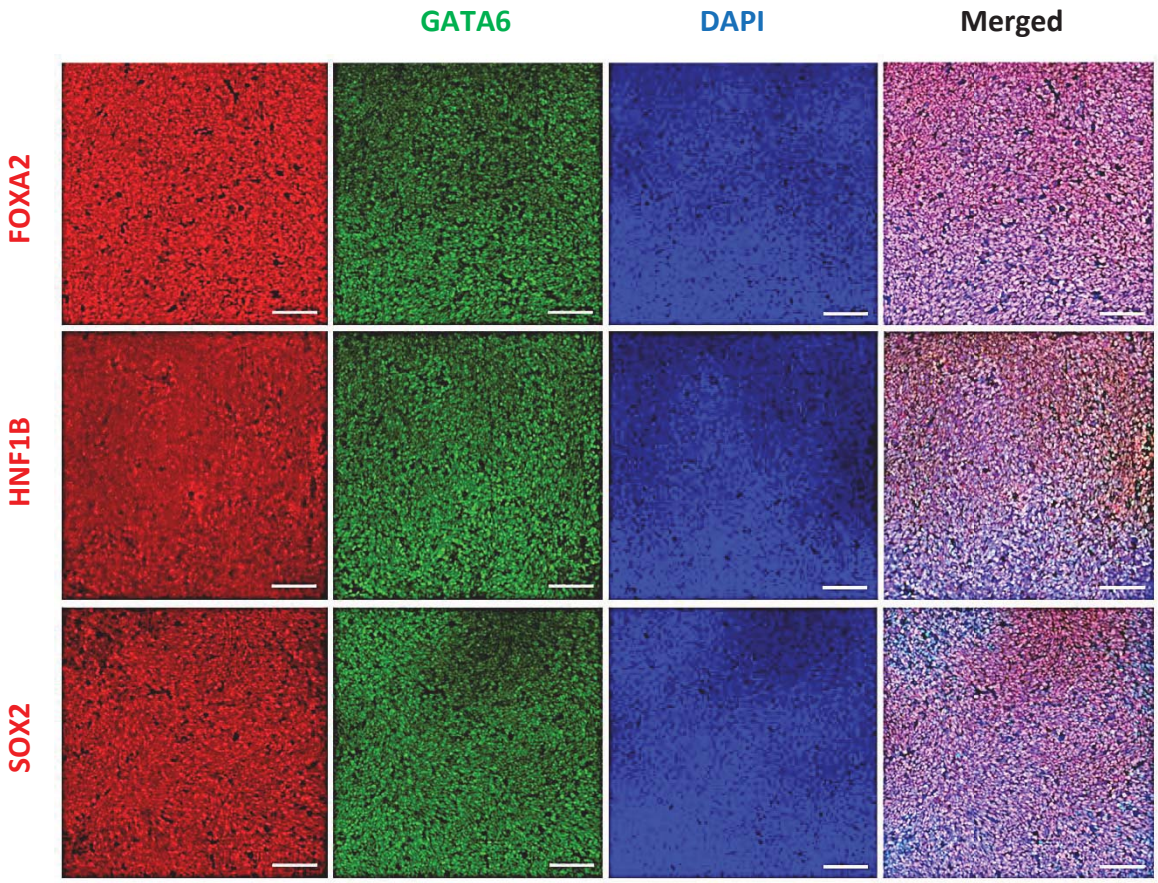


Figure 36. GATA6 is co-localised with key markers of the posterior foregut. Cells were grown in culture conditions that specified them toward the foregut lineage and analysed via immunofluorescence. Cells were fixed on day 9 and were stained for key markers FOXA2, HNF1B and SOX2. Scale bar, 100 µm.

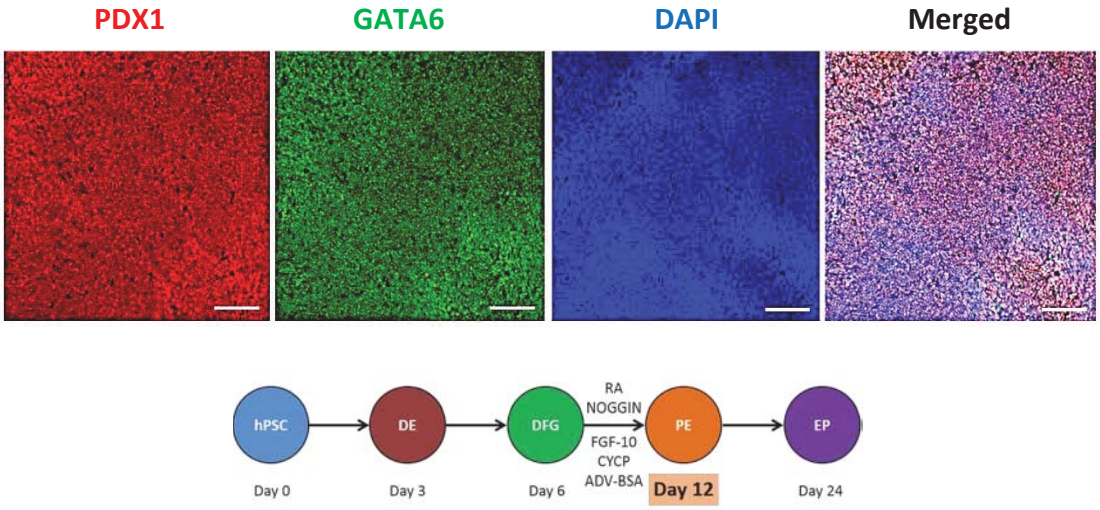


Figure 37. GATA6 is co-localised with key marker of the pancreatic endoderm. Cells were grown in culture conditions that specified them toward the pancreatic endoderm lineage and analysed via immunofluorescence. Cells were fixed on day 12 and were stained for key pancreatic marker PDX1. Scale bar, 100 µm.

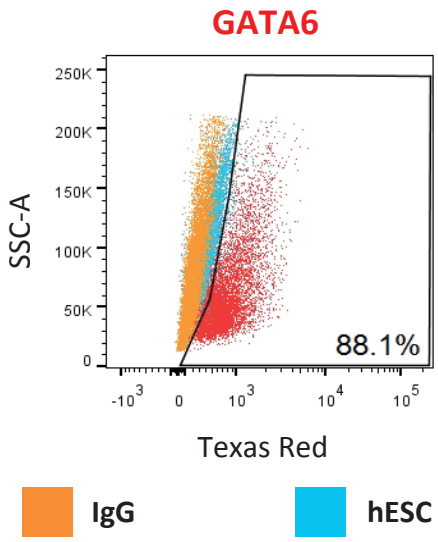


Figure 38. FACS of GATA6+ H9 cells at day 12. Cells were grown in culture conditions that specified them toward the pancreatic endoderm lineage and then analysed by FACS. Cells were fixed on day 12 and were stained for GATA6. Data show results of one experiment that is representative of at least 3 independent experiments.

On day 24, GATA6 continues to be expressed at high levels as shown by ICC (Figure 39) and qRT-PCR (Figure 33). Together, these results suggest that GATA6 plays an important role during pancreatic specification.

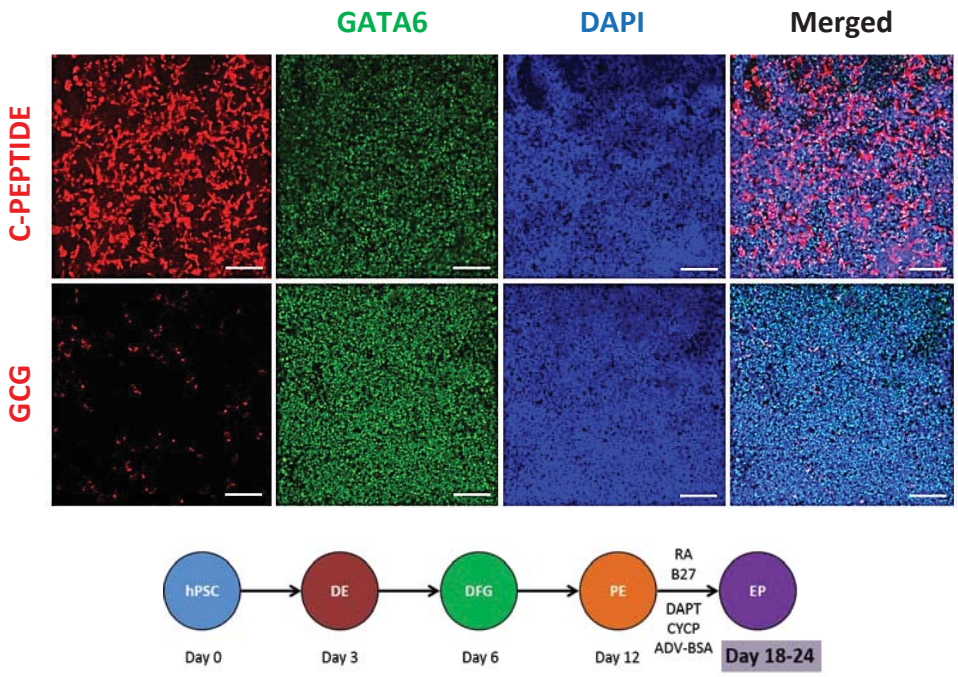


Figure 39. GATA6 is co-localised with key markers of endocrine progenitors. Cells were grown in culture conditions that allowed them to mature into endocrine progenitors and analysed via immunofluorescence. Cells were fixed on day 24 and were stained for key markers C-PEPTIDE and GCG. Scale bar, 100 µm.

3.2. Successful derivation of *GATA6* mutant lines

Using an efficient endocrine progenitor differentiation protocol, I have determined the expression profile of the transcription factor *GATA6*, where its expression is first detected during the formation of the DE and continues to be expressed throughout human pancreas development. Next, I sought to generate heterozygous and homozygous *GATA6* mutant cell lines in order to further investigate the role of *GATA6* in the formation of the human pancreas. To generate *GATA6* mutants, I employed TALENs to induce mutations at two specific target sites that were selected based on their close proximity downstream of the first and second start codons of the *GATA6* gene to avoid generating partial protein products.

3.2.1. NHEJ pathway

To generate *GATA6* mutants via the NHEJ pathway, I employed TALENs to induce mutations at both target sites in the *GATA6* gene as described in Chapter 2.3.1 in both H9 and FSPS13.B cell hPSC lines. In the first instance, H9 cells were targeted at both *GATA6* TALEN1 and TALEN2 sites (Figure 40) and as described in Chapters 2.3.1 and 2.4.1. After TALEN targeting via electroporation, antibiotic selection and recovery, approximately 20-80 colonies derived from single cells were present in each culture dish. To reveal potential heterozygous or homozygous clones, restriction digest was performed to screen for mutations. Genomic DNA extracted from each colony was digested using restriction enzymes AfeI and PstI for TALEN1 and TALEN2 respectively. Colonies that were successfully cut by the TALENs and incorporated mutations upon DNA repair, and thus and remained undigested due to a disruption within the restriction enzyme site. Colonies that did not harbour any mutations have the restriction enzyme site intact, and therefore resulted in smaller and more numerous DNA fragments. This is apparent in Figure 41 where colonies numbered in red are colonies with a disrupted AfeI site showing a band resembling the undigested wild-type band, hence indicating the occurrence of a likely mutational event in an allele. These clones also showed the PCR fragments resulting from AfeI digestion on a wild-type allele, suggesting that they were likely heterozygous *GATA6* mutants.

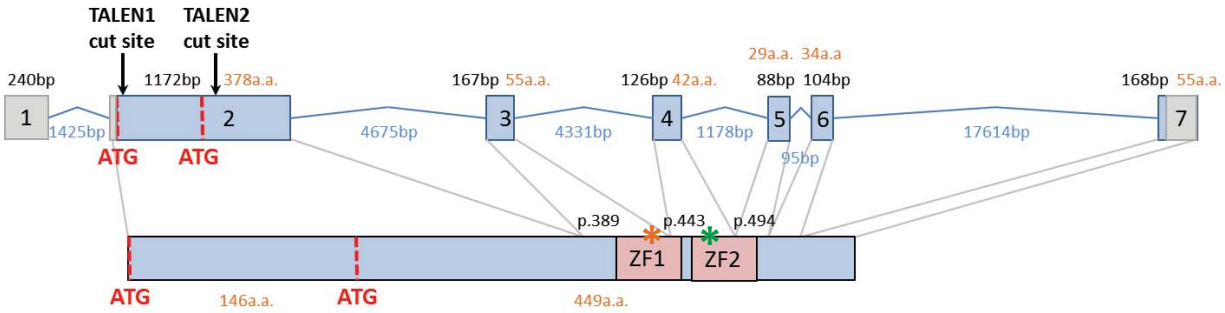


Figure 40. Schematic of TALEN1 and TALEN2 cut sites on the GATA6 locus. Top row represents *GATA6* cDNA showing the exons as boxes numbered accordingly and introns as lines. Grey regions of the exons represent non-coding regions while blue regions represent coding regions. The number of base pairs in each exon is shown in black text; the number of base pairs in each intron is shown in blue text; number of amino acids in each exon is shown in orange text. Bottom row illustrates the GATA6 protein and the grey lines connecting the cDNA to protein represents the corresponding area of the protein that each exon codes for. Protein translation commencing from the first ATG forms the long isoform of GATA6, which has a length of 595 amino acids. Protein translation commencing from the second ATG forms the short isoform of GATA6, which has a length of 449 amino acids. Green and orange asterisks indicate the locations of Patient A, *GATA6*$^{R465C/+}$ and Patient B, *GATA6*$^{c.1136\text{-}2A>G/+}$ mutations respectively.

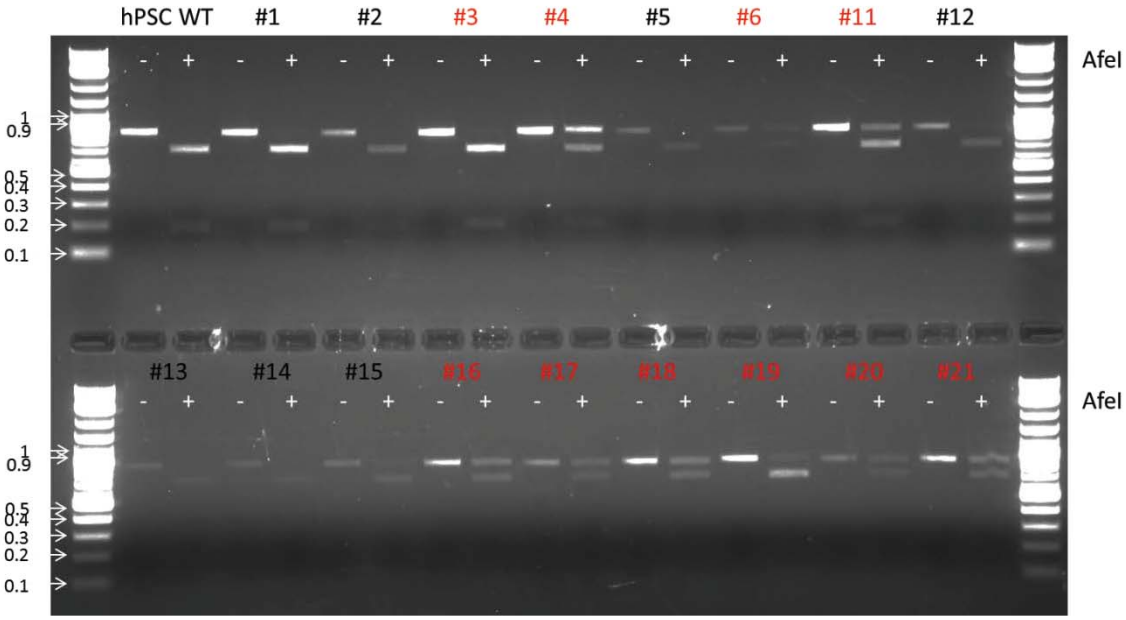


Figure 41. Representative DNA agarose gel picture of colonies screened via restriction enzyme digest. Genomic DNA was extracted from picked colonies and subjected to restriction digest using AfeI. Undigested and digested DNAs were run simultaneously. Colonies numbered red indicate the presence of heterozygous mutations while colonies numbered in black indicate wild-type.

24 clones were screened for each TALEN1 and TALEN2 targeting, with an observed cutting efficiency of approximately 80% at either targeting site, and a cutting efficiency of approximately 5% on both alleles at the TALEN2 target site. Unfortunately, no homozygous mutant for the TALEN1 cut site was generated in H9 cells. The colonies were then sequenced via Sanger sequencing to confirm their genotype. One heterozygous mutant derived from TALEN1 and TALEN2 target sites and one homozygous mutant derived from TALEN2 target site that harboured out of frame frameshift mutations resulting in a premature stop codon were selected for further western blotting validation (Figure 42). The TALEN-derived mutant lines were labelled according to their TALEN cut site (e.g. T1 for TALEN1 cut site or T2 for TALEN2 cut site) followed by the colony number.

To control for any potential off-target effects, I selected one colony that had been treated similarly to the mutants i.e. had undergone the TALEN targeting, antibiotic selection and re-growth process but harboured no observable mutations (Figure 42). All the TALEN-generated lines were karyotyped regularly to ensure chromosomal stability.

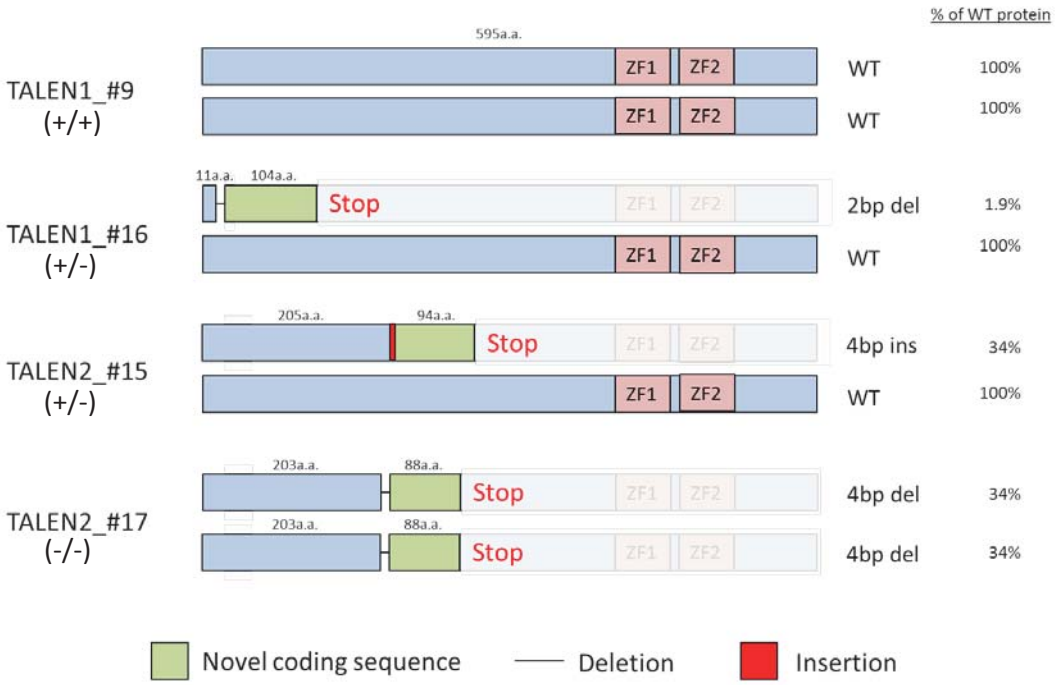


Figure 42. Schematic of selected H9 TALEN-derived *GATA6* cell lines. Four TALEN-derived cell lines were selected for use in downstream experiments. TALEN1_#9 is a TALEN-targeted but wild-type line; TALEN1_#16 contains a 2 base pair deletion frameshift mutation at the TALEN1 cut site; TALEN2_#15 contains a 4 base pair insertion frameshift mutation at the TALEN2 cut site; TALEN2_#17 contains an identical 4 base pair deletion on both *GATA6* alleles at the TALEN 2 cut site. All frameshift mutations were out of frame resulting in a premature stop codon.

Next, I performed western blotting to investigate GATA6 protein level in the TALEN-derived mutant lines. Cells were harvested on day 0 (D0) for negative control and day 3 (D3), as GATA6 is highly expressed on this day. Two antibodies, one recognising the N-terminus of the GATA6 protein and the other recognising the C-terminus of the GATA6 protein, were used. Expectedly, the N-terminal GATA6 antibody detected a partial protein product (PPP) for the mutants derived from the TALEN2 cut site, confirming the presence a truncated GATA6 protein caused by a premature stop codon in one *GATA6* allele (T2_#15) or both *GATA6* alleles (T2_#17) (Figure 43). The absence of GATA6 protein for T2_#17 mutant line using the C-terminal GATA6 antibody confirmed that this line was a homozygous mutant (Figure 44). The wild-type *GATA6* allele of heterozygous mutants T1_#16 and T2_#15 appeared similar to the wild-type lines H9 and T1_#9 (Figure 43 and Figure 44).

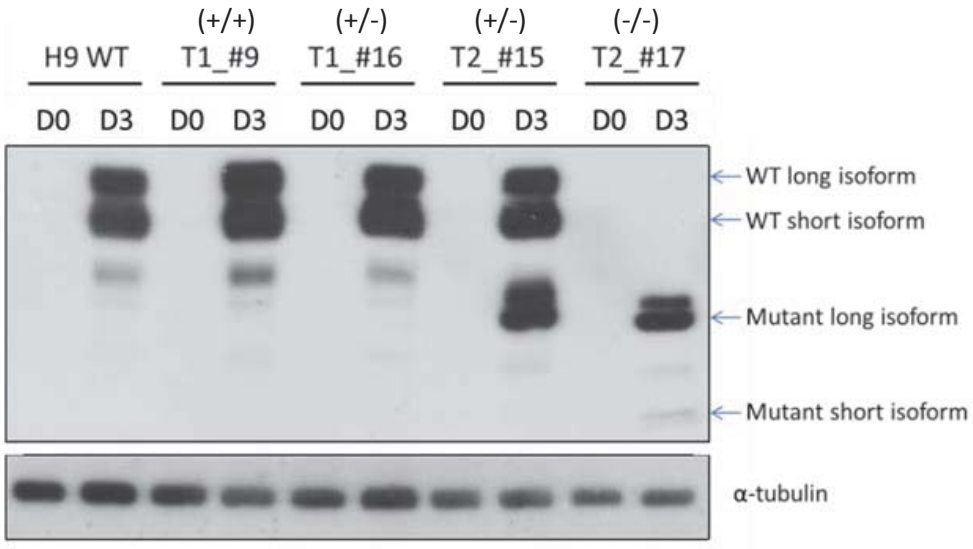


Figure 43. Western blot analysis of GATA6 protein levels in TALEN-derived H9 mutant lines using an N-terminal GATA6 antibody. Cells were harvested on day 0 (undifferentiated) and day 3 (DE). Alpha-tubulin was used as a loading control. Long and short isoforms of wild-type GATA6 are 64 kDa and 52 kDa respectively; the partial protein products for T2_#15 and T2_#17 are approximately 30 kDa and 18 kDa for the long and short isoforms respectively.

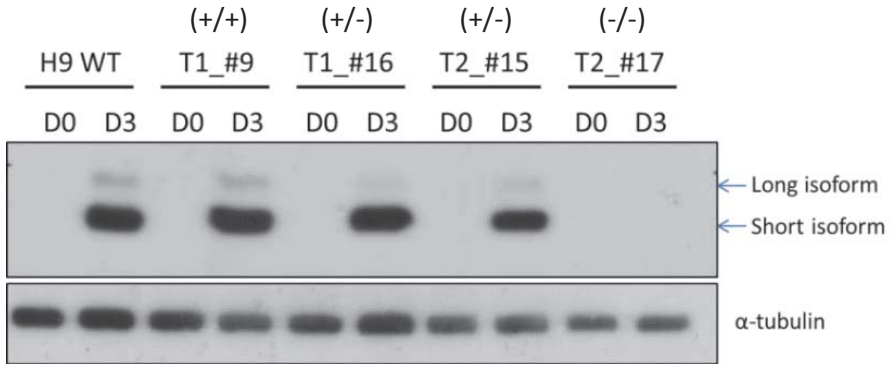


Figure 44. Western blot analysis of GATA6 protein levels in TALEN-derived H9 mutant lines using a C-terminal GATA6 antibody. Cells were harvested on day 0 (undifferentiated) and day 3 (DE). Alpha-tubulin was used as a loading control. Long and short isoforms of wild-type GATA6 are 64 kDa and 52 kDa respectively. No GATA6 protein was present for the T2_#17 mutant.

Gene editing was next performed in FSPS13.B cells. I focused on using the TALEN1 cut site with the intention of eliminating the PPP and obtaining complete loss of function *GATA6* alleles. 78 clones were screened with an observed cutting efficiency of approximately 10% on one allele. I obtained an identical mutant for FSPS13.B, referred to as T1_#50, that contained the same 2 base pair deletion as T2_#16 in H9 cells (Figure 45). Similar to H9 cells, no homozygous mutant for the TALEN1 cut site was recovered. Thus, I subjected the T1_#50 mutant line to re-targeting at the TALEN1 site. 48 clones were screened with an observed cutting efficiency of approximately 2% of the remaining wild-type allele. A homozygous mutant, referred to as T1_#50_#42, containing the same 2 base pair deletion on both alleles was obtained (Figure 45).

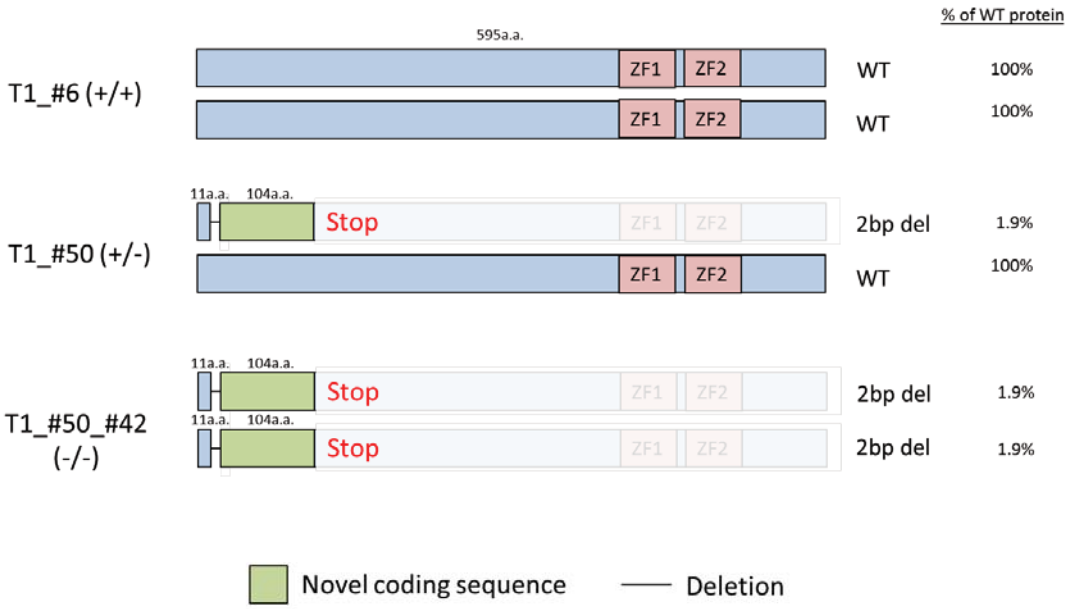


Figure 45. Schematic of selected FSPS13.B TALEN-derived *GATA6* cell lines at the TALEN1 cut site. Three TALEN-derived cell lines were selected for use in downstream applications. T1_#6 is a TALEN-targeted but wild-type line; T1_#50 contains a 2 base pair deletion frameshift mutation; T1_#50_#42 contains an identical 2 base pair deletion frameshift mutation on both *GATA6* alleles. All frameshift mutations were out of frame resulting in a premature stop codon.

Surprisingly, upon performing western blot to verify the complete loss of GATA6 protein in the T1_#50_#42 mutant, the short protein isoform of GATA6 was observed using both N- and C-terminus GATA6 antibodies (Figure 46). The simplest explanation for this finding is ribosomal read through and translational initiation at the second *GATA6* ATG (Figure 47).

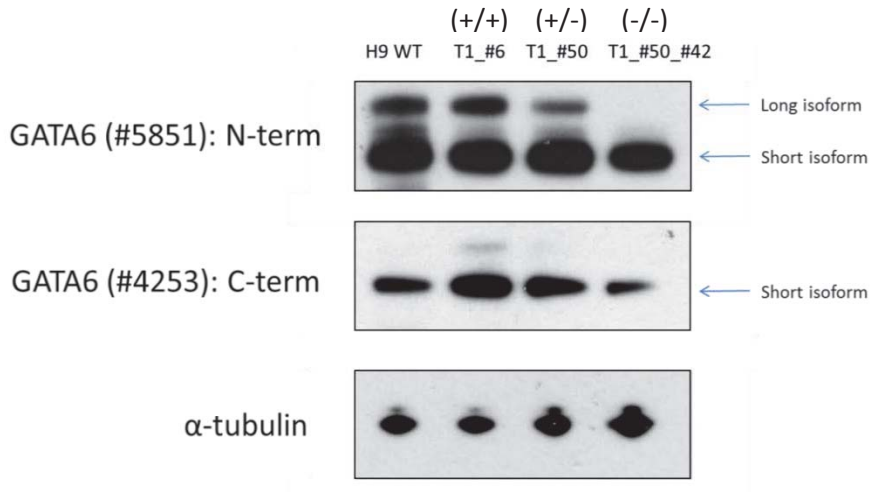


Figure 46. Western blot analysis of GATA6 protein expression in TALEN-derived FSPS13.B mutant lines using N- and C-terminal antibodies. Cells were harvested on day 3 (DE). Alpha-tubulin was used as a loading control. Long and short isoforms of wild-type GATA6 are 64 kDa and 52 kDa respectively. The short isoform of GATA6 protein was present for the T1_#50_#42 mutant.

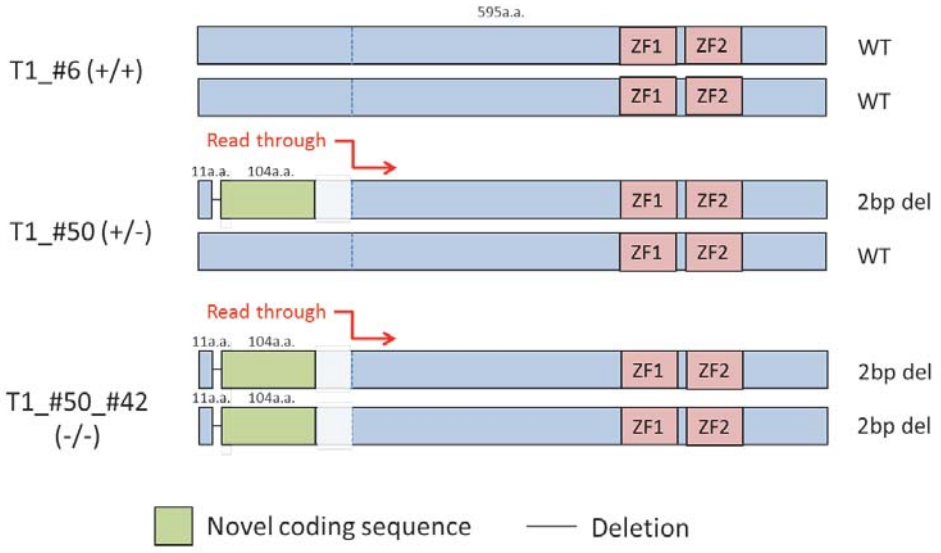


Figure 47. Schematic of selected FSPS13.B TALEN-derived *GATA6* cell lines at the TALEN1 cut site with read-through. The ribosomal read through and translational initiation at the second *GATA6* ATG led to the translation of the short GATA6 isoform in the mutants. Blue dotted lines represent the second start codon.

Since it is now clear that the TALEN1 cut site is not suitable for generating *GATA6* loss of function mutant lines, I proceeded to target the FSPS13.B cell line at the TALEN2 cut site. 33 clones were screened with an observed cutting efficiency of 30% of one allele. I selected three colonies which after sequencing were found to harbour out of frame heterozygous frameshift mutations resulting in premature stop codons (Figure 48). Again, no homozygous mutants were obtained and so I proceeded to retarget mutant T2_#8 which harbours a 14 base pair out of frame frameshift mutation. 24 clones were screened with an observed cutting efficiency of 30% of the remaining wild-type allele. A homozygous mutant, referred to as T2_#8_#4, containing an 11 base pair out of frame frameshift mutation in the other *GATA6* allele was selected for use in downstream experiments (Figure 48).

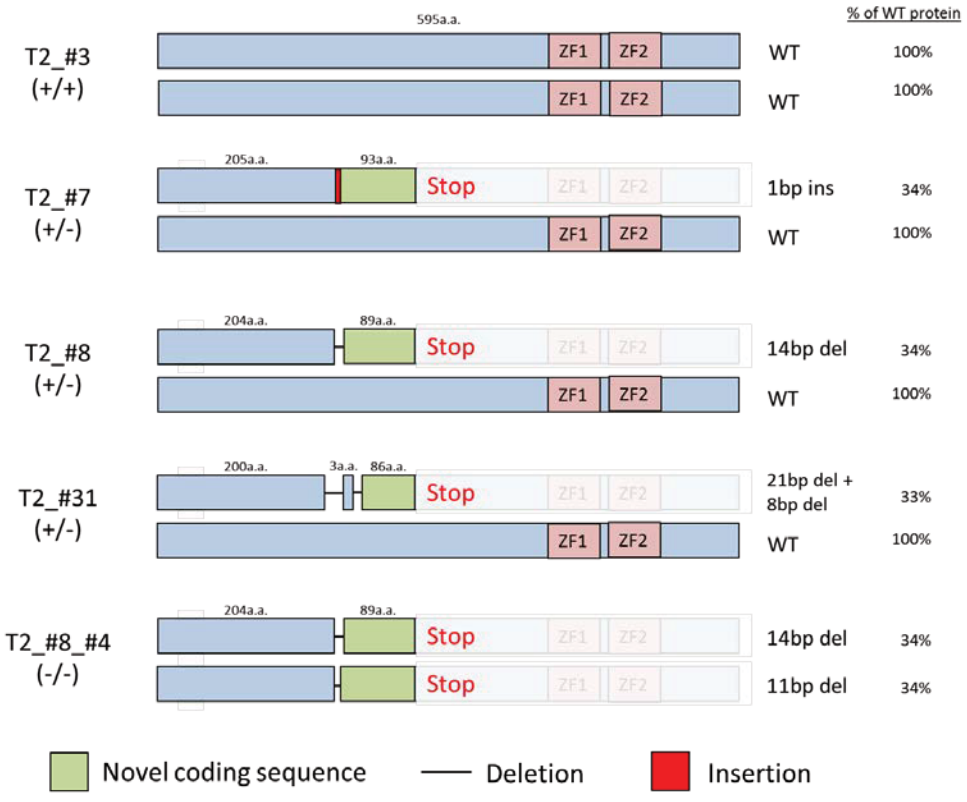


Figure 48. Schematic of selected FSPS13.B TALEN-derived *GATA6* cell lines at the TALEN2 cut site. Four TALEN-derived cell lines were selected for use in downstream applications. T2_#3 is a TALEN-targeted but wild-type line; T2_#7 contains a 1 base pair insertion frameshift mutation; T2_#8 contains a 14 base pair deletion frameshift mutation; T2_#31 contains a 21 and 8 base pair deletion frameshift mutation; T2_#8_#4 contains a 14 base pair deletion frameshift mutation in one *GATA6* allele and 11 in the other. All frameshift mutations were out of frame resulting in a premature stop codon.

GATA6 heterozygous and homozygous mutant H9 and FSPS13.B cell lines derived from TALEN targeting at the TALEN2 site that were selected for subsequent experiments are summarised in Table 51 with precise details of their genotypes. The nomenclature for the mutants that is used throughout the remainder of this thesis is listed in Table 51. All mutant lines were verified by western blotting and sequencing. They were also karyotyped and found to be normal and absent of any chromosomal abnormality.

Table 51. Summary of selected H9 and FSPS13.B mutants generated via NHEJ pathway

Cell line	TALEN cut site and colony number	Genotype	Nomenclature
H9	T1_#9	*GATA6* $^{+/+}$ (wild-type)	H9* (WT)
H9	T2_#15	*GATA6* $^{c.618\_619insTGCA/+}$	*GATA6* $^{4ins/+}$ (Het)
H9	T2_#17	*GATA6* $^{c.611\_614delACCT/c.611\_614delACCT}$	*GATA6* $^{\Delta4/\Delta4}$ (Hom)
FSPS13.B	T2_#3	*GATA6* $^{+/+}$ (wild-type)	FSPS13.B* (WT)
FSPS13.B	T2_#7	*GATA6* $^{c.615\_616insC/+}$	*GATA6* $^{1ins/+}$ (Het)
FSPS13.B	T2_#8	*GATA6* $^{c.del614\_627TGCAGGGGTCGGGC/+}$	*GATA6* $^{\Delta14/+}$ (Het)
FSPS13.B	T2_#31	*GATA6* $^{c.600\_621delinsTGGGCCAG/+}$	*GATA6* $^{\Delta21\_8ins/+}$ (Het)
FSPS13.B	T2_#8_#4	*GATA6* $^{c.del614\_627TGCAGGGGTCGGGC/}$ $^{c.del613\_623CTGCAGGGGTC}$	*GATA6* $^{\Delta14/\Delta11}$ (Hom)

3.2.2. HR pathway

Next, I generated *GATA6* knockout hPSC lines that are entirely devoid of a PPP. To accomplish this, I performed TALEN targeting at the TALEN1 site and introduced a donor/targeting vector (TG) that contained an emerald GFP (emGFP) reporter in frame with the endogenous *GATA6* ATG and a puromycin-resistance cassette (Figure 49). TALEN cutting and successful HR via the 5' and 3' homology arms results in either a heterozygous or homozygous loss-of-function *GATA6* mutation (Figure 49).

I targeted both H9 and FSPS13.B cells. Initial PCR screens on genomic DNA extracted from picked colonies using forward and reverse primers within exon 1 and intron 2 respectively as indicated by the red arrows in Figure 49 revealed the "knock-in" of the donor template from the targeting vector (Figure 50). For H9 cells, 12 clones were screened with an observed HR efficiency of approximately 20% on one allele and 10% on two alleles. For FSPS13.B cells, 24 clones were screened with an observed HR efficiency of 50% of one allele. Unfortunately, no homozygous mutants were obtained even after I attempted to re-target a heterozygous mutant and screened through 24 clones.

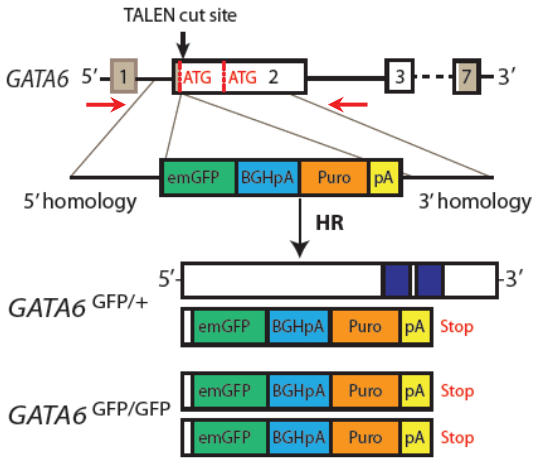


Figure 49. Schematic of generating heterozygous or homozygous loss-of-function *GATA6* mutations via HR. A "knock-in" vector that introduces an emerald GFP (emGFP) reporter in-frame with the first *GATA6* ATG and a puromycin resistance cassette into *GATA6* exon 2 is depicted. Successful homologous recombination resulted in both heterozygous (*GATA6* $^{GFP/+}$) and homozygous (*GATA6* $^{GFP/GFP}$) mutant cells. Red arrows indicate the positions of the primers used for PCR screening.

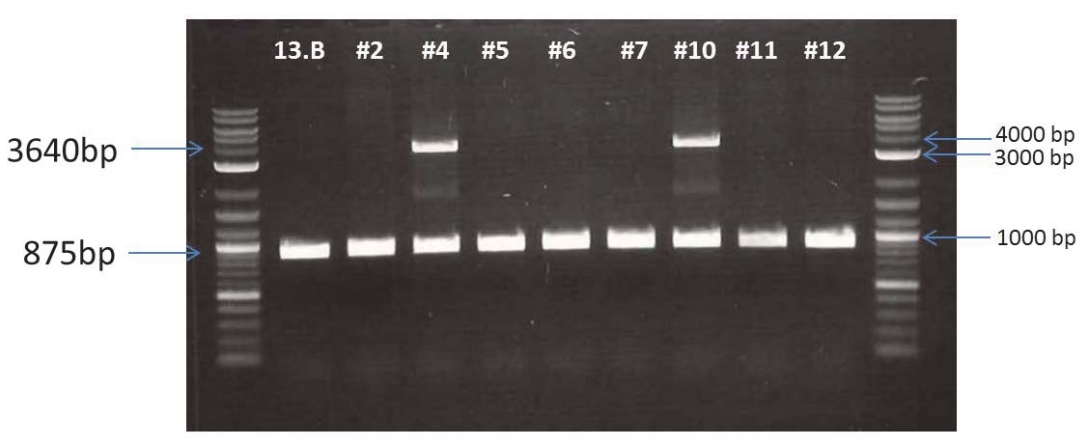


Figure 50. Representative DNA agarose gel picture of colonies screened via PCR to assess for successful HR. Genomic DNA was extracted from picked colonies and subjected to PCR using forward and reverse primers within exon 1 and intron 2 of the *GATA6* gene respectively. Successful HR of the donor template resulted in a 2765bp insertion as seen in colonies #4 and #10.

For H9 cells, I selected two heterozygous mutants and one homozygous mutant for use in downstream applications (Figure 51). For FSPS13.B cells, I selected two heterozygous mutants for use in downstream applications. To control for any potential off-target effects, I selected one colony that had been treated similarly to the mutants but harboured no observable HR or insertion of the donor template.

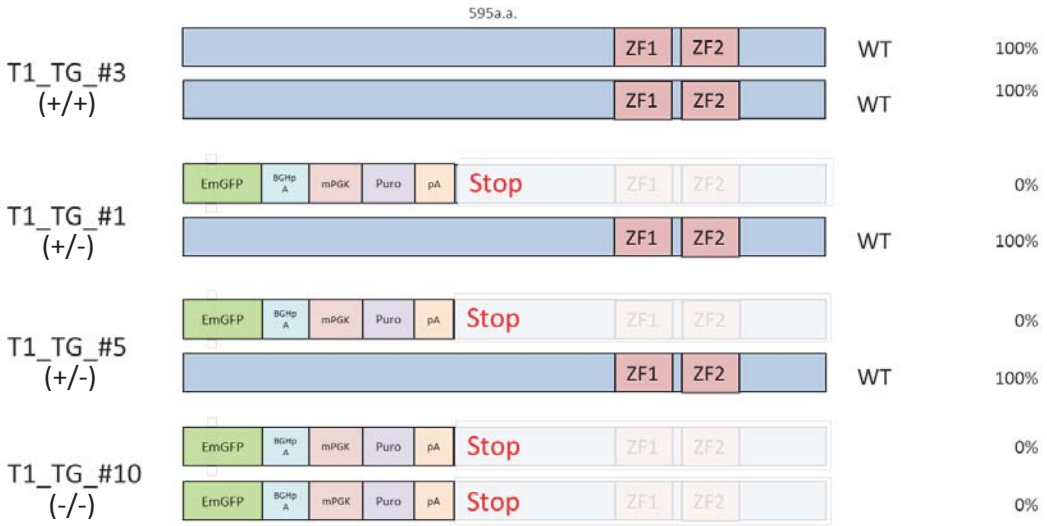


Figure 51. Schematic of selected H9 TALEN-derived *GATA6* cell lines via HR. Four TALEN-derived cell lines were selected for use in downstream applications. T1_TG_#3 is a TALEN-targeted but wild-type line; T1_TG_#1 and T1_TG_#5 are heterozygous *GATA6* knockouts; T1_TG_#10 is a homozygous *GATA6* knockout.

Western blotting using both N- and C-terminal antibodies confirmed the complete absence of GATA6 protein in *GATA6* $^{GFP/GFP}$ (Figure 52). All mutant lines were sequenced to ensure correct insertion of the donor template. This was done by sequencing the 3640bp PCR product containing the inserted fragment from the donor vector as shown in colonies #4 and #10 (Figure 50). They were also karyotyped and found to be normal and absent of any chromosomal abnormality. The mutants were further validated by the expression of emGFP (Figure 53).

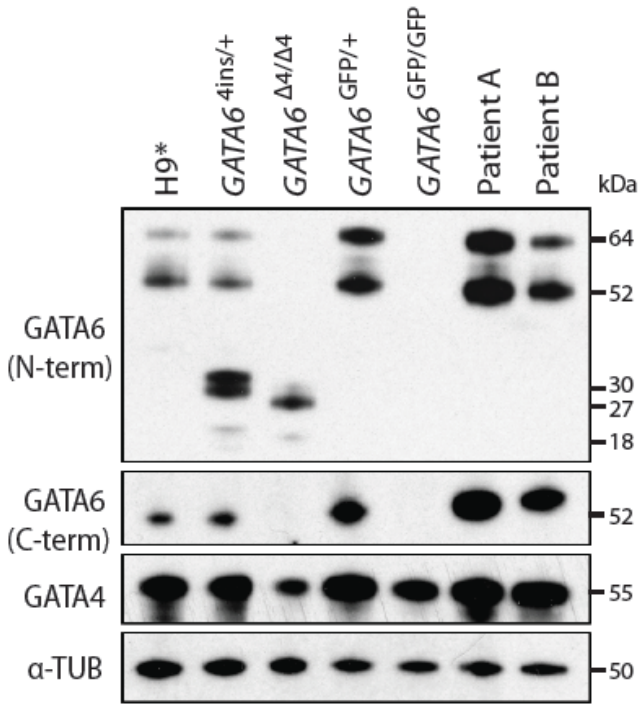


Figure 52. Western blot analysis of GATA6 and GATA4 protein levels in GATA6 mutant lines. H9*, *GATA6* $^{4ins/+}$, *GATA6* $^{\Delta4/\Delta4}$, *GATA6* $^{GFP/+}$, *GATA6* $^{GFP/GFP}$ mutant cells as well as Patients A and B were differentiated to day 3 (DE) and harvested. Alpha-tubulin was used as a loading control. Long and short isoforms of wild-type GATA6 are 64 kDa and 52 kDa respectively; the partial protein products for GATA6$^{4ins/+}$ are 30 kDa and 18 kDa for the long and short isoforms respectively; the partial protein products for GATA6$^{\Delta4/\Delta4}$ are 27 kDa and 15 kDa for the long and short isoforms respectively. No GATA6 protein was present for the *GATA6* $^{GFP/GFP}$ mutant.

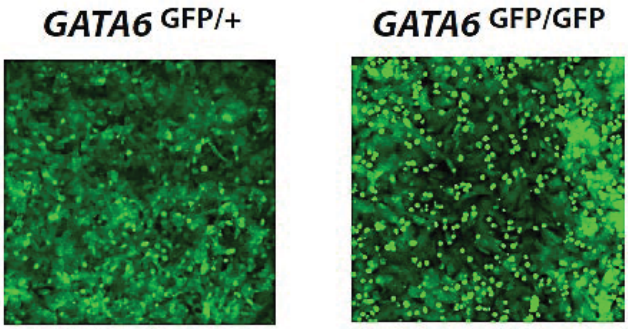


Figure 53. Immunofluorescence showing emGFP-expressing heterozygous *GATA6* $^{GFP/+}$ and homozygous *GATA6* $^{GFP/GFP}$ mutant cells on day 3. Cells were differentiated toward the DE and checked for the expression of emGFP.

GATA6 heterozygous and homozygous mutant H9 and FSPS13.B cell lines derived from TALEN targeting at TALEN1 site following HR repair pathway that were selected for use in downstream experiments are summarised in Table 52. The nomenclature for the mutants that is used for the subsequent sections of this report is also listed in Table 52.

Table 52. Summary of selected H9 and FSPS13.B mutants generated via HR pathway

Cell line	TALEN cut site and colony number	Nomenclature
H9	T1_TG_#3	H9* (WT)
H9	T1_TG_#1	*GATA6* $^{GFP/+}$ (Clone 1) (Het)
H9	T1_TG_#5	*GATA6* $^{GFP/+}$ (Clone 2) (Het)
H9	T1_TG_#10	*GATA6* $^{GFP/GFP}$ (Hom)
FSPS13.B	T1_TG_#2	FSPS13.B* (WT)
FSPS13.B	T1_TG_#4	*GATA6* $^{GFP/+}$ (Clone 1) (Het)
FSPS13.B	T1_TG_#10	*GATA6* $^{GFP/+}$ (Clone 2) (Het)

3.2.3. Reprogramming of GATA6 patient fibroblasts

In addition to the TALEN-derived GATA6 mutants, I also obtained fibroblasts from two *GATA6* patients in collaboration with Professor Andrew Hattersley.

Genotyping via Sanger sequencing revealed that Patient A (*GATA6* $^{R456C/+}$) has a missense mutation within the second zinc finger DNA-binding domain of the GATA6 protein, while Patient B (*GATA6* $^{c.1136\text{-}2A>G/+}$) contains a splice acceptor mutation in exon 3 (Figure 54). After reprogramming of the fibroblasts, three independent hiPSC clones were selected for each patient line, and these clones were also monitored for absence of the Sendai virus (Figure 55). All patient-derived cell lines were karyotypically normal and pluripotent (Figure 56).

Western blot analyses revealed the presence of both short and long isoforms of the GATA6 protein in both Patient A and B cell lines, suggesting the absence of a nonsense mediated decay (Figure 52).

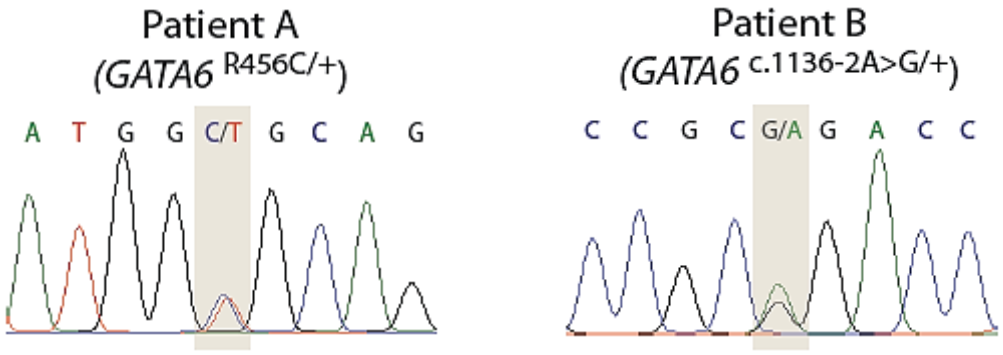


Figure 54. Genotype confirmation of Patients A and B by Sanger sequencing. Two *GATA6* patient-derived hiPSC lines—Patient A, *GATA6*$^{R465C/+}$ and Patient B, *GATA6*$^{c.1136\text{-}2A>G/+}$ were sequenced to confirm their heterozygous mutations.

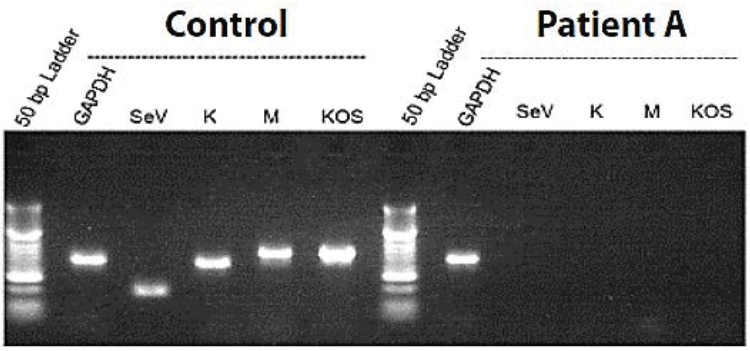


Figure 55. PCR showing loss of transgenes in Patient A mutant line, clone 1 compared with positive control. Picture is representative of 3 clones derived from each Patient A and Patient B mutant lines.

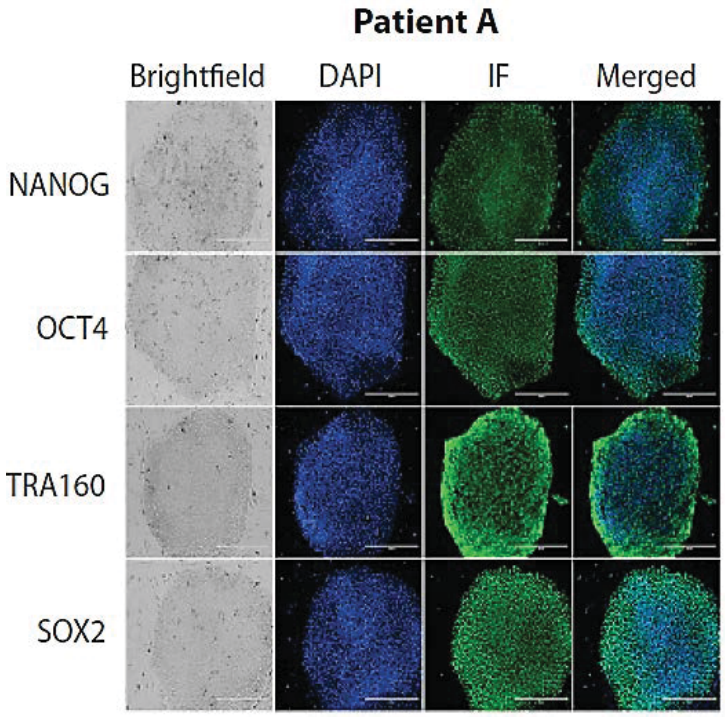


Figure 56. Immunofluorescence showing successful reprogramming of patient-derived Patient A (*GATA6* $^{R465C/+}$) mutant line via expression of pluripotency markers. Scale bars, 400 µm. Picture is representative of 3 clones derived from each Patient A and Patient B (*GATA6* $^{c.1136\text{-}2A>G/+}$) mutant lines.

3.2.4. Genome editing does not affect pluripotency

Next, I wanted to determine whether genome editing of the *GATA6* gene would affect pluripotency in the TALEN-derived *GATA6* mutants. ICC analyses on undifferentiated H9 *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ mutant cells showed high expression levels of key pluripotency markers NANOG, OCT4 and SOX2 and absence of key DE marker SOX17 as well as GATA6, indicating that the cells are indeed in a pluripotent state (Figure 57 and Figure 58).

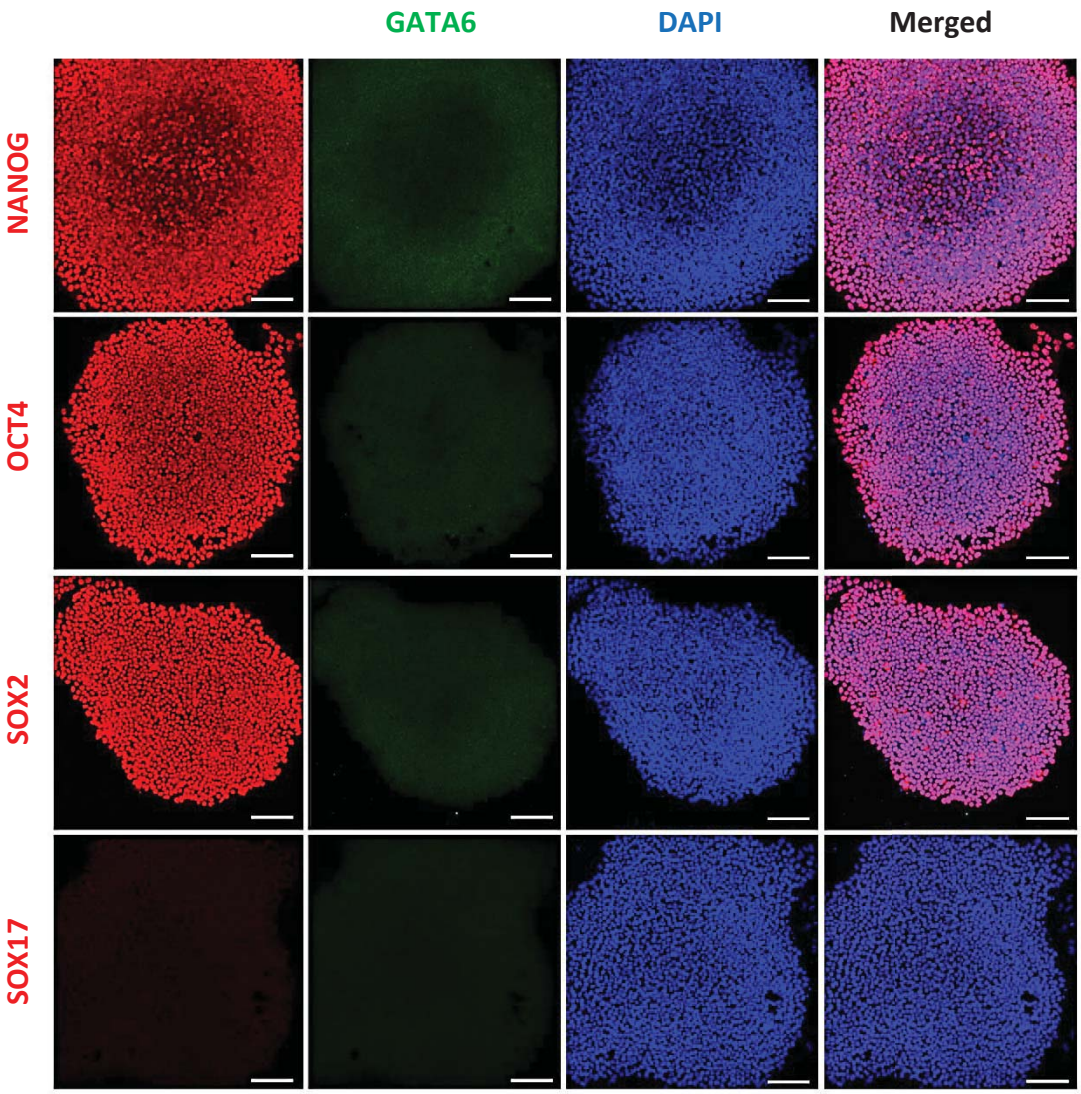


Figure 57. Pluripotency is maintained in *GATA6* $^{4ins/+}$ H9 cells. Cells were grown in culture conditions that maintained pluripotency and analysed via immunofluorescence. Cells were fixed on day 0 and were stained for the pluripotency markers *NANOG*, *OCT4* and *SOX2* and early DE marker *SOX17*. Scale bar, 100 µm.

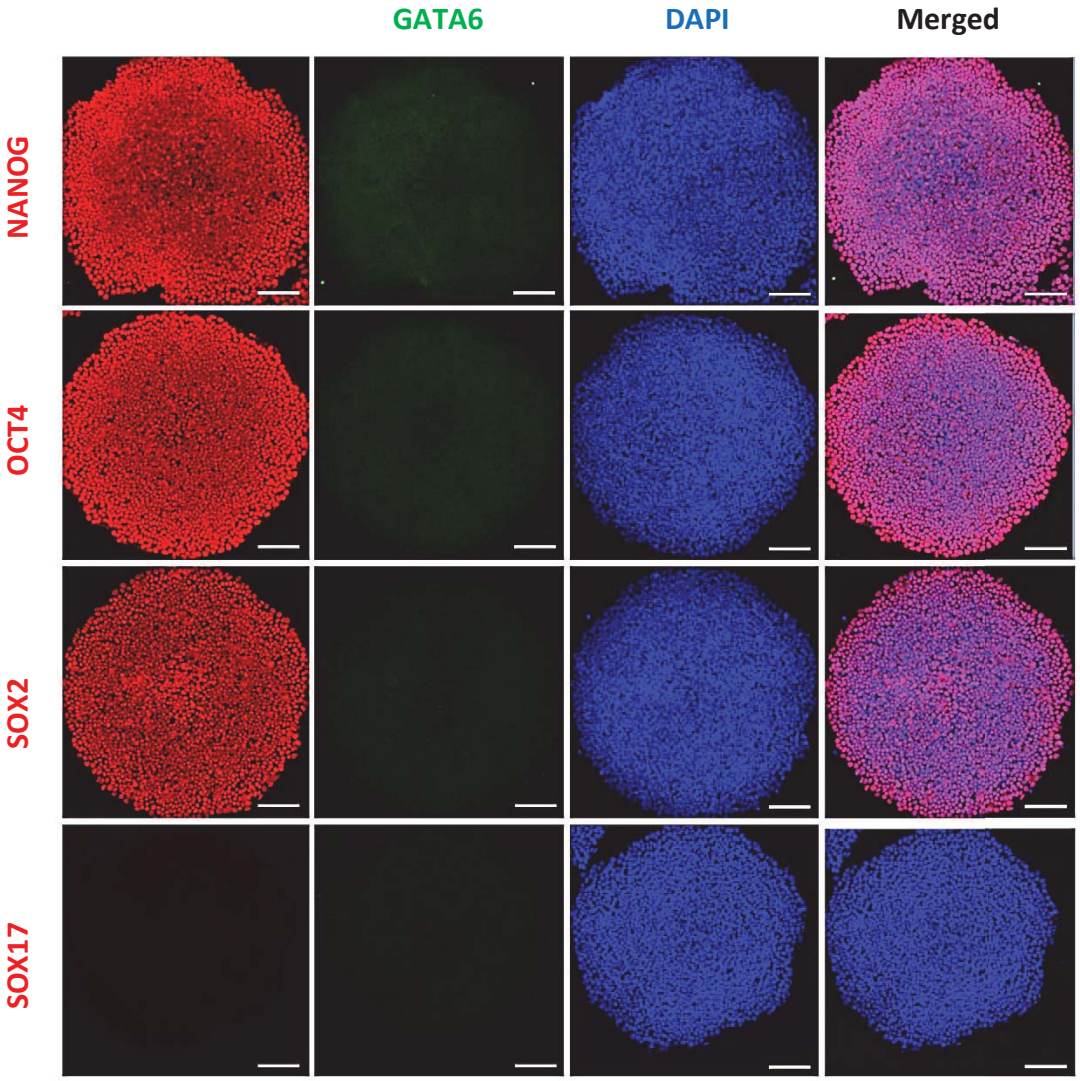


Figure 58. Pluripotency is maintained in *GATA6* $^{\Delta4/\Delta4}$ H9 cells. Cells were grown in culture conditions that maintained pluripotency and analysed via immunofluorescence. Cells were fixed on day 0 and were stained for the pluripotency markers *NANOG*, *OCT4* and *SOX2* and early DE marker *SOX17*. Scale bar, 100 µm.

Subsequent qRT-PCR analyses of the H9, FSPS13.B and patient *GATA6* mutant lines at D0 focusing on pluripotency genes such as *OCT4* and *SOX2* indicated that the mutant cell lines were similarly pluripotent , as described further in the next sections. Together, these results suggest genome editing on the *GATA6* gene does not affect pluripotency.

3.2.5. TALEN-derived wild-type cell lines resemble untargeted hPSCs

To assay for any potential off target effects, I differentiated TALEN-derived wild-type cells lines designated H9* and FSPS13.B* alongside their respective unmanipulated wild-type counterparts (H9 and FSPS13.B). qRT-PCR and FACS results show similar levels of key markers between the respective cell lines, indicating that DE and pancreatic differentiation were unaffected by the TALEN targeting, and further suggesting that there were no observable off target effects from the TALEN targeting (Figure 59, Figure 60 and Figure 61).

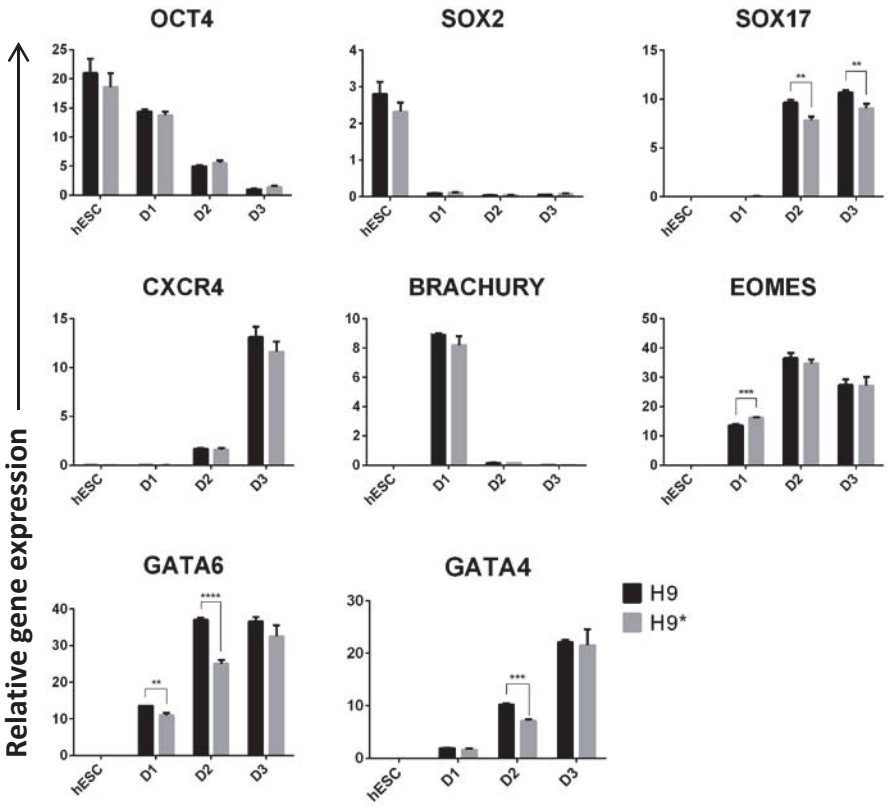


Figure 59. qRT-PCR analyses of H9 and H9* cells on days 1, 2 and 3. RNA was extracted at specific stages and the expression patterns of key markers were determined. Data are triplicate samples of one experiment and representative of three independent experiments. Error bars indicate standard deviation.

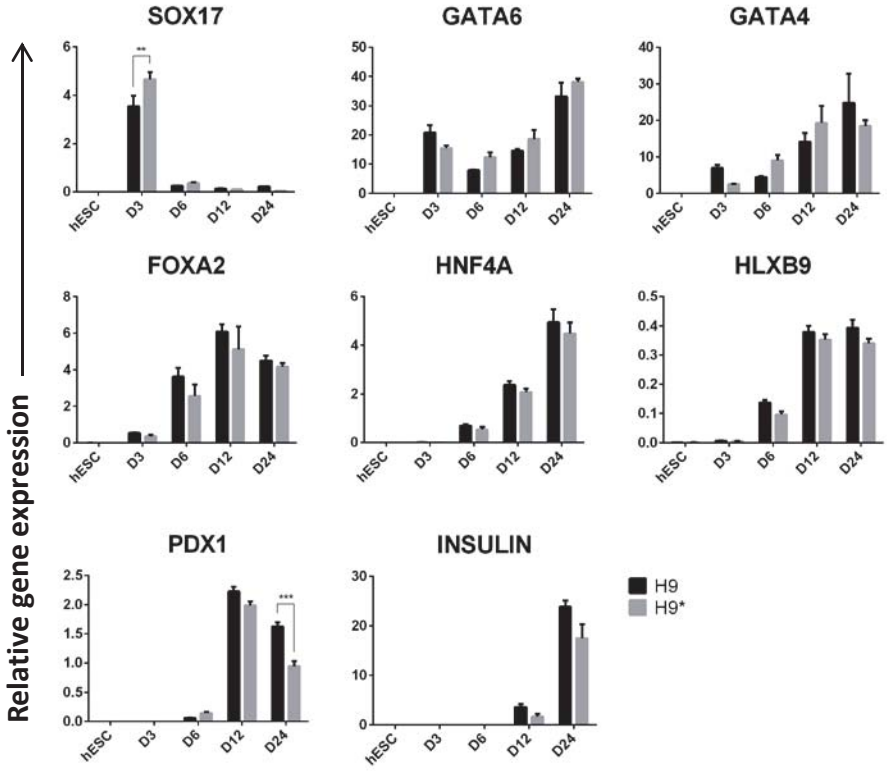


Figure 60. qRT-PCR analyses of H9 and H9* cells on days 3, 6, 12 and 24. RNA was extracted at specific stages and the expression patterns of key markers were determined. Data are triplicate samples of one experiment and representative of three independent experiments. Error bars indicate standard deviation.

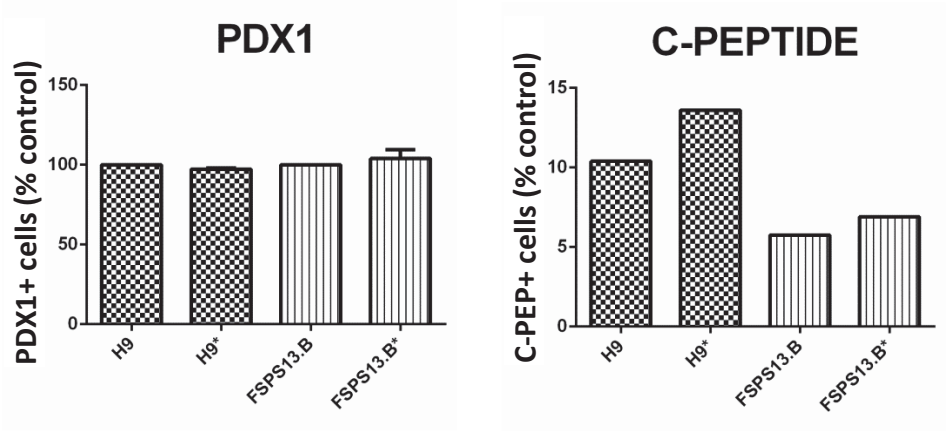


Figure 61. Summary of PDX1+ (day 12) and C-PEPTIDE+ (day 24) cells via FACS for H9, H9*, FSPS13.B and FSPS13.B. For PDX1, H9 and FSPS13.B cells are normalised to 100% and the relative PDX1+ cells to their respective mutant cell lines are shown. Data show results of three independent experiments and error bars indicate standard deviation. For C-PEPTIDE, absolute percentages of C-PEPTIDE+ cells in all cell lines are shown. Each bar represents one biological sample, and the graph was taken from one experiment, which is representative of three independent experiments.

3.3. Endodermal formation is inconsistently impaired by heterozygous *GATA6* mutations

Next, to study the effects of heterozygous or homozygous loss of *GATA6* during human pancreatic development, I subjected the mutant lines to pancreatic differentiation in order to characterise their phenotypes and perform phenotypical comparisons between the various genotypes with an initial focus at the DE stage.

3.3.1. Biallelic loss of N-terminal of GATA6 protein impairs DE formation

To determine if loss of the N-terminal GATA6 protein upstream of the second start codon affects the formation of the DE, I differentiated the FSPS13.B lines derived from TALEN1 targeting toward the DE and performed FACS of key DE marker CXCR4 at day 3 (refer to Figure 47). Results from FACS suggested that in FSPS13.B cells, the loss of GATA6 protein between the first and second start codon on one allele does not affect DE differentiation (84% of CXCR4+ cells for T1_#50), but the biallelic loss of GATA6 protein between the first and second start codon impairs DE specification by approximately 30% (57% of CXCR4+ cells for T1_#50_#42) (Figure 62). This is in concordance to an early study where it was reported that the N-terminal 146 amino acids of GATA-6 contains transactivational activity, and its deletion reduced luciferase expression by 50% (Brewer et al., 1999).

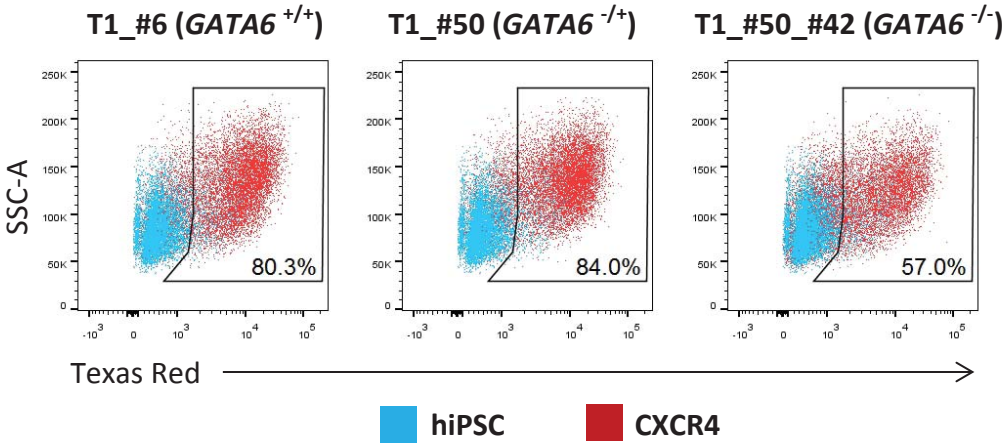


Figure 62. FACS of CXCR4+ cells for FSPS13.B TALEN1-targeted mutant cells on day 3. Cells were grown in culture conditions that specified them toward the DE lineage and they were analysed via FACS. Cells were fixed on day 3 and were stained for CXCR4. Gates are set according to undifferentiated FSPS13.B cells. Data show results of one experiment that is representative of at least 3 independent experiments.

3.3.2. Homozygous *GATA6* mutants fail to form the DE

It was reported that mouse embryos lacking both copies of *Gata6* die during gastrulation (Morrisey et al., 1998, Koutsourakis et al., 1999). This early embryonic lethality is believed to be a consequence of extraembryonic endoderm dysfunction and arrest development at E6.5-7, as this deficiency could be overcome by providing *Gata6*-null embryos with a wild-type extraembryonic endoderm with the use of tetraploid embryo complementation (Zhao et al., 2005).

To determine if this early endoderm dysfunction is recapitulated in humans using our *in vitro* model system, I differentiated TALEN-derived homozygous mutants in both H9 (*GATA6* $^{\Delta4/\Delta4}$ and *GATA6* $^{GFP/GFP}$) and FSPS13.B (*GATA6* $^{\Delta14/\Delta11}$) cell lines. At day 3, the loss of pluripotency marker NANOG in all cell lines (Figure 63) and the high expression of key DE marker SOX17 in H9* cells (Figure 64) indicated the successful differentiation of cells toward the DE lineage. H9-derived *GATA6* $^{\Delta4/\Delta4}$ and *GATA6* $^{GFP/GFP}$ homozygous mutants, however, displayed negligible expression of SOX17 as shown by ICC (Figure 64). FACS analyses showed a 90-100% decrease of SOX17+ and CXCR4+ cells (Figure 65 and Figure 66). FACS analyses of FSPS13.B-derived *GATA6* $^{GFP/GFP}$ homozygous mutant also showed a similar result (Figure 67). The loss of DE markers in these homozygous mutants was confirmed by qRT-PCR analyses (Figure 68, Figure 69 and Figure 70). That CXCR4+ cells were similarly abolished in H9-derived *GATA6* $^{\Delta4/\Delta4}$ and *GATA6* $^{GFP/GFP}$ mutants suggest that the PPP in *GATA6* $^{\Delta4/\Delta4}$ mutant cells does not have a functional role during differentiation (Figure 71). Expectedly, the homozygous mutants failed to develop into the subsequent key stages toward pancreatic development such as the dorsal foregut (Figure 72, Figure 73 and Figure 74).

Taken together, these results are in concordance with previously published data in mice where *Gata6* $^{-/-}$ embryos displayed early embryonic lethality, which is believed to be a consequence of extraembryonic endoderm dysfunction. My results show that absence of *GATA6* in *GATA6* $^{-/-}$ hPSC mutants generated from TALEN genome editing abolished the formation of the DE.

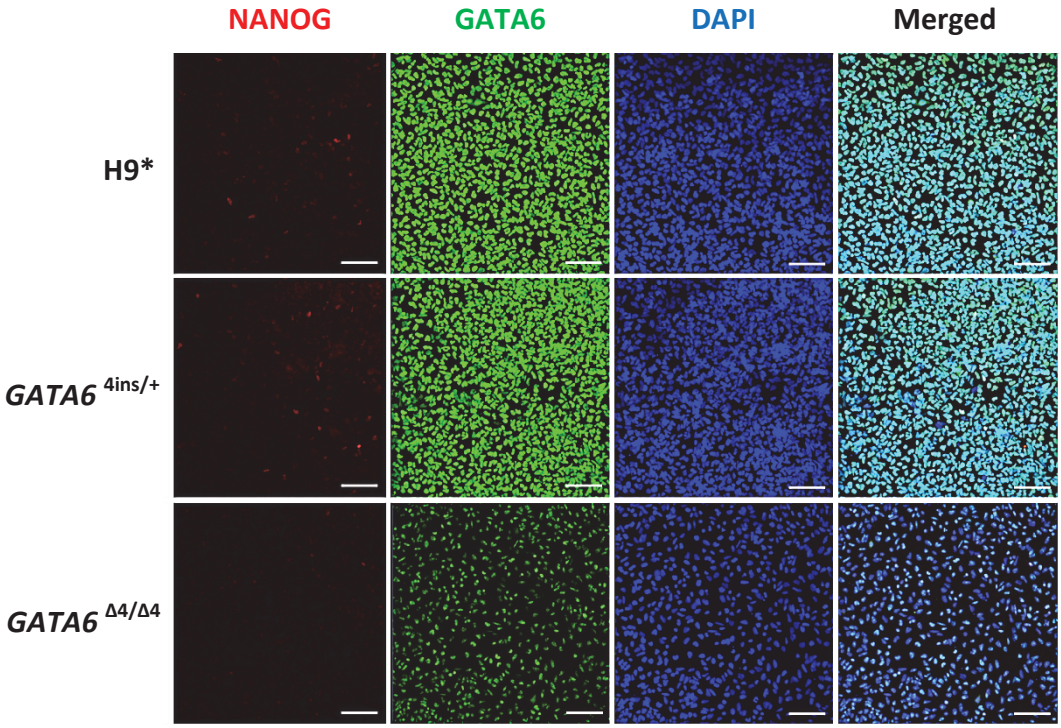


Figure 63. *GATA6* $^{4ins/+}$ and *GATA6* $^{Δ4/Δ4}$ mutants had the capacity to differentiate. Cells were grown in culture conditions that specified them toward the DE lineage and analysed via immunofluorescence. Cells were fixed on day 3 and were stained for the pluripotency marker NANOG. Scale bar, 100 µm.

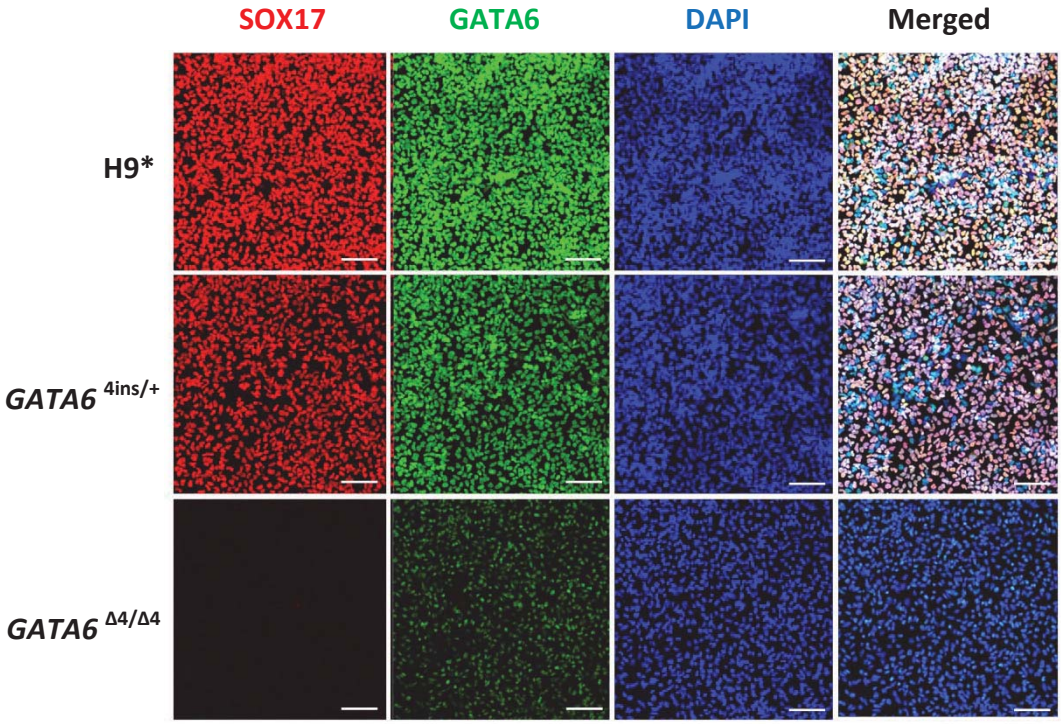


Figure 64. SOX17 expression is abolished *GATA6* $^{Δ4/Δ4}$ cells. Cells were grown in culture conditions that specified them toward the DE lineage and analysed via immunofluorescence. Cells were fixed on day 3 and were stained for the DE marker SOX17. Scale bar, 100 µm.

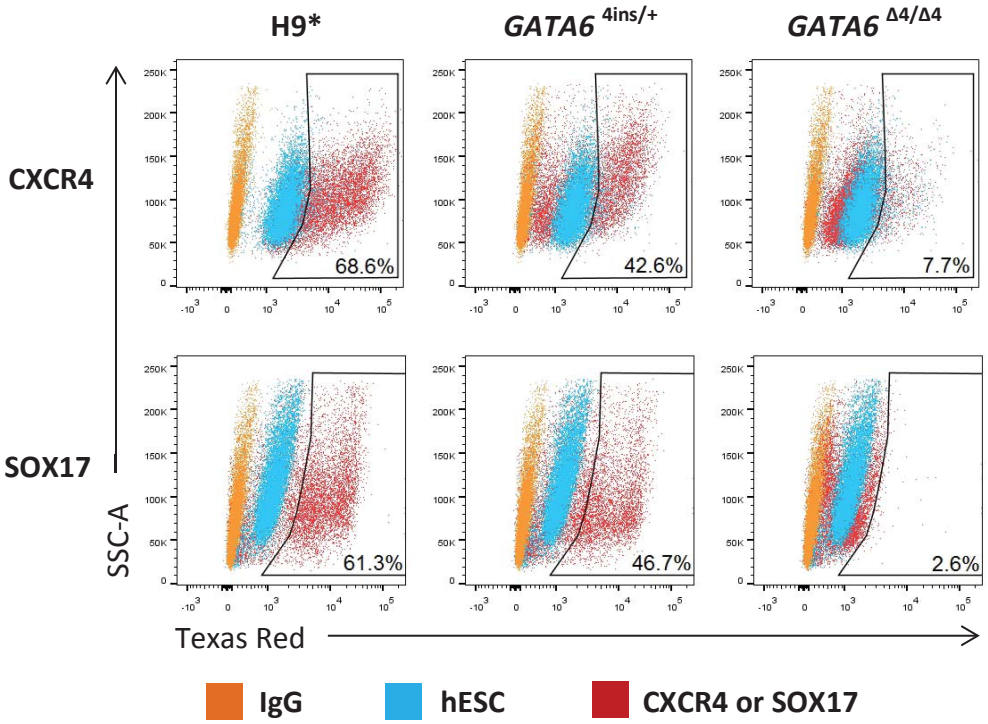


Figure 65. The number of CXCR4+ and SOX17+ cells are decreased in *GATA6* $^{4ins/+}$ cells and are almost completely absent in *GATA6* $^{\Delta4/\Delta4}$ H9 cells. Cells were fixed on day 3 and were stained for the DE markers SOX17 and CXCR4. Data show results of one experiment that is representative of at least 3 independent experiments.

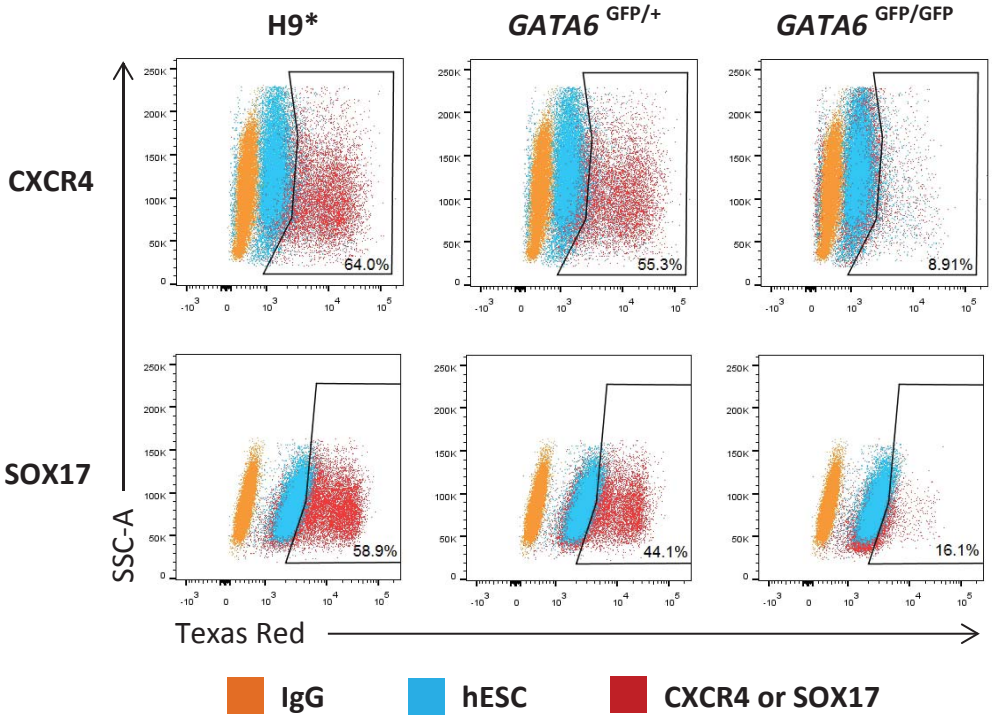


Figure 66. The number of CXCR4+ and SOX17+ cells are decreased in *GATA6* $^{GFP/+}$ cells and are almost completely absent in *GATA6* $^{GFP/GFP}$ H9 cells. Cells were fixed on day 3 and were stained for the DE markers SOX17 and CXCR4. Data show results of one experiment that is representative of at least 3 independent experiments.

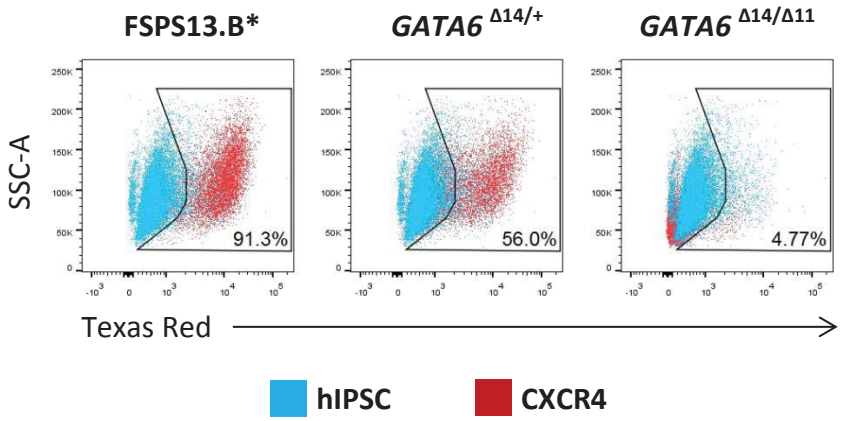


Figure 67. The number of CXCR4+ is decreased in *GATA6* $^{\Delta 14/+}$ cells and are almost completely absent in *GATA6* $^{\Delta 14/\Delta 11}$ FSPS13.B cells. Cells were fixed on day 3 and were stained for the DE marker CXCR4. Data show results of one experiment that is representative of at least 3 independent experiments.

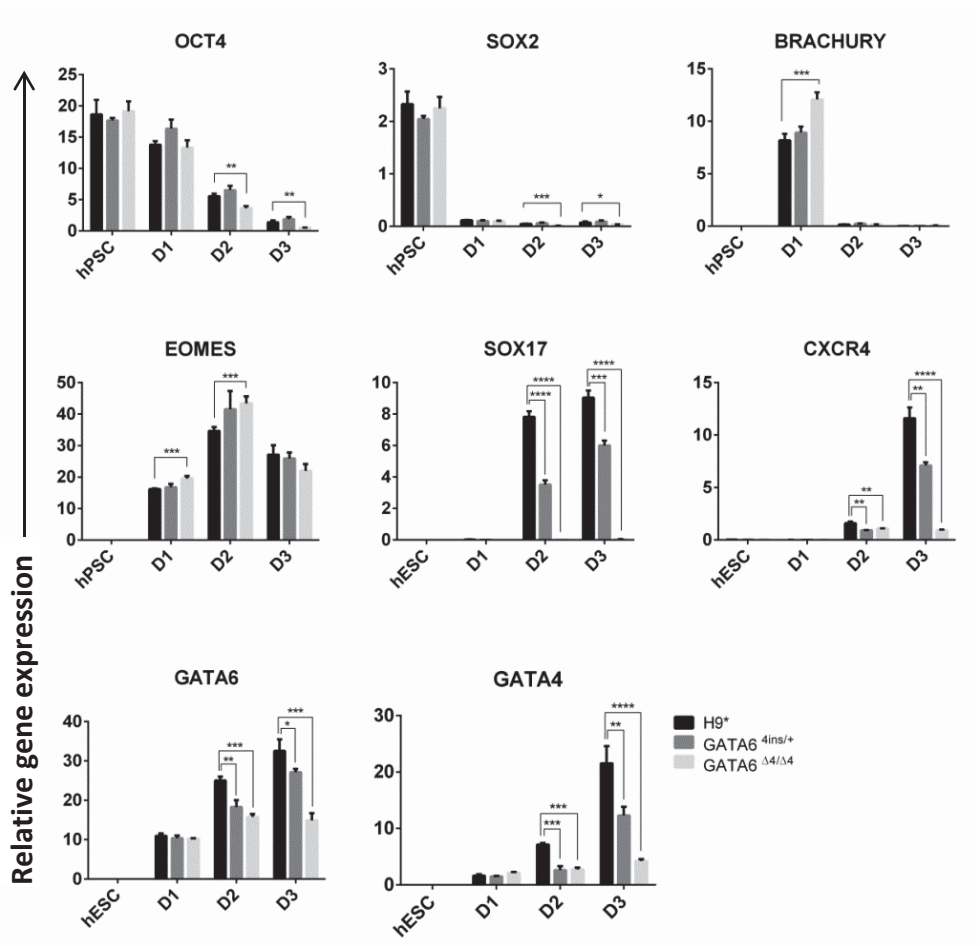


Figure 68. qRT-PCR analyses of H9*, *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta 4/\Delta 4}$ cells on days 1, 2 and 3. RNA was extracted at specific stages and the expression patterns of key markers were determined. Data are triplicate samples of one experiment and representative of three independent experiments. Error bars indicate standard deviation.

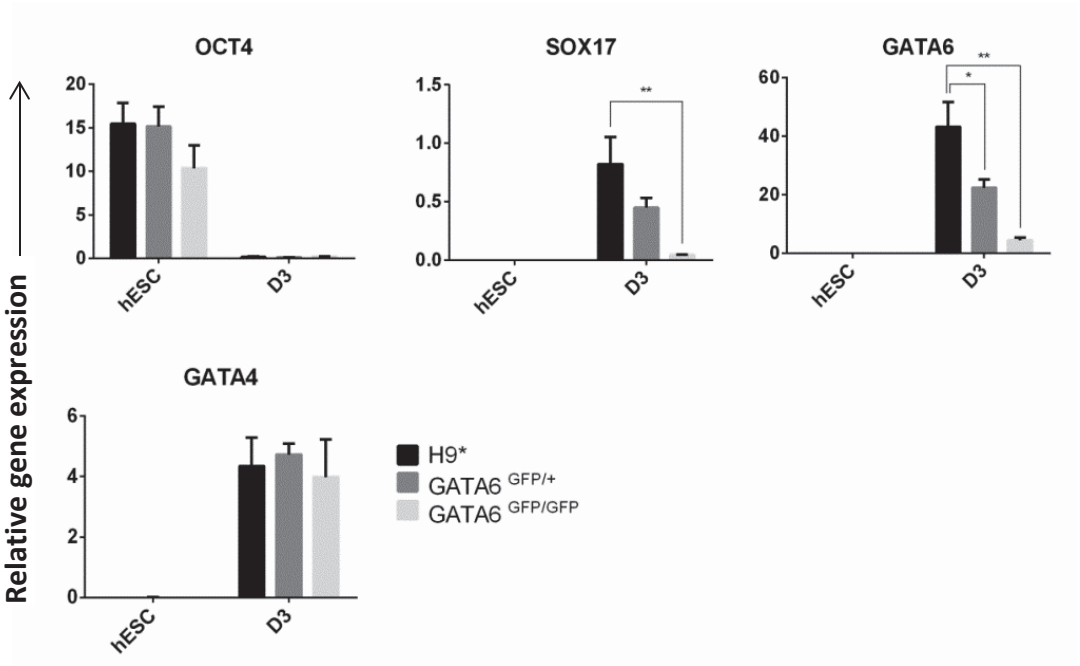


Figure 69. qRT-PCR analyses of H9*, *GATA6* $^{GFP/+}$ and *GATA6* $^{GFP/GFP}$ cells on day 3. RNA was extracted on day 3 and the expression patterns of key endoderm markers were determined. Data are triplicate samples of one experiment and representative of three independent experiments. Error bars indicate standard deviation.

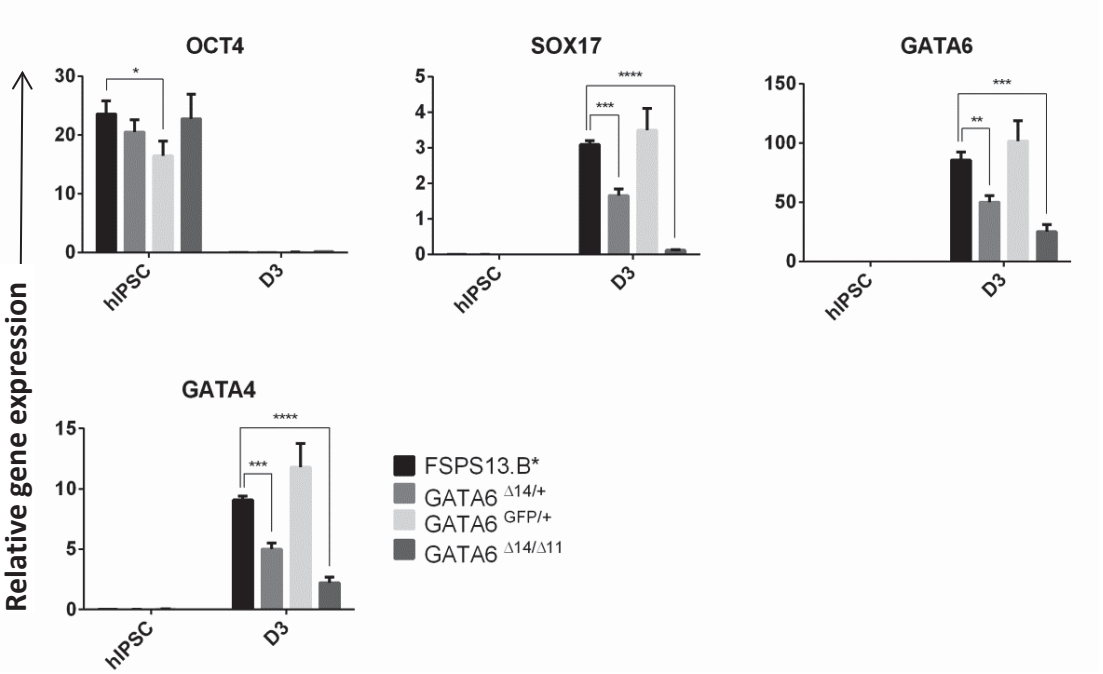


Figure 70. qRT-PCR analyses of FSPS13.B*, *GATA6* $^{\Delta 14/+}$, *GATA6* $^{GFP/+}$ and *GATA6* $^{\Delta 14/\Delta 14}$ cells on day 3. RNA was extracted on day 3 and the expression patterns of key endoderm markers were determined. Data are triplicate samples of one experiment and representative of three independent experiments. Error bars indicate standard deviation.

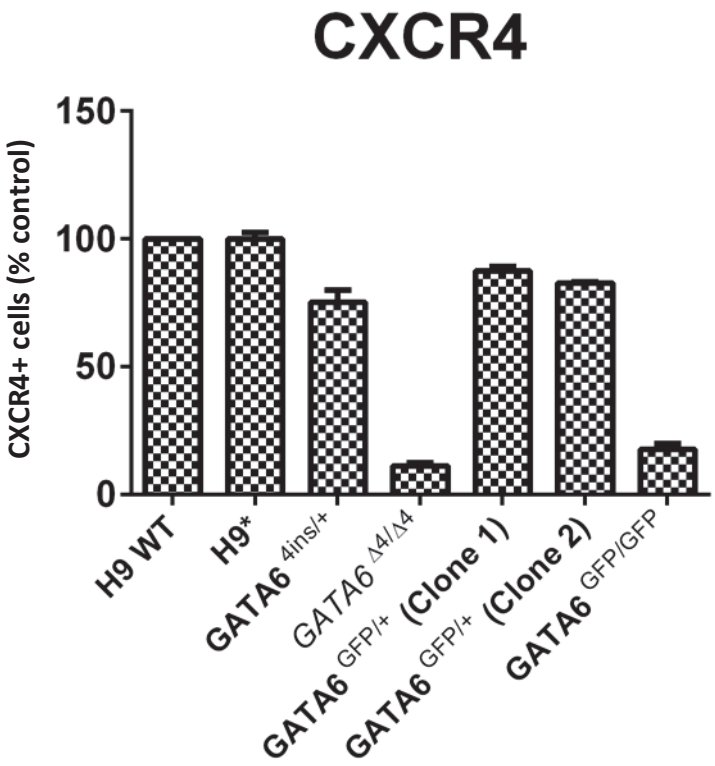


Figure 71. Summary of CXCR4+ cells via FACS for all H9-derived mutant cells on day 3. Wild-type H9 cells are normalised to 100% and the relative CXCR4+ cells of H9* or mutant cells are shown. Data show results of two experiments and error bars indicate standard deviation.

3.3.3. *GATA6* heterozygous mutants exhibit endodermal defects using lab-derived protocol

After determining the phenotypic consequence of homozygous *GATA6* mutations, I here shift my focus to heterozygous *GATA6* hPSC lines.

In H9-derived *GATA6* $^{4ins/+}$ heterozygous mutant cells, SOX17 is expressed on day 3, indicating that DE formation is not abolished (Figure 64). FACS analyses on day 3 for key DE markers CXCR4 and SOX17 for *GATA6* $^{4ins/+}$ and *GATA6* $^{GFP/+}$ mutant cells shows a moderate decrease in the number of CXCR4+ and SOX17+ cells of approximately 20% (Figure 65 and Figure 66). The relative number of CXCR4+ cells from FACS analyses comparing the *GATA6* $^{4ins/+}$ and *GATA6* $^{GFP/+}$ clones 1 and 2 mutants to H9* cells (normalised to 100%) indicate that the population of CXCR4+ cells for heterozygous mutants are consistently approximately 80% (Figure 71). The phenotypic similarity between *GATA6* $^{4ins/+}$ and both clones 1 and 2 of *GATA6* $^{GFP/+}$ also suggests that the one copy of the PPP present in the *GATA6* $^{4ins/+}$ mutant cells is non-functional during differentiation. ICC analyses showing expression of FOXA2 on days 3 and 6 (Figure 72), the absence of CDX2 (Figure 75) and HEX (Figure 76) on day 6 and expression of HNF1B on day 6 and *SOX2* on day 9 (Figure 77) indicate that *GATA6* $^{4ins/+}$ mutant cells are capable of differentiating into the dorsal foregut and foregut lineages.

Using the H9-derived *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ lines, I next asked whether *GATA6* haploinsufficiency impacts the very early differentiation from mesendoderm (corresponding to days 1 and 2) to DE (day 3). qRT-PCR analyses revealed the expression of primitive streak (*BRACHYURY*) and mesendoderm (*EOMESODERMIN* (*EOMES)*) markers were relatively unchanged across the control H9* and mutant lines, suggesting that early mesendoderm formation was not affected by either single or biallelic loss of *GATA6* (Figure 68). In contrast, consistent with FACS analyses, key DE markers *SOX17* and *CXCR4* were downregulated beginning on day 2 by approximately 30% in *GATA6* $^{4ins/+}$ and clones 1 and 2 of *GATA6* $^{GFP/+}$ cells (Figure 68 and Figure 69).

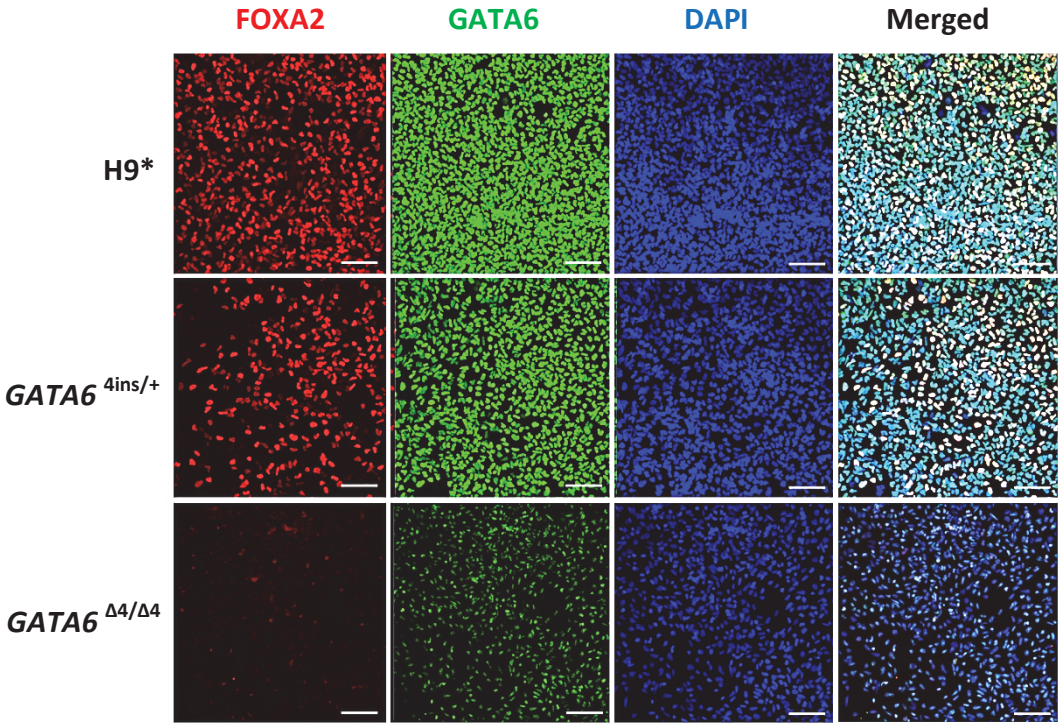


Figure 72. FOXA2 expression is abolished in *GATA6* $^{\Delta4/\Delta4}$ cells and decreased in *GATA6* $^{4ins/+}$ cells on day 3. Cells were grown in culture conditions that specified them toward the DE lineage and analysed via immunofluorescence. Cells were fixed on day 3 and were stained for the foregut marker FOXA2. Scale bar, 100 µm.

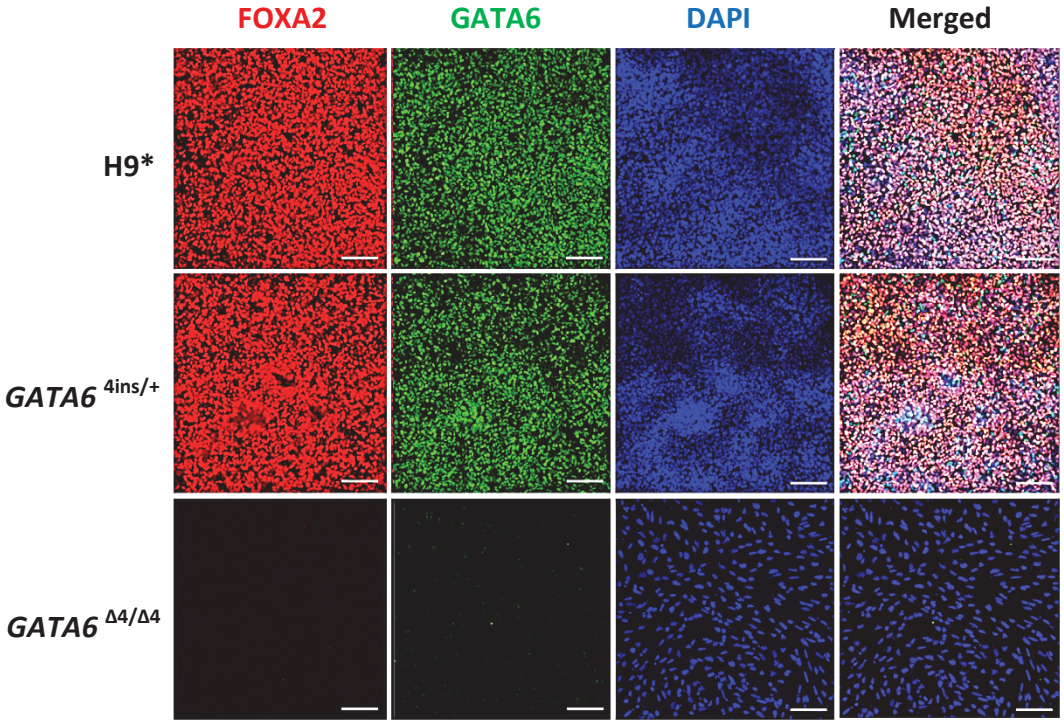


Figure 73. FOXA2 expression is abolished in *GATA6* $^{\Delta4/\Delta4}$ cells on day 6. Cells were grown in culture conditions that specified them toward the dorsal foregut lineage and analysed via immunofluorescence. Cells were fixed on day 6 and were stained for the foregut marker FOXA2. Scale bar, 100 µm.

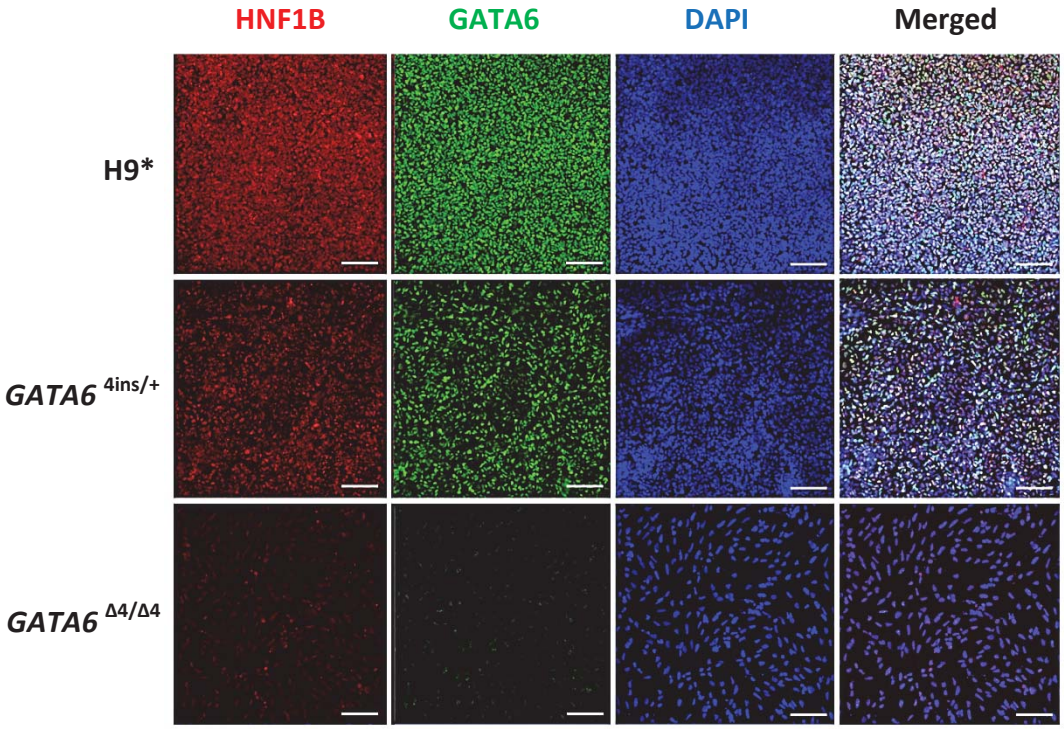


Figure 74. HNF1B expression is abolished in *GATA6* $^{\Delta4/\Delta4}$ cells on day 6. Cells were grown in culture conditions that specified them toward the dorsal foregut lineage and analysed via immunofluorescence. Cells were fixed on day 6 and were stained for the foregut marker HNF1B. Scale bar, 100 µm.

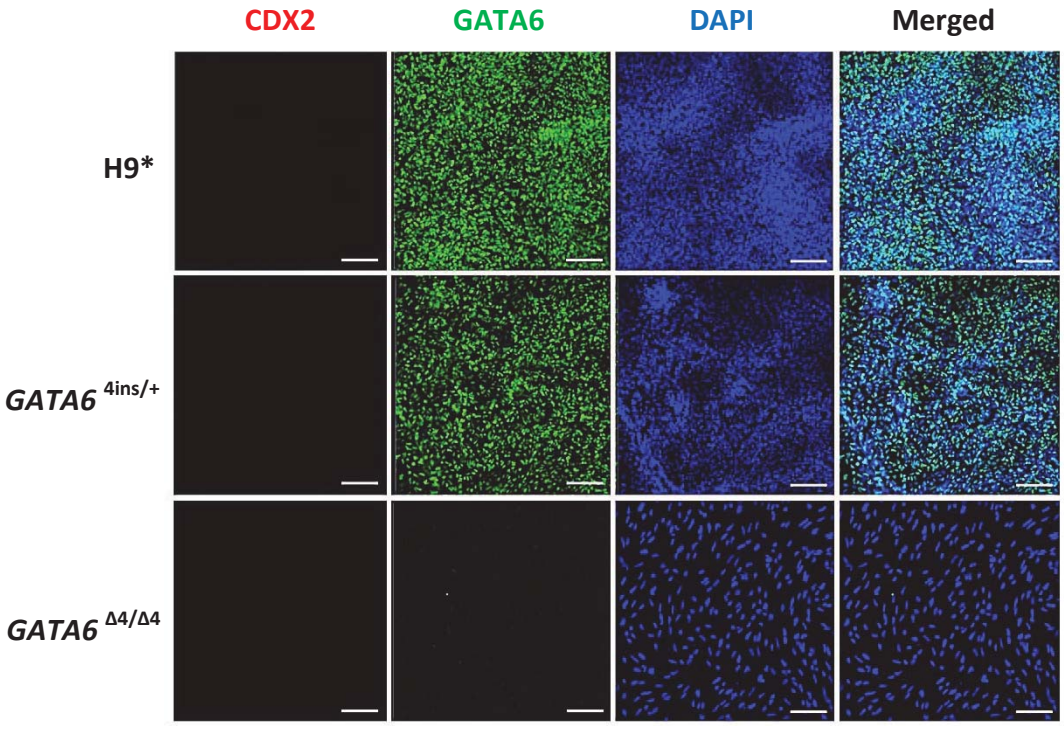


Figure 75. CDX2 remains unexpressed in all cells on day 6. Cells were grown in culture conditions that specified them toward the dorsal foregut lineage and analysed via immunofluorescence. Cells were fixed on day 6 and were stained for the hindgut marker CDX2. Scale bar, 100 µm.

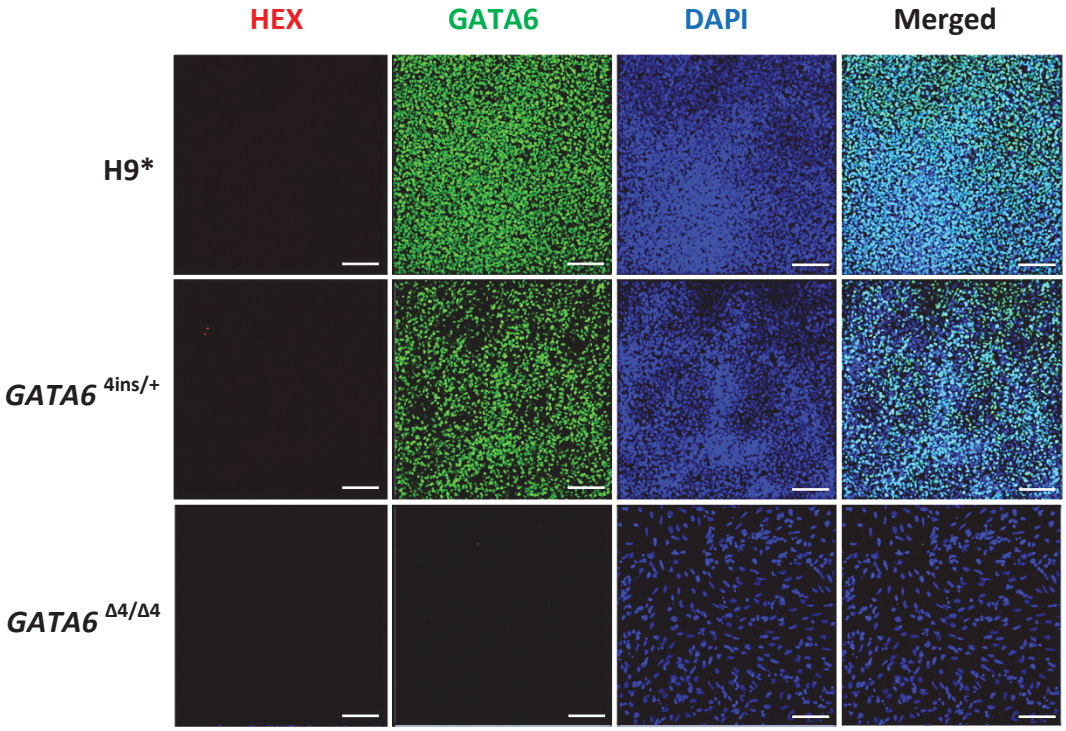


Figure 76. HEX remains unexpressed in all cells on day 6. Cells were grown in culture conditions that specified them toward the dorsal foregut lineage and analysed via immunofluorescence. Cells were fixed on day 6 and were stained for the ventral foregut marker HEX. Scale bar, 100 µm.

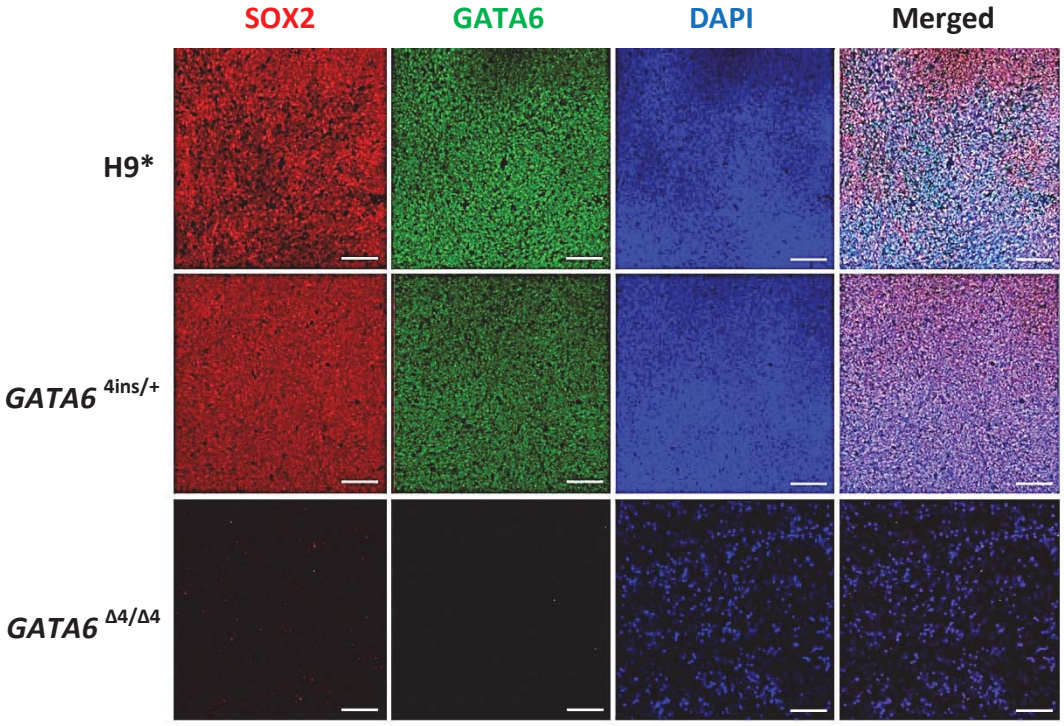


Figure 77. SOX2 expression is abolished in *GATA6* $^{\Delta4/\Delta4}$ cells on day 9. Cells were grown in culture conditions that specified them toward the foregut lineage and analysed via immunofluorescence. Cells were fixed on day 9 and were stained for the foregut marker SOX2. Scale bar, 100 µm.

As *GATA6* and *GATA4* have been shown to interact with each other (Charron et al., 1999), I looked at the mRNA levels of *GATA4* to investigate whether its regulation is affected by mutations in the *GATA6* gene. Notably, on day 3 at the transcriptional level, I observed that *GATA4* expression was decreased in H9-derived *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ mutant cells (Figure 68), suggesting that the GATA6 PPPs generated from these mutant alleles interact with *GATA4* to negatively impact its expression, leading to its down-regulation. This was in contrast to the *GATA4* levels observed in *GATA6* $^{GFP/+}$ and *GATA6* $^{GFP/GFP}$ cells, where *GATA4* levels remained relatively unchanged at day 3 (Figure 69). Nevertheless, a similar DE phenotype was seen with H9-derived *GATA6* $^{GFP/+}$ and *GATA6* $^{GFP/GFP}$ cells based on *SOX17* levels (Figure 69), suggesting that even though the PPPs seemed to decrease *GATA4* levels, the defect on DE development is independent of this decrease.

Interestingly, in FSPS13.B-derived heterozygous mutants, differing DE phenotypes were observed as compared to H9-derived heterozygous mutants. FACS analyses of CXCR4 on day 3 in *GATA6* $^{1ins/+}$, *GATA6* $^{\Delta14/+}$ and *GATA6* $^{\Delta21\_8ins/+}$ mutant cells showed a 40-60% decrease in CXCR4+ cells (Figure 67 and Figure 79), indicating a stronger defect in these cells compared to the respective H9 mutant (*GATA6* $^{4ins/+}$). qRT-PCR was consistent with this observation, showing an approximately 50% decrease of *SOX17* expression levels in *GATA6* $^{\Delta14/+}$ mutant cells (Figure 70). At the transcriptional level on day 3, similar to H9-derived *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ mutant cells, I also observed that *GATA4* expression was decreased in *GATA6* $^{\Delta14/+}$ and *GATA6* $^{\Delta14/\Delta11}$ cells (Figure 70).

In contrast, FACS analyses of the key DE markers CXCR4 and SOX17 on day 3 in FSPS13.B-derived heterozygous mutants *GATA6* $^{GFP/+}$ clones 1 and 2 both displayed no defects in DE formation (Figure 78 and Figure 79). This data was recapitulated by qRT-PCR analyses where expression levels of *SOX17*, *GATA6* and *GATA4* had no significant changes (Figure 70).

Consistent with H9-derived homozygous *GATA6* $^{\Delta4/\Delta4}$ and *GATA6* $^{GFP/GFP}$ mutant cells, FSPS13.B-derived homozygous *GATA6* $^{\Delta14/\Delta11}$ mutant cells failed to form the DE as shown by FACS analyses of CXCR4 at day 3 (Figure 67 and Figure 79). qRT-PCR analyses also showed almost complete ablation of *SOX17*, and strong decrease of *GATA6* and *GATA4* (Figure 70).

Patient A (clones 1-3) and Patient B (clones 1-3) displayed a similar DE phenotype to FSPS13.B-derived heterozygous *GATA6* $^{1ins/+}$, *GATA6* $^{\Delta14/+}$ and *GATA6* $^{\Delta21\_8ins/+}$ mutants cells, where FACS analyses of the key DE marker CXCR4 at day 3 showed a 50-60% decrease in CXCR4+ cells (Figure 80, Figure 81, Figure 82 and Figure 83). qRT-PCR analyses also showed a strong down-regulation of DE markers *SOX17* and *CXCR4* in both patient lines (Figure 84). *GATA4* levels, and to a lesser extent *GATA6*, in both patient lines were also down-regulated (Figure 84).

Taken together, these finding suggest that complete loss of *GATA6* in TALEN-edited hPSCs significantly perturbs the gene regulatory network (GRN) required for DE specification, resulting in the failure of DE formation. Heterozygous loss of *GATA6* seems to impair DE formation with a varying penetrance of phenotype from 20% in H9-derived heterozygous mutants, to 40-60% in FSPS13.B-derived heterozygous mutants and Patients A and B, and to no impairment in FSPS13.B-derived *GATA6* $^{GFP/+}$ mutants (summarised in Figure 85).

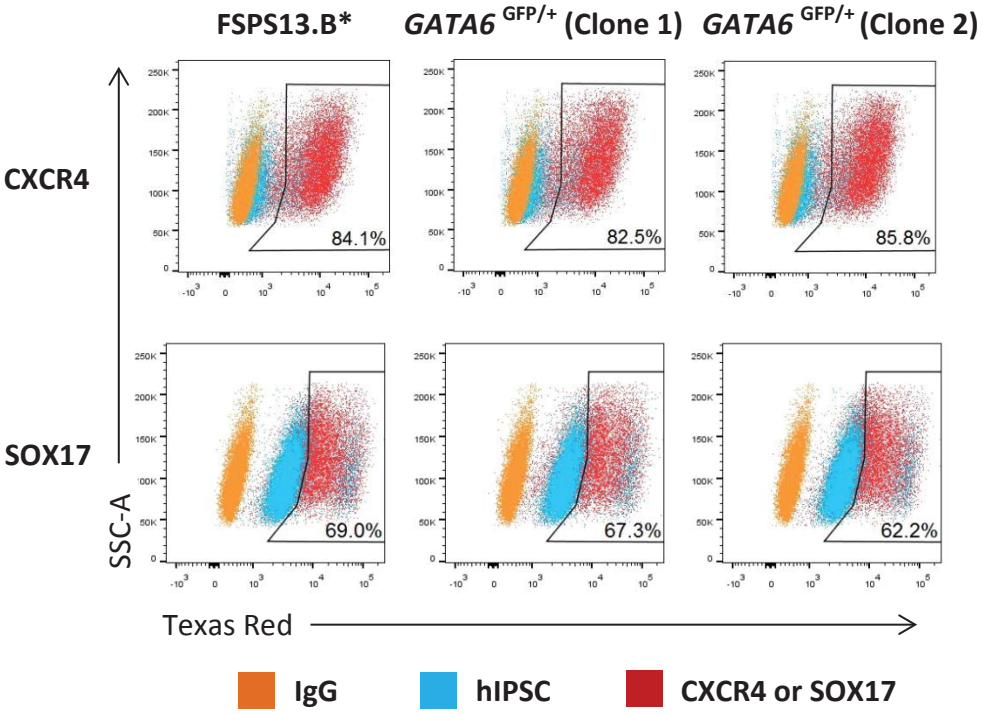


Figure 78. The number of CXCR4+ and SOX17+ cells is not decreased in both clones 1 and 2 of *GATA6* $^{GFP/+}$ FSPS13.B cells. Cells were fixed on day 3 and were stained for the DE markers SOX17 and CXCR4. Data show results of one experiment that is representative of at least 2 independent experiments.

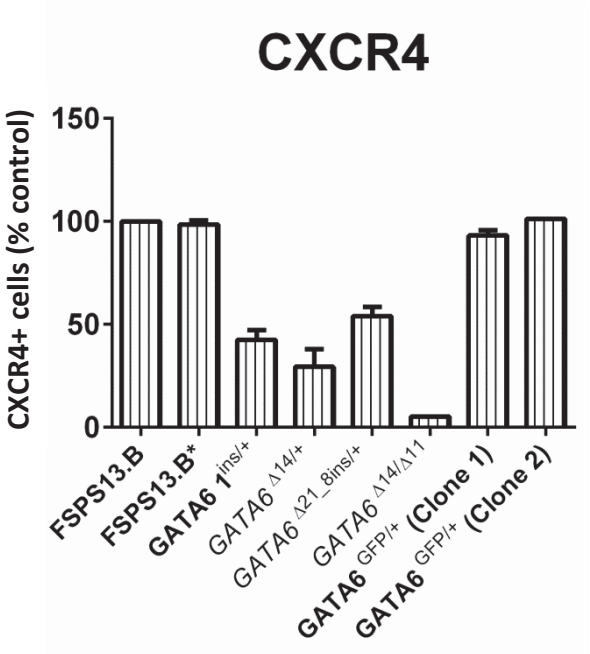


Figure 79. Summary of CXCR4 levels via FACS for all FPSP13.B-derived mutant cells on day 3. Wild-type FSPS13.B cells are normalised to 100% and the relative CXCR4 expression levels of FSPS13.B* or mutant cells are shown. Data show results of two experiments and error bars indicate standard deviation.

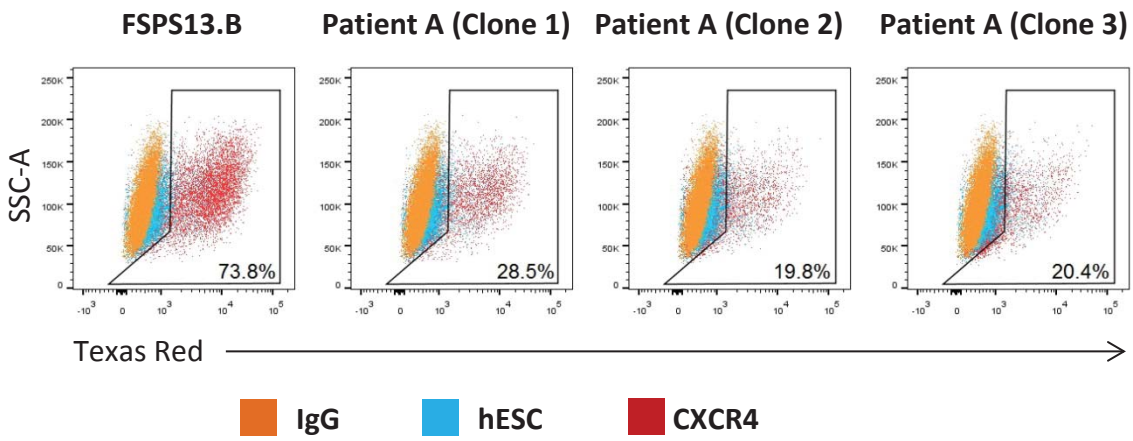


Figure 80. The number of CXCR4+ cells is decreased in Patient A. Cells were fixed on day 3 and were stained for the DE marker CXCR4. Data show results of one experiment that is representative of at least 3 independent experiments.

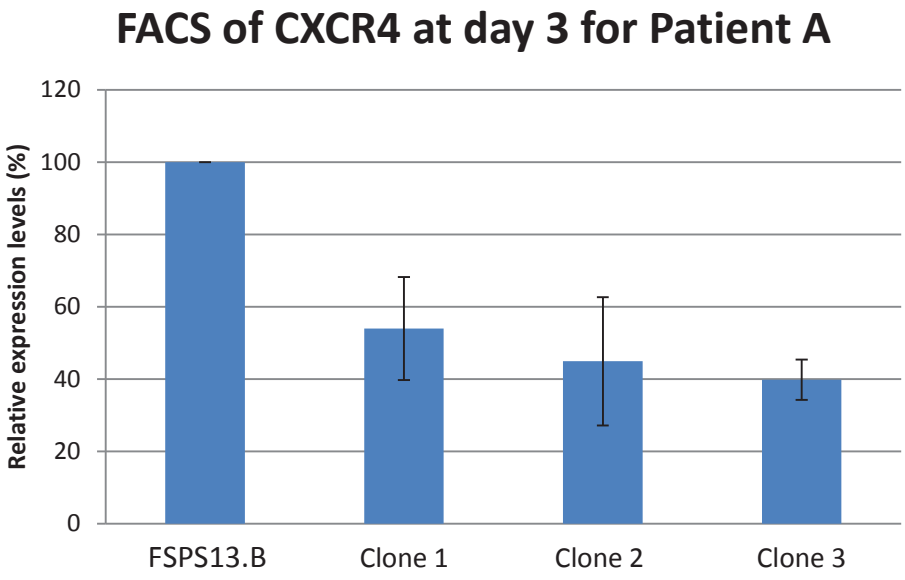


Figure 81. Summary of CXCR4 levels via FACS for Patient A cells on day 3. Wild-type FSPS13.B cells are normalised to 100% and the relative CXCR4 expression levels of clones 1, 2 and 3 of Patient A are shown. Data show results of two experiments and error bars indicate standard deviation.

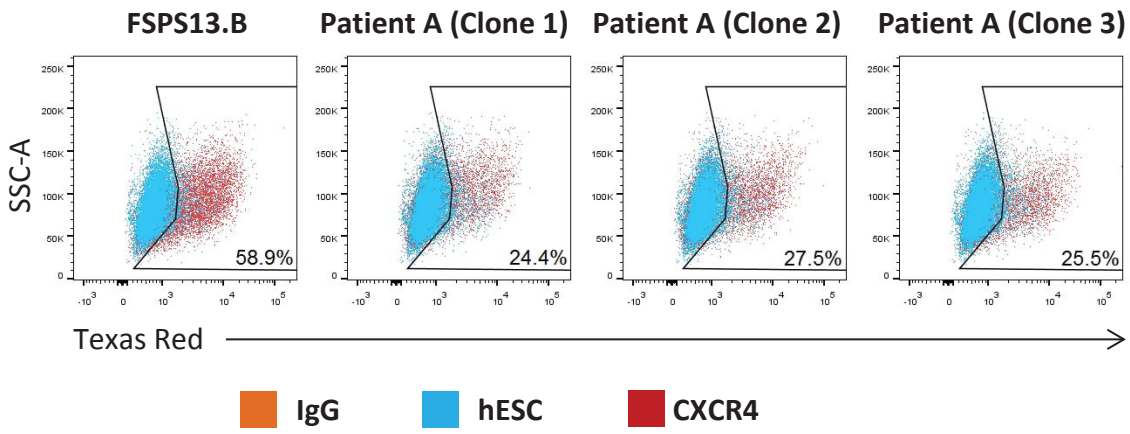


Figure 82. The number of CXCR4+ cells is decreased in Patient B. Cells were fixed on day 3 and were stained for the DE marker CXCR4. Data show results of one experiment.

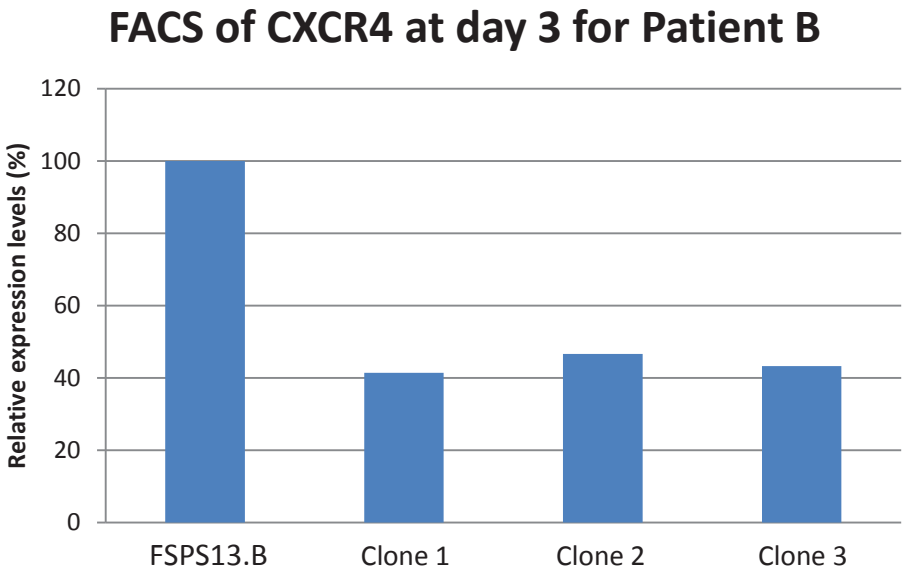


Figure 83. Summary of CXCR4 levels via FACS for Patient B cells on day 3. Wild-type FSPS13.B cells are normalised to 100% and the relative CXCR4 expression levels of clones 1, 2 and 3 of Patient B are shown. Data show results of one experiment.

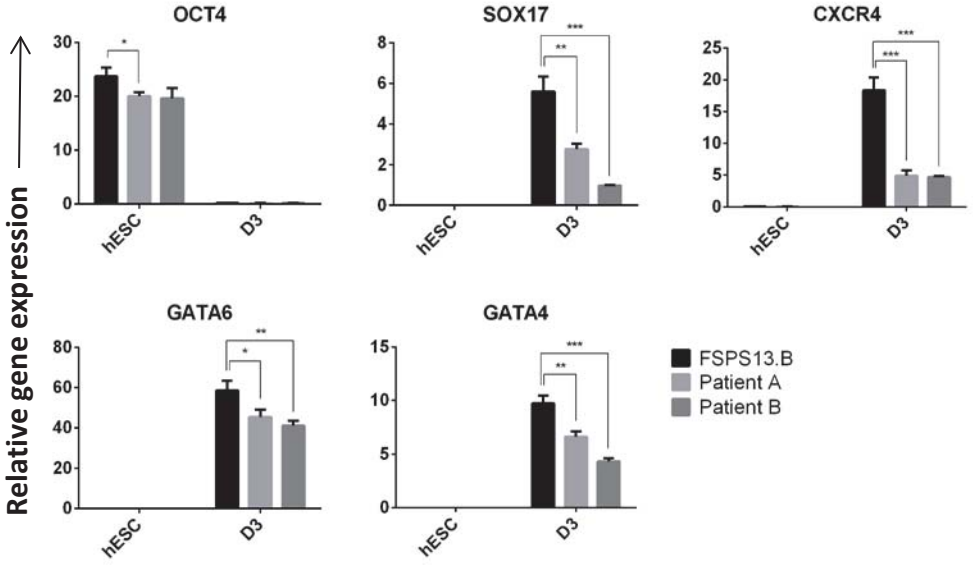


Figure 84. qRT-PCR analyses of Patient A and B cells on day 3. RNA was extracted on day 3 and the expression patterns of key endoderm markers were determined. Data are triplicate samples of one experiment and representative of two independent experiments (Patient A) and one experiment (Patient B). Error bars indicate standard deviation.

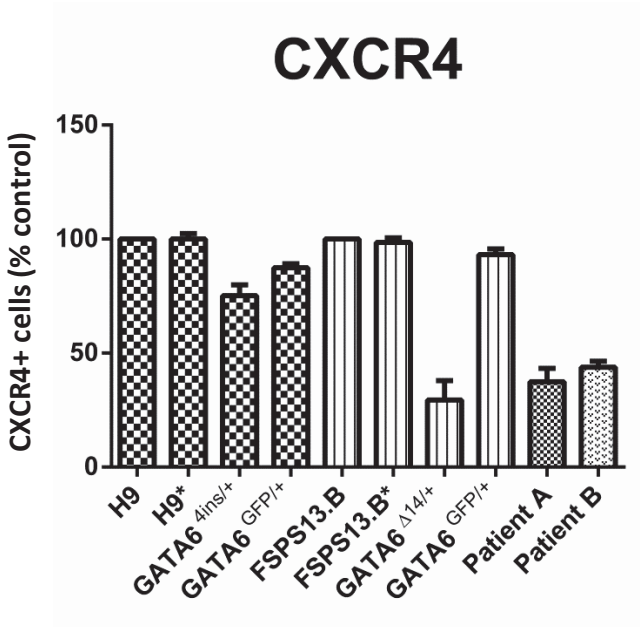


Figure 85. Summary of CXCR4 levels via FACS for H9 and FSPS13.B selected heterozygous mutant cells, and Patients A and B on day 3. H9 and FSPS13.B cells are normalised to 100% and the relative CXCR4 expression levels of their respective mutant cell lines are shown. Data show results of three independent experiments and error bars indicate standard deviation.

3.3.4. *GATA6* heterozygous mutants display similar endodermal defects using a commercial kit from STEMCELL Technologies

I further validated these results using the STEMdiff Definitive Endoderm kit from Stem Cell Technologies (SCT) using H9*, H9-derived *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ lines. A similar endoderm defect was observed in the H9-derived *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ cells using this protocol as compared to the lab protocol. FACS analyses at day 3 showed that SOX17+ and CXCR4+ cells were decreased by approximately 30-50% in *GATA6* $^{4ins/+}$ cells and 90% in *GATA6* $^{\Delta4/\Delta4}$ cells (Figure 86). This was consistently seen in qRT-PCR analyses (Figure 87). Expression patterns of *OCT4*, *GATA6* and *GATA4* on days 3 and 6 also resembled that of the lab's in-house protocol (Figure 87). These results using a commercially available kit align well with those presented earlier in this chapter and give me confidence in my conclusion that GATA6 haploinsufficiency indeed impacts early formation of the DE lineage.

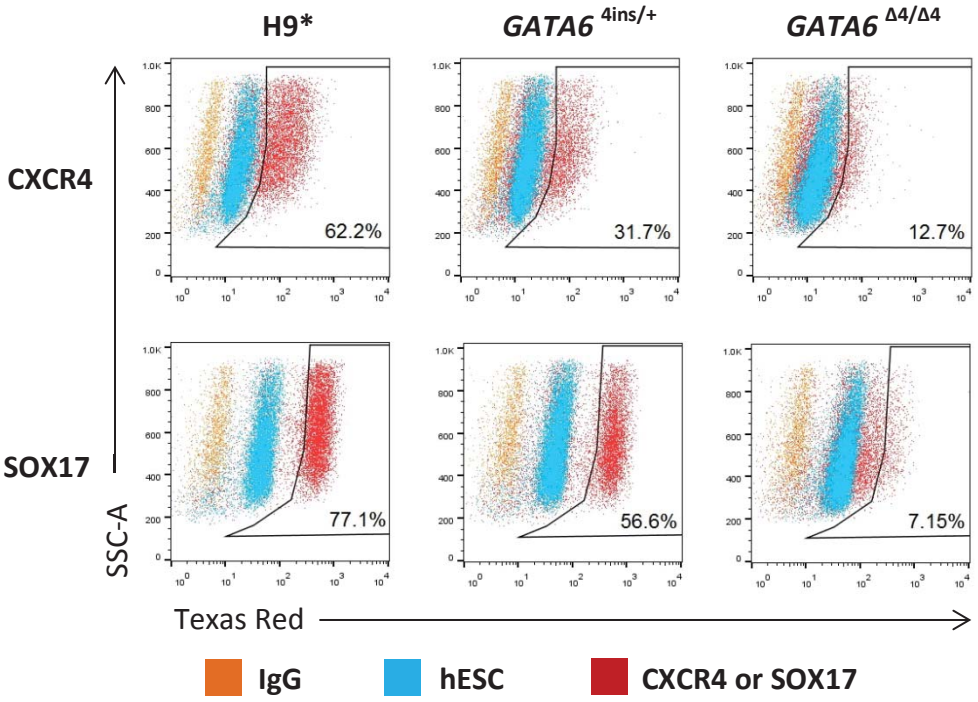


Figure 86. The number of CXCR4+ and SOX17+ cells is decreased in *GATA6* $^{4ins/+}$ cells and are almost completely absent in *GATA6* $^{\Delta4/\Delta4}$ H9 cells differentiated via STEMCELL Technologies kit. Cells were fixed on day 3 and were stained for the DE markers SOX17 and CXCR4. Data show results of one experiment that is representative of at least two independent experiments.

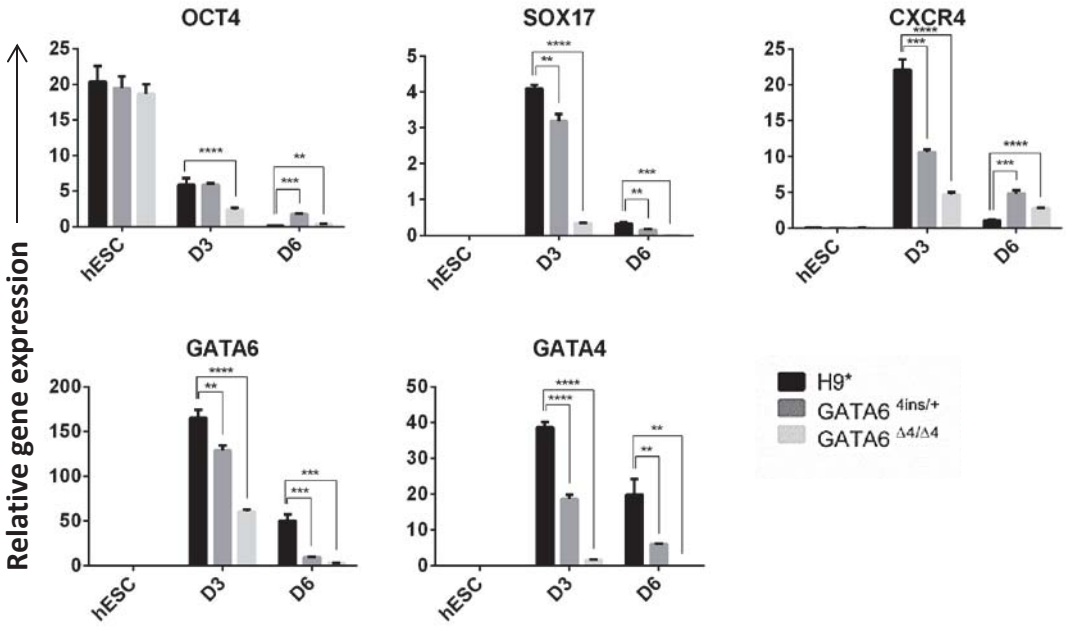


Figure 87. qRT-PCR analyses of H9*, *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ cells on days 3 and 6 differentiated via STEMCELL Technologies kit . RNA was extracted at specific stages and the expression patterns of key markers were determined. Data are triplicate samples of one experiment and representative of three independent experiments. Error bars indicate standard deviation.

3.3.5. *GATA6* heterozygous mutants did not exhibit endodermal defects using PSC Definitive Endoderm Induction Kit from Life Technologies

In addition to the commercial kit from SCT, I also performed DE differentiation using PSC Definitive Endoderm Induction Kit from Life Technologies. H9 and FPSP13.B cells were differentiated alongside the TALEN-derived and Patient B lines as a control for DE differentiation efficiency compared to the lab and the STEMdiff Definitive Endoderm kit from SCT. From FACS analyses, CXCR4+ cells were consistently over 80% in both H9 and FSPS13.B cells, suggesting a slightly more efficient DE differentiation than the lab and SCT protocols (Figure 88).

TALEN-derived FSPS13.B*, *GATA6* $^{\Delta 14/+}$ and *GATA6* $^{GFP/+}$, *GATA6* $^{\Delta 14/\Delta 11}$ cell lines and Patient B (clone 1) were differentiated using the PSC DE induction kit. Interestingly, FACS analyses revealed no endodermal defect in the *GATA6* $^{\Delta 14/+}$ and *GATA6* $^{GFP/+}$ and Patient B heterozygous mutant cell lines (Figure 88). qRT-PCR analyses recapitulate results from FACS, except for Patient B (Figure 89). However, qRT-PCR was only performed once, whereas FACS was performed twice. So, it is possible that there could have been an error with the qRT-PCR result of Patient B. FACS and qRT-PCR results for *GATA6* $^{\Delta 14/\Delta 11}$ cell line was consistent with both lab and SCT protocols, with an approximately 90% decrease of CXCR4+ cells (Figure 88 and Figure 89).

I attempted to pursue this inconsistency between protocols further by requesting to know the components of the PSC DE Kit, but Life Technologies was unwilling to share the formulation of their kit. Hence, I was unable to compare the components between the protocols and determine which growth factor(s) and/or pathway inhibitors could have attributed to this discrepancy in the results. Taken together, the results from the PSC DE kit suggest that the impaired DE phenotype seen in the heterozygous *GATA6* mutant lines could be attributed to a protocol-dependent defect.

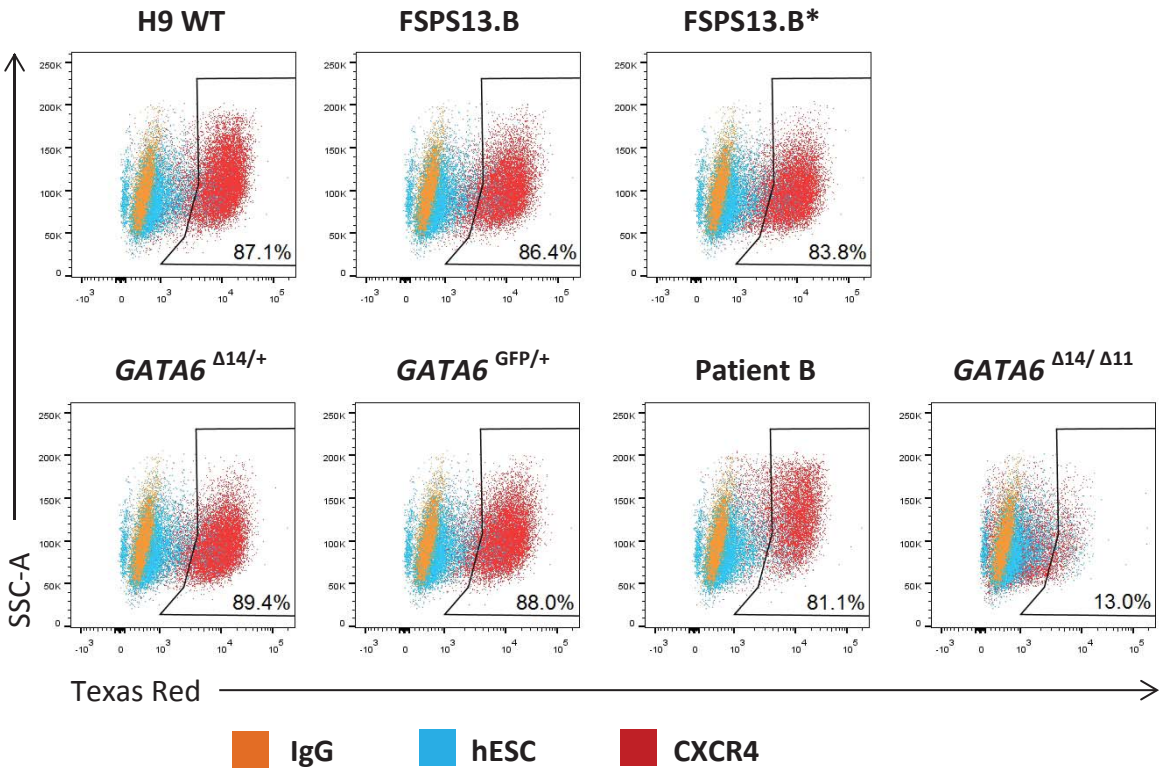


Figure 88. The number of CXCR4+ cells is not decreased in *GATA6* heterozygous FSPS13.B mutant cells and Patient B differentiated via PSC Definitive Endoderm Induction Kit from Life Technologies. Cells were fixed on day 3 and were stained for the DE marker CXCR4. Gates were set according to hESC. Data show results of one experiment that is representative of two independent experiments.

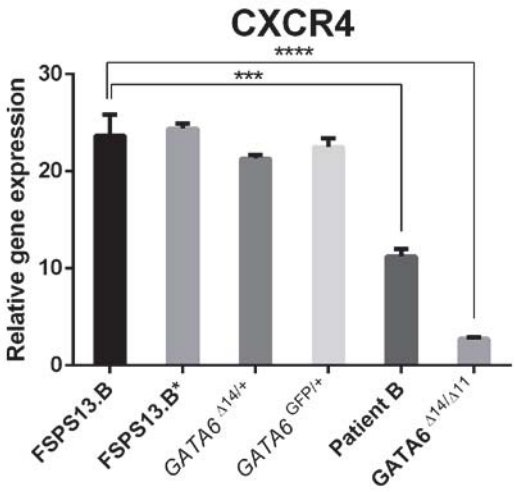


Figure 89. qRT-PCR analyses of FSPS13.B, FPSP13.B*, FSPS13.B-derived mutant cells and Patient B on day 3 differentiated via PSC Definitive Endoderm Induction Kit from Life Technologies. RNA was extracted on day 3 and the expression of *CXCR4* was determined. Data are triplicate samples of one experiment, error bars indicate standard deviation.

3.4. *GATA6* is required for differentiation into the pancreatic lineage

To further study the effects of heterozygous and homozygous loss of *GATA6* during human pancreatic development, I continued the differentiation process with the mutant lines in order to characterise their later phenotypes during acquisition of pancreatic identity (pancreatic endoderm; PE on day 12) and allocation to the endocrine lineage (endocrine progenitors; EP on day 24).

3.4.1. Homozygous *GATA6* mutants fail to enter the pancreatic lineage

To characterise the phenotype of homozygous loss of *GATA6* using our *in vitro* model system, I differentiated H9 TALEN-derived *GATA6* $^{\Delta4/\Delta4}$ and *GATA6* $^{GFP/GFP}$ mutant cells toward the EP stage. Expectedly, ICC analyses of *GATA6* $^{\Delta4/\Delta4}$ cells showed the absence of key markers PDX1 at day 12 (Figure 90), NGN3 at day 15 (Figure 91) and C-PEPTIDE, SST and GCG at day 24 (Figure 92), indicating that the *GATA6* $^{\Delta4/\Delta4}$ mutant entirely failed to enter the pancreatic lineage.

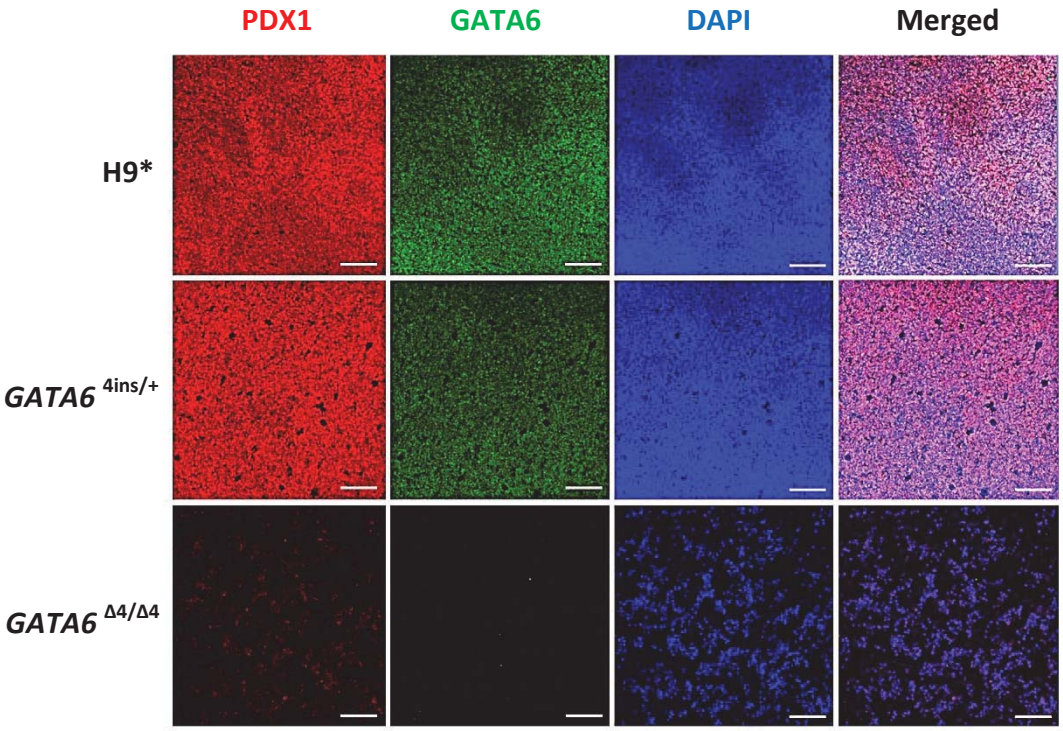


Figure 90. PDX1 is not activated in *GATA6* $^{\Delta4/\Delta4}$ cells by day 12 of differentiation. Cells were grown in culture conditions that specified them toward the pancreatic progenitor lineage and analysed via immunofluorescence. Cells were fixed on day 12 and were stained for the pancreatic marker PDX1. Scale bar, 100 µm.

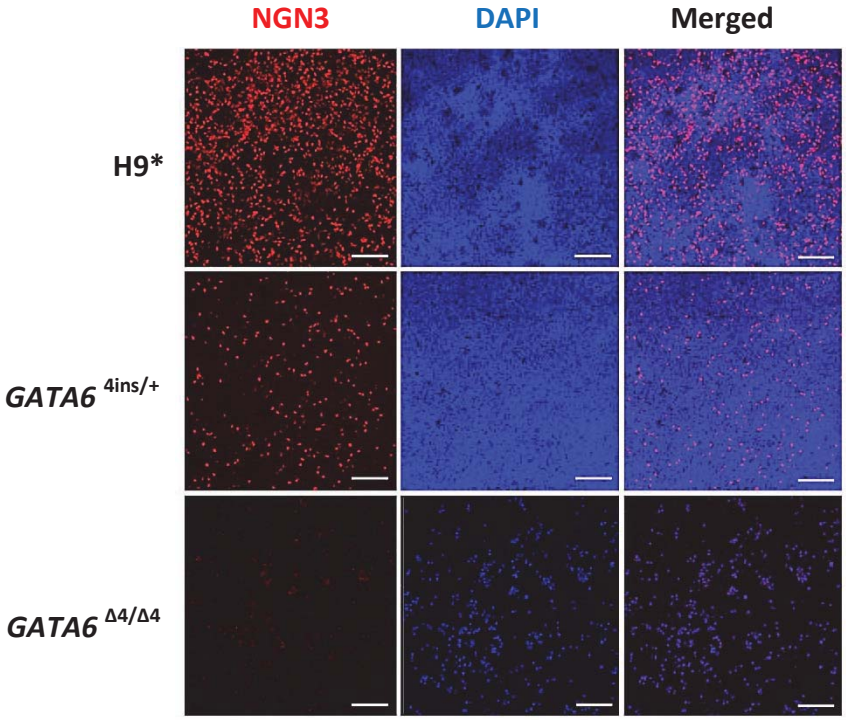


Figure 91. NGN3 is not activated in *GATA6* $^{\Delta4/\Delta4}$ cells by day 15 of differentiation. Cells were grown in culture conditions that specified them toward the endocrine lineage and analysed via immunofluorescence. Cells were fixed on day 15 and were stained for the key endocrine marker NGN3. Scale bar, 100 µm.

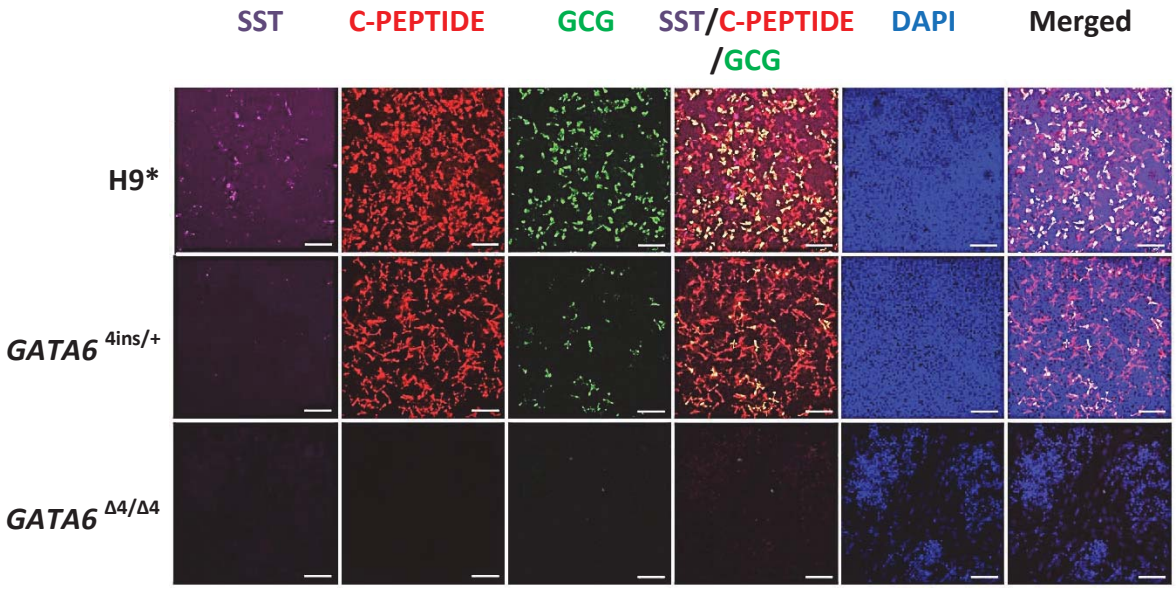


Figure 92. C-PEPTIDE, SST and GCG expression are not activated in *GATA6* $^{\Delta4/\Delta4}$ cells by day 24 of differentiation. Cells were grown in culture conditions that mature into endocrine progenitors and analysed via immunofluorescence. Cells were fixed on day 24 and were stained for key markers C-PEPTIDE, SST and GCG. Scale bar, 100 µm.

FACS analyses of *GATA6* $^{\Delta4/\Delta4}$ and *GATA6* $^{GFP/GFP}$ mutant cells on day 12 showed a 90-100% loss of PDX1+ cells compared to H9 wild-type and H9* cells (Figure 93 and Figure 94). On day 24, C-PEPTIDE levels in *GATA6* $^{GFP/GFP}$ cells were almost negligible (Figure 95 and Figure 96). In some experiments, *GATA6* $^{\Delta4/\Delta4}$ cells did not survive up to day 24 and died around day 18 (Figure 95).

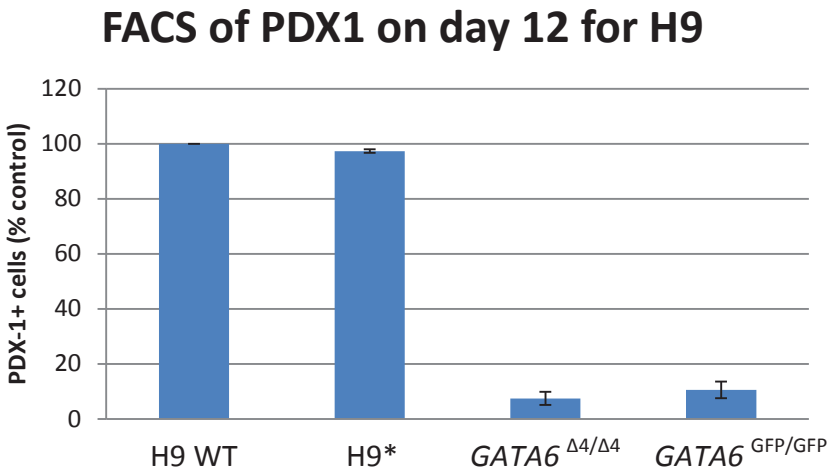


Figure 93. Summary of PDX1 levels via FACS for H9 homozygous mutant cells on day 12. Wild-type H9 cells are normalised to 100% and the relative PDX1 levels of H9*, H9-derived *GATA6* $^{\Delta4/\Delta4}$ and *GATA6* $^{GFP/GFP}$ mutant cells are shown. Data show results of two independent experiments and error bars indicate standard deviation.

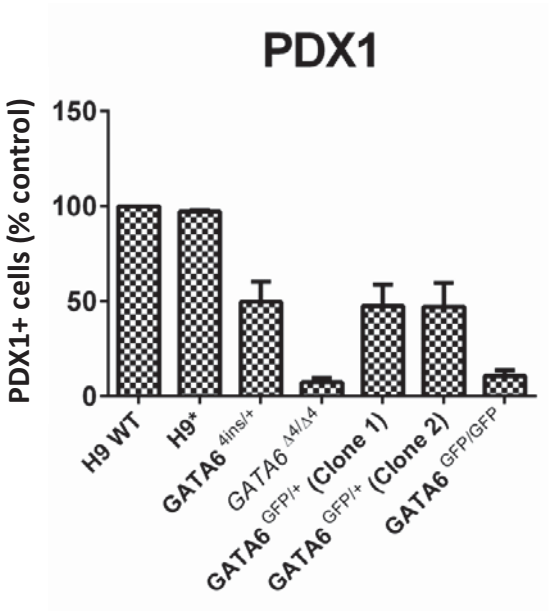


Figure 94. Summary of PDX1 levels via FACS for all H9-derived mutants cells on day 12. Wild-type H9 cells are normalised to 100% and the relative PDX1 levels of H9* or mutant cells are shown. Data show results of two experiments and error bars indicate standard deviation.

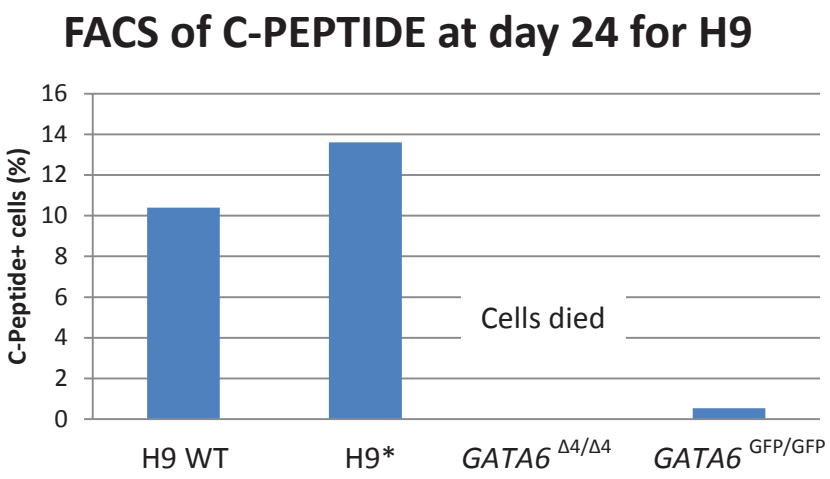


Figure 95. Summary of C-PEPTIDE levels via FACS for H9 homozygous mutant cells on day 24. Absolute percentage of C-PEPTIDE-positive cells in H9 wild-type, H9* and H9-derived *GATA6* $^{\Delta4/\Delta4}$ and *GATA6* $^{GFP/GFP}$ mutant cells are shown. Each bar represents one biological sample, and the graph was taken from one experiment, which is representative of two independent experiments.

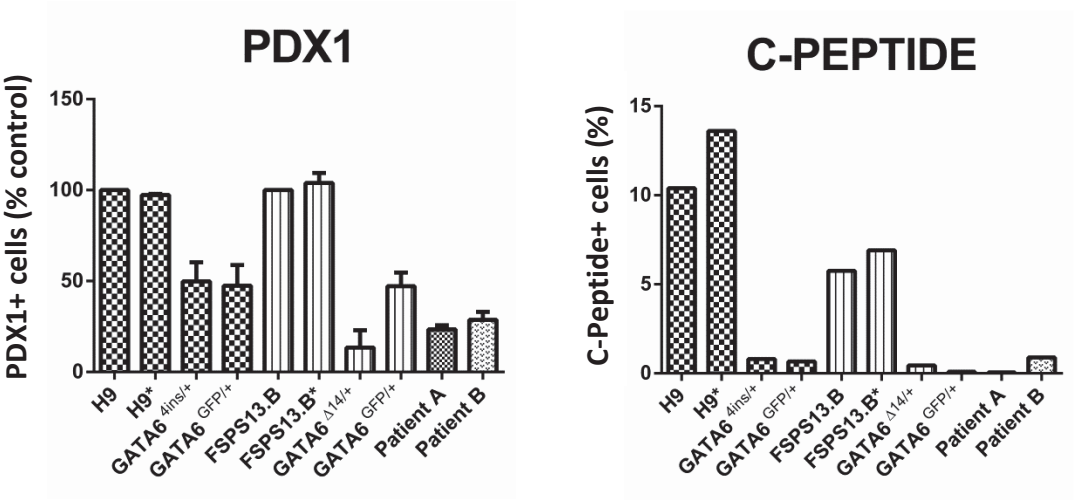


Figure 96. Summary of PDX1 (day 12) and C-PEPTIDE (day 24) expression via FACS for H9 and FSPS13.B selected heterozygous mutant cells, and Patients A and B. For PDX1, H9 and FSPS13.B cells are normalised to 100% and the relative PDX1 levels of their respective mutant cell lines are shown. Data show results of three independent experiments and error bars indicate standard deviation. For C-PEPTIDE, absolute percentages of C-PEPTIDE-positive cells in all cell lines are shown. Each bar represents one biological sample, and the graph was taken from one experiment, which is representative of three independent experiments.

3.4.2. Heterozygous *GATA6* mutants elicit a pancreatic defect in all protocols

Next, to characterise the phenotype of heterozygous loss of *GATA6* using our *in vitro* model system, I first differentiated H9 TALEN-derived *GATA6* $^{4ins/+}$ and clones 1 and 2 of *GATA6* $^{GFP/+}$ mutant cells via the lab-derived protocol toward the EP stage. FACS analyses of key PE marker PDX1 at day 12 for *GATA6* $^{4ins/+}$ and clones 1 and 2 of *GATA6* $^{GFP/+}$ mutant cells indicate a decrease in PDX1+ cells by about 50% in both lines (Figure 94 and Figure 96), even though *GATA6* $^{GFP/+}$ seemed to have lower levels of *PDX1* (Figure 97). In contrast, ICC analyses of *GATA6* $^{4ins/+}$ cells did not show an observable difference in PDX1 expression levels compared to H9* cells (Figure 90). This observation could be due to the fact that the cells analysed by ICC are grown very densely, resulting in overlapping growth that may mask PDX1-negative cells.

On day 24, C-PEPTIDE levels by FACS were almost negligible in both *GATA6* $^{4ins/+}$ and *GATA6* $^{GFP/+}$ mutant lines (Figure 96). Similar to *PDX1*, *GATA6* $^{GFP/+}$ have lower levels of *Insulin* mRNA compared to *GATA6* $^{4ins/+}$ (Figure 97). qRT-PCR analyses of key genes showed a similar down-regulation of *HNF4A* and *GATA4* in both *GATA6* $^{4ins/+}$ and *GATA6* $^{GFP/+}$ mutant lines at day 24 (Figure 97). *HLXB9*, on the other hand, was strongly decreased in *GATA6* $^{4ins/+}$ mutant cells but not in *GATA6* $^{GFP/+}$ mutant cells at day 24 (Figure 97).

In addition, I performed C-peptide ELISA on H9*, *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ mutant cells. Similar to the H9 wild-type cells that was described earlier in Chapter 3.1.1, H9* elicited an inverse C-peptide release response upon glucose stimulation (Figure 98). The levels of C-Peptide stimulation and down-regulation in H9*is very similar to H9 wild-type, suggesting the lack of off target effects from the TALEN targeting, if any. In contrast, *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ mutant lines did not elicit any glucose-stimulated C-peptide release, and the basal C-Peptide levels these lines are much lower than H9 wild-type and H9* (Figure 98).

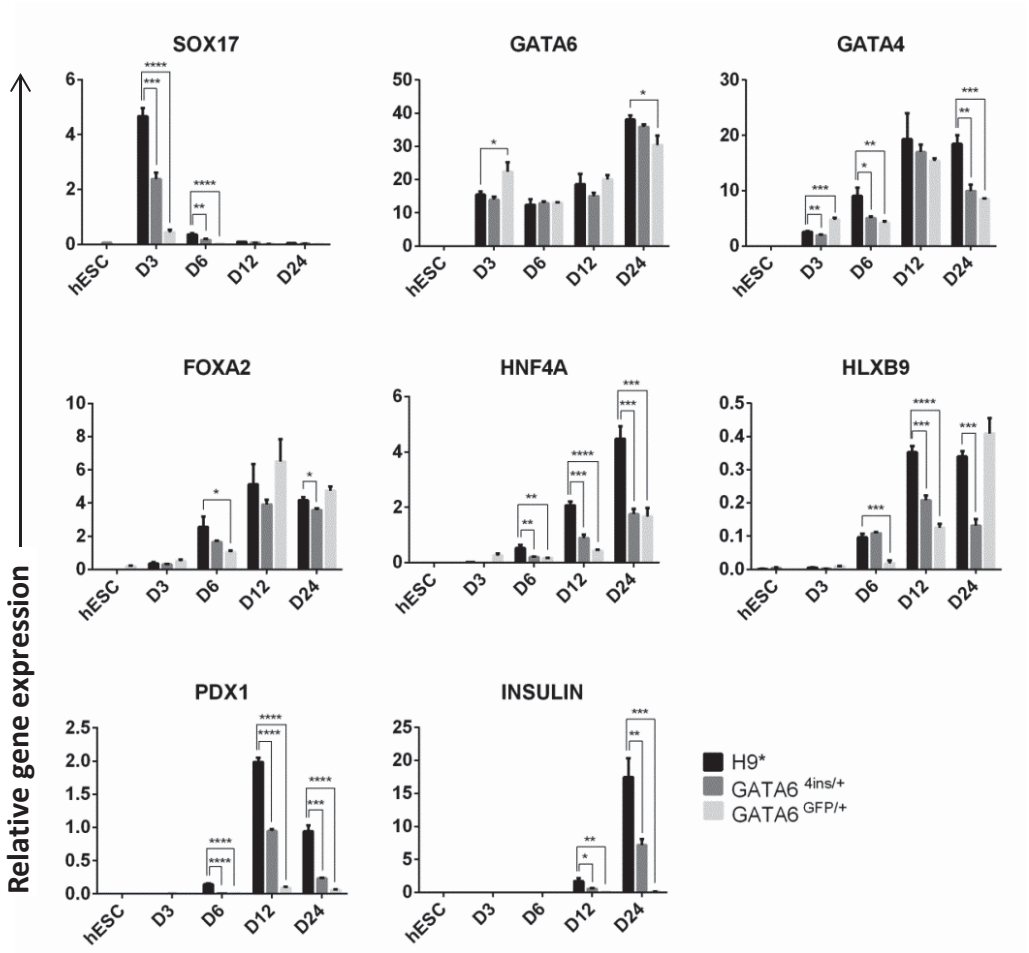


Figure 97. qRT-PCR analyses of H9*, *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ cells on days 3, 6, 12 and 24. RNA was extracted at specific stages and the expression patterns of key markers were determined. Data are triplicate samples of one experiment and representative of three independent experiments. Error bars indicate standard deviation.

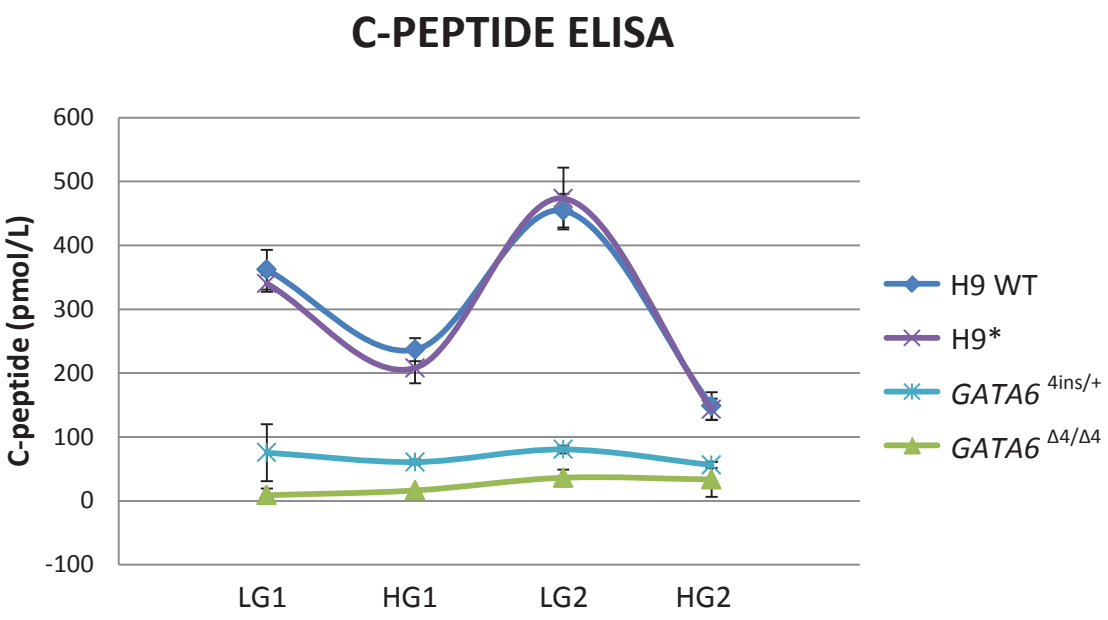


Figure 98. C-peptide secretion of TALEN-derived H9 mutants upon glucose stimulation on day 24. Cells were differentiated using the lab-derived protocol and assayed on day 24. Data are presented as the average of 3 biological replicates of one experiment and representative of two independent experiments. Error bars indicate standard deviation. LG1 is first incubation of low glucose; HG1 is first incubation of high glucose; LG2 is second incubation of low glucose; HG2 is second incubation of high glucose.

In FSPS13.B TALEN-derived *GATA6* $^{\Delta 14/+}$ and *GATA6* $^{GFP/+}$ mutant cells, *PDX1* levels were also strongly down-regulated on days 12 and 24 by qRT-PCR (Figure 99). FACS analyses showed an approximately 80-90% decrease of PDX1+ cells in *GATA6* $^{\Delta 14/+}$ mutant cells and 50% decrease of PDX1+ cells in *GATA6* $^{GFP/+}$ mutant cells on day 12 (Figure 96). PDX1 FACS analyses for all the other heterozygous mutant lines are shown in Figure 100. qRT-PCR analyses of key genes showed down-regulation of *HNF4A* and *Insulin* on days 12 and 24, but no change for *HLXB9* (Figure 99). *GATA4* expression was down-regulated in *GATA6* $^{\Delta 14/+}$ mutant cells at all time points but remained relatively unchanged in *GATA6* $^{GFP/+}$ mutant cells (Figure 99). On day 24, C-PEPTIDE+ cells in both *GATA6* $^{\Delta 14/+}$ and *GATA6* $^{GFP/+}$ mutant lines numbered near zero (Figure 96).

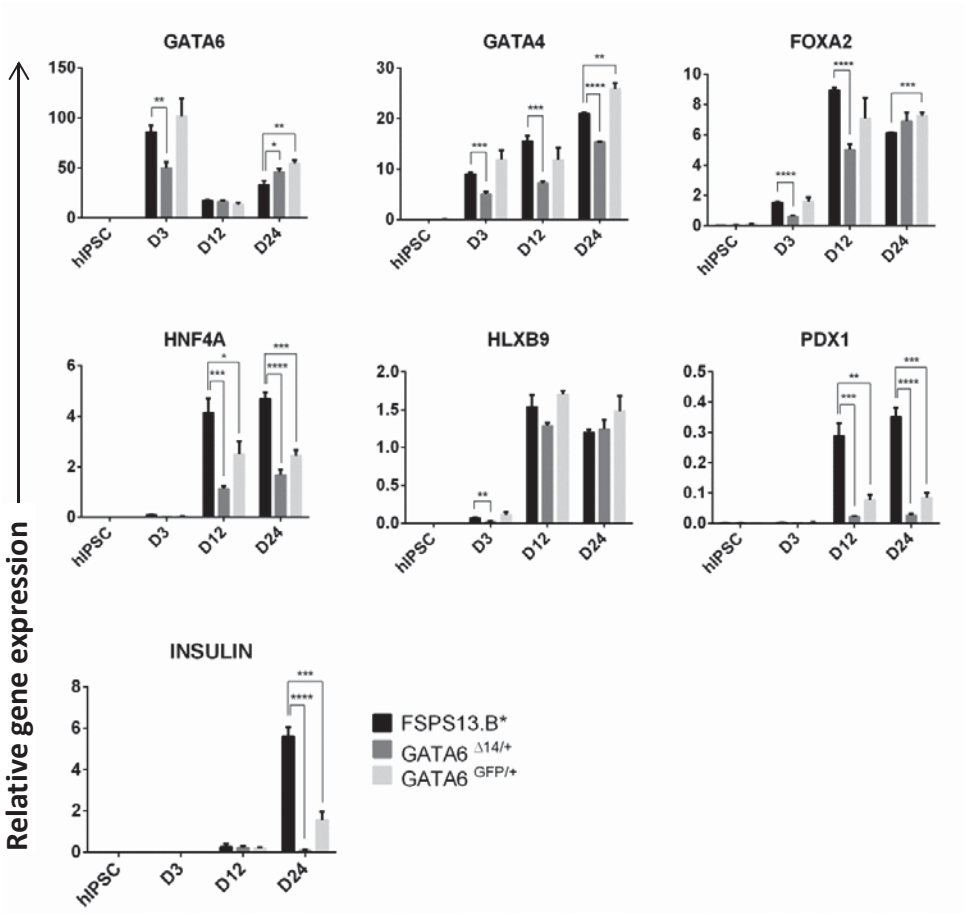


Figure 99. qRT-PCR analyses of FSPS13.B*, *GATA6* $^{\Delta 14/+}$ and *GATA6* $^{GFP/+}$ cells on days 3, 12 and 24. RNA was extracted at specific stages and the expression patterns of key markers were determined. Data are triplicate samples of one experiment and representative of three independent experiments. Error bars indicate standard deviation.

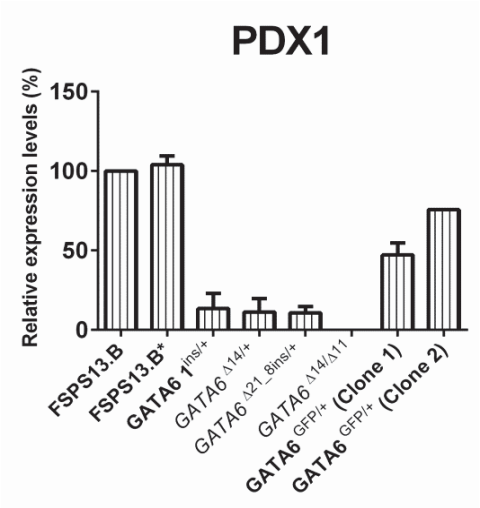


Figure 100. Summary of *CXCR4* levels via FACS for all FPSP13.B-derived mutant cells on day 12. Wild-type FSPS13.B cells are normalised to 100% and the relative *CXCR4* levels of FSPS13.B* or mutant cells are shown. Data show results of two experiments and error bars indicate standard deviation.

In Patients A and B, *PDX1* levels were also strongly down-regulated on days 12 and 24, as shown by qRT-PCR (Figure 101). FACS analyses showed an approximately 70-80% decrease of PDX1+ cells in both Patients A and B lines on day 12 (Figure 96). qRT-PCR analyses of key genes showed down-regulation of *GATA4*, *FOXA2*, *HNF4A*, *HLXB9* and *Insulin* on days 12 and 24 (Figure 101). On day 24, FACS of C-PEPTIDE+ cells were near zero in both Patient A and B lines (Figure 96).

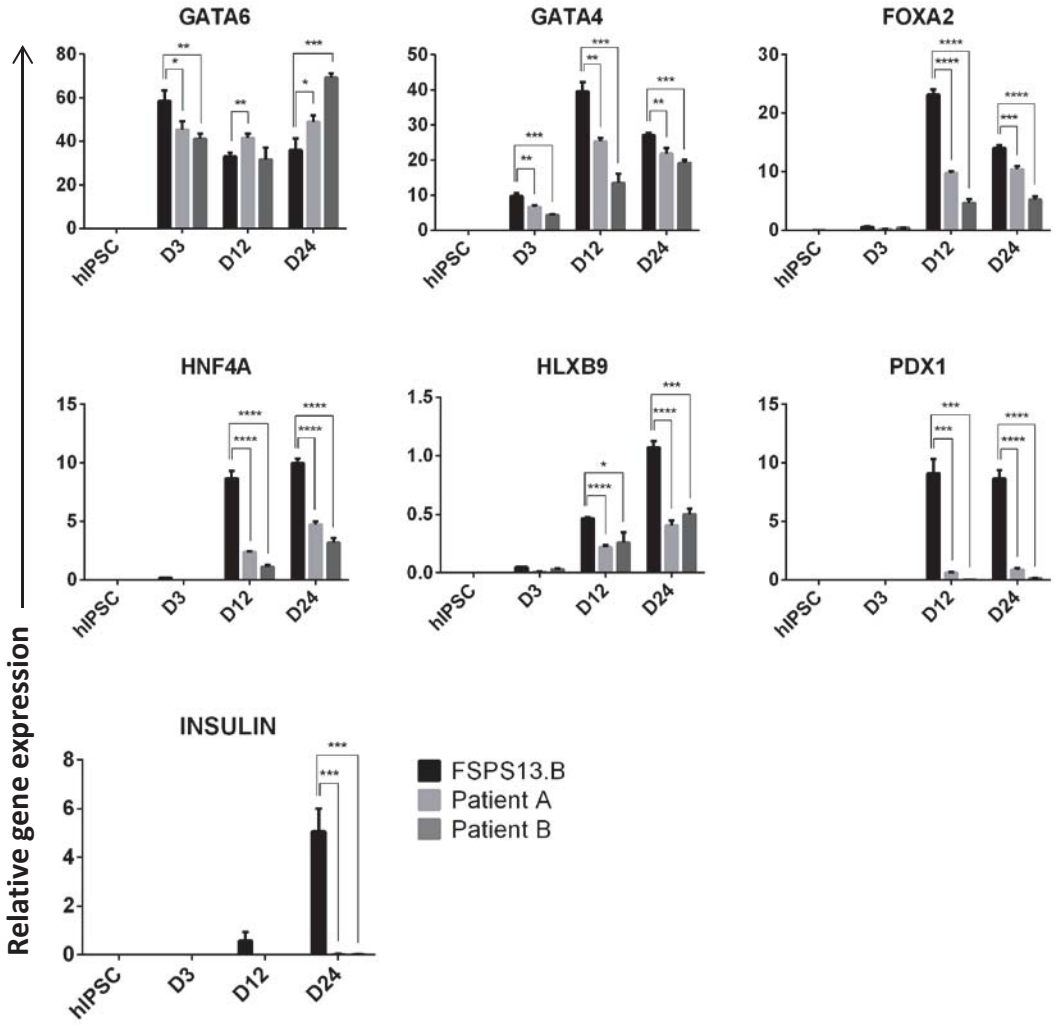


Figure 101. qRT-PCR analyses of FSPS13.B, Patient A and Patient B cells on days 3, 12 and 24. RNA was extracted at specific stages and the expression patterns of key markers were determined. Data are triplicate samples of one experiment and representative of three independent experiments. Error bars indicate standard deviation.

Using the STEMdiff pancreatic progenitor kit from SCT, I differentiated H9*, H9-derived *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ lines toward the PE lineage. A similar pancreatic defect was observed in the H9-derived *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ cells using this protocol as compared to the lab protocol. FACS analyses on day 12 showed that PDX1+ cells were decreased by approximately 50% in *GATA6* $^{4ins/+}$ cells and absent in *GATA6* $^{\Delta4/\Delta4}$ cells (Figure 102). This was consistently seen in qRT-PCR analyses where key markers such as *GATA4*, *HNF4A*, *PDX1* and *NKX6-1* were strongly down-regulated in *GATA6* $^{4ins/+}$ mutant cells and absent in *GATA6* $^{\Delta4/\Delta4}$ cells (Figure 103). These results complement the results derived from the lab protocol.

Using the PSC Definitive Endoderm Induction Kit from Life Technologies, I differentiated FSPS13.B* and FSPS13.B-derived mutants *GATA6* $^{\Delta14/+}$, *GATA6* $^{GFP/+}$ and *GATA6* $^{\Delta14/\Delta11}$ mutant cell lines and Patient B toward the PE lineage. Interestingly, despite the heterozygous mutants showing no impairment of DE formation (described in Chapter 3.3.5), a very similar pancreatic defect was observed in the heterozygous mutants using this protocol as compared to the lab protocol. FACS analyses on day 12 showed that PDX1+ cells were decreased by approximately 80-90% in *GATA6* $^{\Delta14/+}$cells and Patient B cells, 60% in *GATA6* $^{GFP/+}$ cells and is almost completely absent in *GATA6* $^{\Delta14/\Delta11}$ cells (Figure 104). qRT-PCR analysis focusing on *PDX1* on day 12 confirms the FACS data (Figure 105).

Taken together, these finding confirm that complete loss of *GATA6* in TALEN-edited hPSCs results in failure of pancreatic development, which is most likely explained by the disruption of the GRN required for earlier DE specification. Heterozygous loss of *GATA6* seems to impair pancreatic formation (day 12) by 50-90%, which is a smaller variation in phenotype penetrance across all heterozygous mutants as compared to DE. By day 24, all heterozygous mutants fail to differentiate into endocrine cells at similar rates, thereby reducing the variation in phenotype penetrance further. The similarity of phenotypes between the lab-derived, SCT and Life Technologies protocols suggest that the pancreatic phenotypes observed in the heterozygous mutants are not a consequence of the differentiation protocol used, and are true effects of heterozygous loss of *GATA6*.

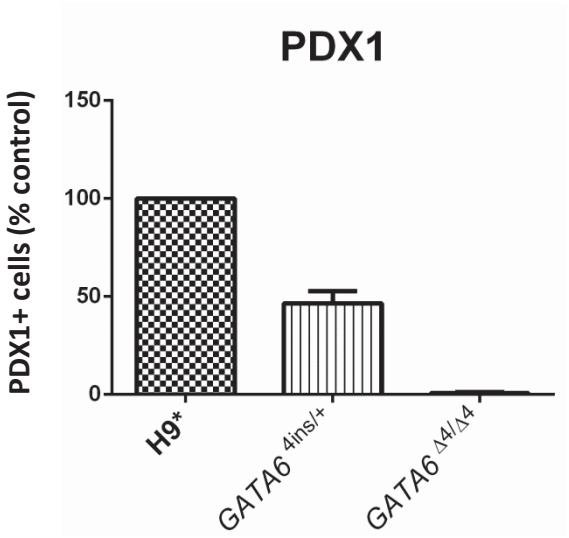


Figure 102. The number of PDX1+ cells is decreased in *GATA6* $^{4ins/+}$ cells and is almost completely zero in *GATA6* $^{\Delta4/\Delta4}$ H9 cells differentiated via STEMCELL Technologies kit. Cells were fixed on day 12 and were stained for PDX1. Data show results of one experiment that is representative of at least two independent experiments.

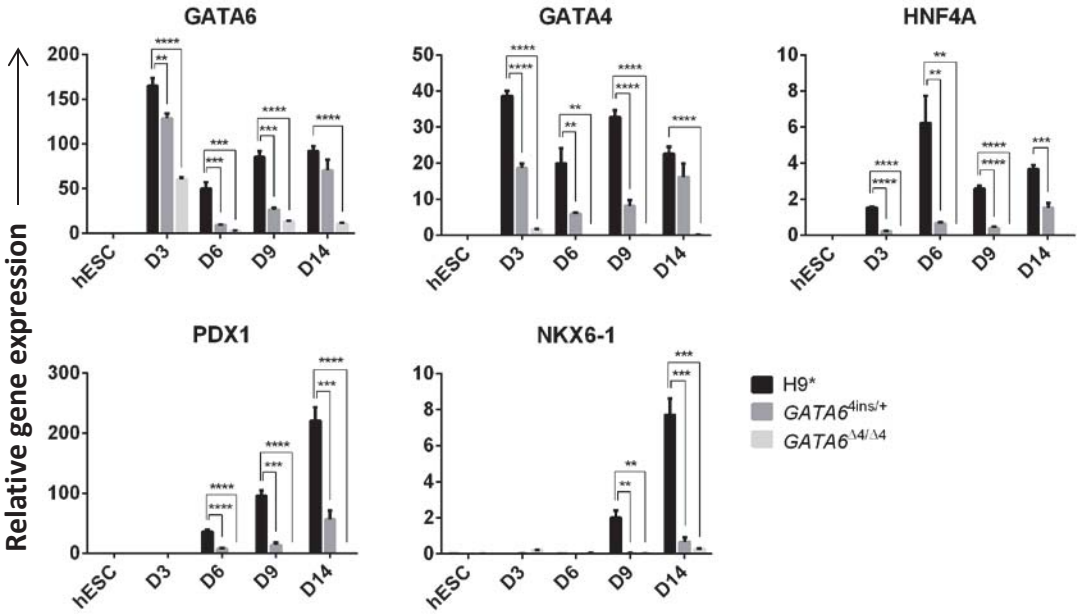


Figure 103. qRT-PCR analyses of H9*, *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ cells on days 3, 6, 9 and 14 differentiated via STEMCELL Technologies kit. RNA was extracted at specific stages and the expression patterns of key markers were determined. Data are triplicate samples of one experiment and representative of three independent experiments. Error bars indicate standard deviation.

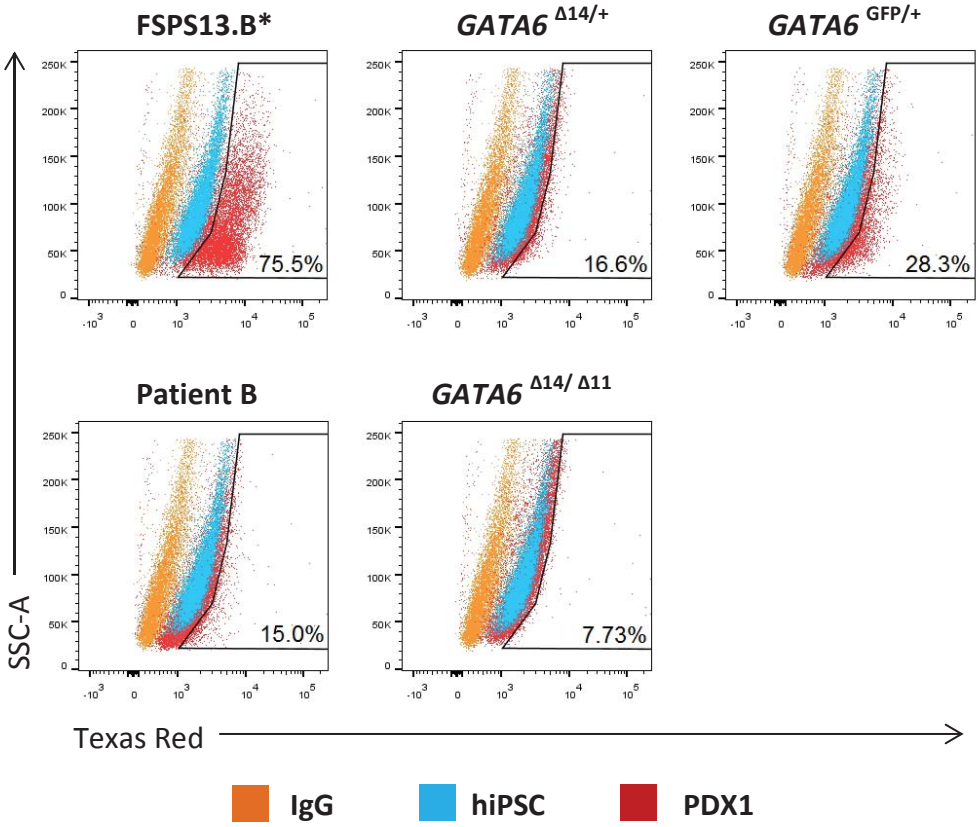


Figure 104. The number of PDX1+ cells is decreased in *GATA6* heterozygous FSPS13.B mutant cells and Patient B differentiated via PSC Definitive Endoderm Induction Kit from Life Technologies. Cells were fixed on day 12 and were stained for PDX1. Data show results of one experiment that is representative of at least two independent experiments.

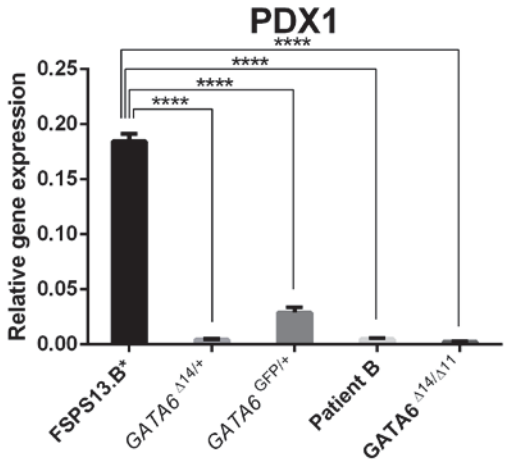


Figure 105. qRT-PCR analyses of FPSP13.B*, FSPS13.B-derived mutant cells and Patient B on day 12 differentiated via PSC Definitive Endoderm Induction Kit from Life Technologies. RNA was extracted on day 12 and the expression of *PDX1* was determined. Data are triplicate samples of one experiment, error bars indicate standard deviation.

3.5. *GATA6* is a key regulator of DE and pancreatic specification

Next, to investigate if *GATA6* is a master regulator of the DE and pancreatic transcriptional networks, I conducted RNA-sequencing (RNA-seq) on H9*, H9-derived *GATA6* $^{4ins/+}$, *GATA6* $^{\Delta4/\Delta4}$ and Patient A (clones 1-3) mutant cells at the DE and pancreatic stages of the lab differentiation protocol. Specifically, I performed RNA-seq for biologically triplicate samples on days 3 and 12 for *GATA6* $^{4ins/+}$ mutant cells, days 3 and 12 for Patient A cells, days 2 and 3 for *GATA6* $^{\Delta4/\Delta4}$ mutant cells and days 2, 3, and 12 for H9* cells. In addition, to identify the direct interacting partners of GATA6, I also perform chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) for biologically duplicate samples on days 3 and 12 for H9* and H9-derived *GATA6* $^{4ins/+}$ mutant cells, and on day 12 for FSPS13.B* and FSPS13.B-derived *GATA6* $^{\Delta14/+}$, *GATA6* $^{GFP/+}$ mutant cells. Pre-processing and downstream analyses of the RNA- and ChIP-seq data were performed with the help of Dr. Pedro Madrigal. These experiments will elucidate the molecular mechanism of *GATA6* and reveal the functional consequences and global transcriptional profiles of *GATA6*.

3.5.1. Loss of GATA6 perturbs the DE transcriptional network and promotes mesoderm formation

Using RNA-seq data, I first performed differential gene expression analyses between *GATA6* $^{\Delta4/\Delta4}$ mutant cells against H9* cells at day 2 to investigate the effect of biallelic loss of *GATA6* on the GRN prior to DE formation. I found 4,679 differentially expressed genes between H9* and *GATA6* $^{\Delta4/\Delta4}$ cells (Table S1) (adjusted p-value 0.01; fold-change ≥ 2). Of these, 3,649 were protein coding genes; 2,239 down-regulated in *GATA6* $^{\Delta4/\Delta4}$ cells and 1,410 up-regulated in *GATA6* $^{\Delta4/\Delta4}$ cells. Interestingly, gene ontology (GO) analysis via DAVID revealed the enrichment of many mesodermal developmental pathways, which were found up-regulated in *GATA6* $^{\Delta4/\Delta4}$ mutant cells (Figure 106). Conversely, endodermal developmental pathways were down-regulated in *GATA6* $^{\Delta4/\Delta4}$ mutant cells. In a heat map showing the top 10 most variable genes, *SOX17* was amongst the most highly down-regulated genes, suggesting that the expression of *SOX17* is regulated by *GATA6* (Figure 107).

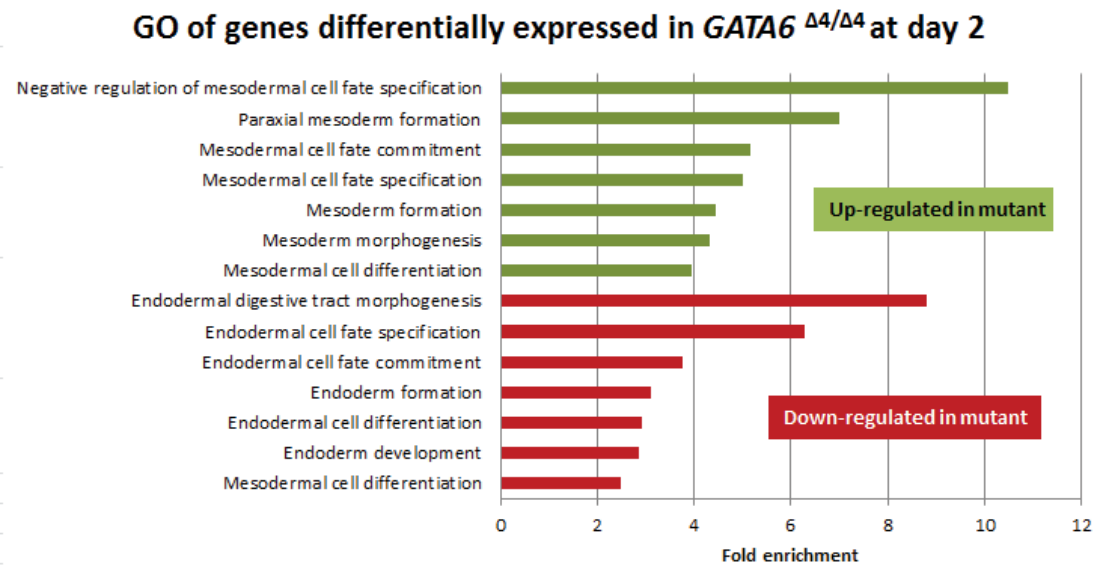


Figure 106. Enriched GO of *GATA6* $^{\Delta4/\Delta4}$ mutant cells against H9* cells on day 2 from RNA-seq. GO shows fold enrichment of many mesodermal and endodermal developmental pathways up- and down-regulated respectively in *GATA6* $^{\Delta4/\Delta4}$ mutant cells. Data show results of triplicate samples in one experiment representative of three independent experiments.

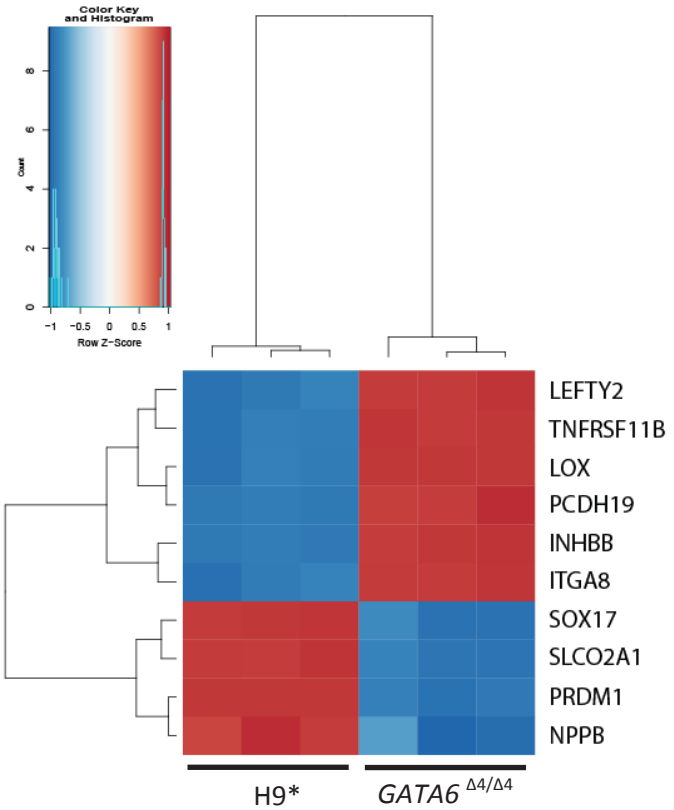


Figure 107. Key DE marker, *SOX17*, is one of the most highly down-regulated genes in *GATA6* $^{\Delta4/\Delta4}$ mutant cells. Heat map illustrating differential gene expression of the top 10 most variable genes in *GATA6* $^{\Delta4/\Delta4}$ mutant cells against H9* cells.

From RNA-seq results on day 3, I found 7,472 differentially expressed genes between H9* and *GATA6* $^{\Delta4/\Delta4}$ mutant cells (Table S2), 2,898 genes between H9* and *GATA6* $^{4ins/+}$ (Table S3), and 6,977 between clones 1-3 of Patient A mutant cells (Table S4). Of the 7,472 genes, 5,393 were protein coding genes; 2,702 down-regulated in *GATA6* $^{\Delta4/\Delta4}$ cells and 2,691 up-regulated in *GATA6* $^{\Delta4/\Delta4}$ cells. Of the 2,898 genes, 1,460 were protein coding genes; 729 down-regulated in *GATA6* $^{4ins/+}$ cells and 731 up-regulated in *GATA6* $^{4ins/+}$ cells. Of the 6,977 genes, 3,981 were protein coding genes; 2,154 down-regulated in Patient A cells and 1,827 up-regulated in Patient A cells.

In a heat map derived from RNA-seq results focusing on genes involved in ectoderm, endoderm and mesoderm development, I observed that consistent with my day 2 RNA-seq data, *GATA6* $^{\Delta4/\Delta4}$ mutant cells displayed a decreased endodermal signature, with a concomitant increase in the expression of mesodermal genes (Figure 108). A similar trend was observed when comparing between H9*, *GATA6*$^{4ins/+}$ and clones 1 to 3 of Patient A (Figure 109).

Next, in order to subsequently identify the direct interacting partners of GATA6, I performed ChIP on H9* cells and H9-derived *GATA6* $^{4ins/+}$ cells which were differentiated to the DE stage via the lab protocol to identify those genes bound directly by GATA6. I validated the ChIP via qPCR using primers recognising a region on the human HNF4α P2 promoter as a GATA6 positive control (primer sequences are listed in Chapter 2.10.5). The primer sequences were kindly provided to us by our collaborator Dr. Santi Rodriguez. Indeed, GATA4 and GATA6 have been shown to bind to HNF4α (Sumi et al., 2007). Results from qPCR show a 4-fold increase in binding in H9* cells and only a 1-fold increase in binding in *GATA6* $^{4ins/+}$ mutant cells compared to IgG (Figure 110). To ensure that this was not a result of a decreased number of cells, I performed FACS on GATA6 at the DE stage and found that both H9* and *GATA6* $^{4ins/+}$ cells similarly express approximately 95% GATA6-positive cells (Figure 111), indicating that the decreased DNA binding observed in *GATA6* $^{4ins/+}$ mutant cells is indeed a consequence of heterozygous loss of *GATA6*.

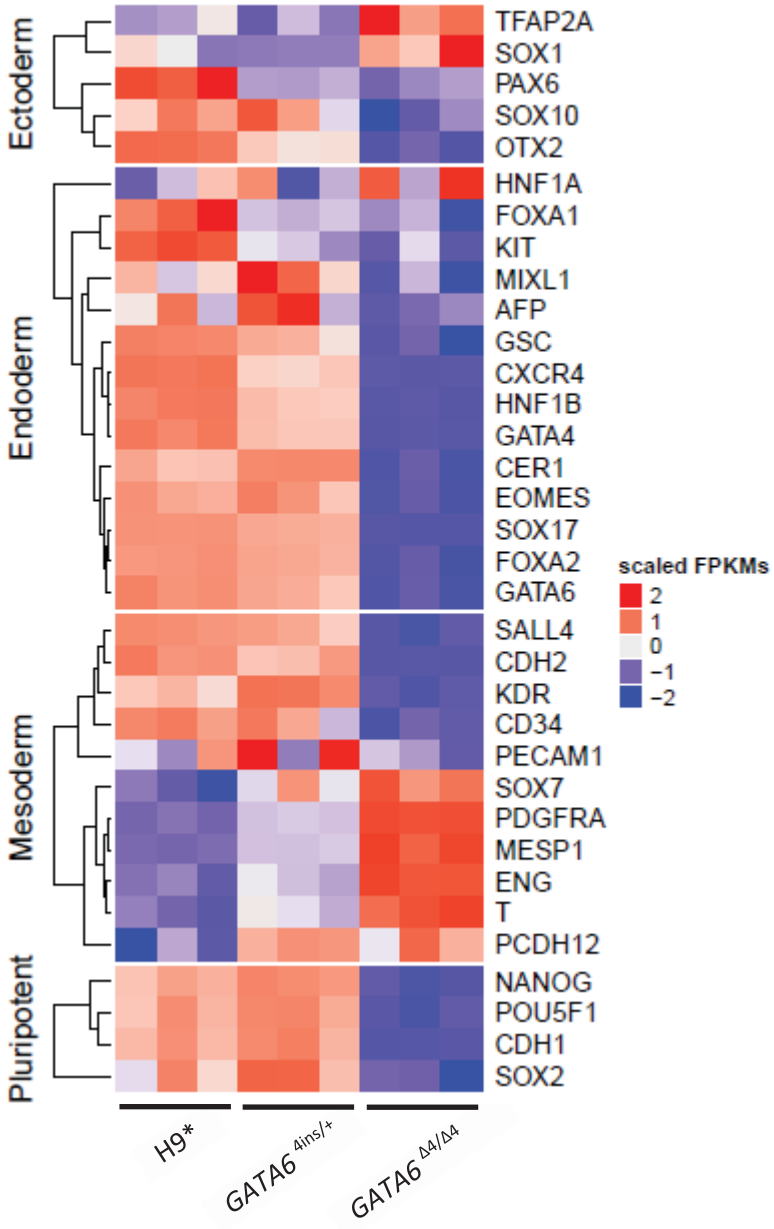


Figure 108. Heat map illustrating differential gene expression of key germ layer markers via RNA-seq between H9*cells and H9-derived *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ mutant cells at the DE stage (day 3). *n* = 3 biological replicates for each cell line.

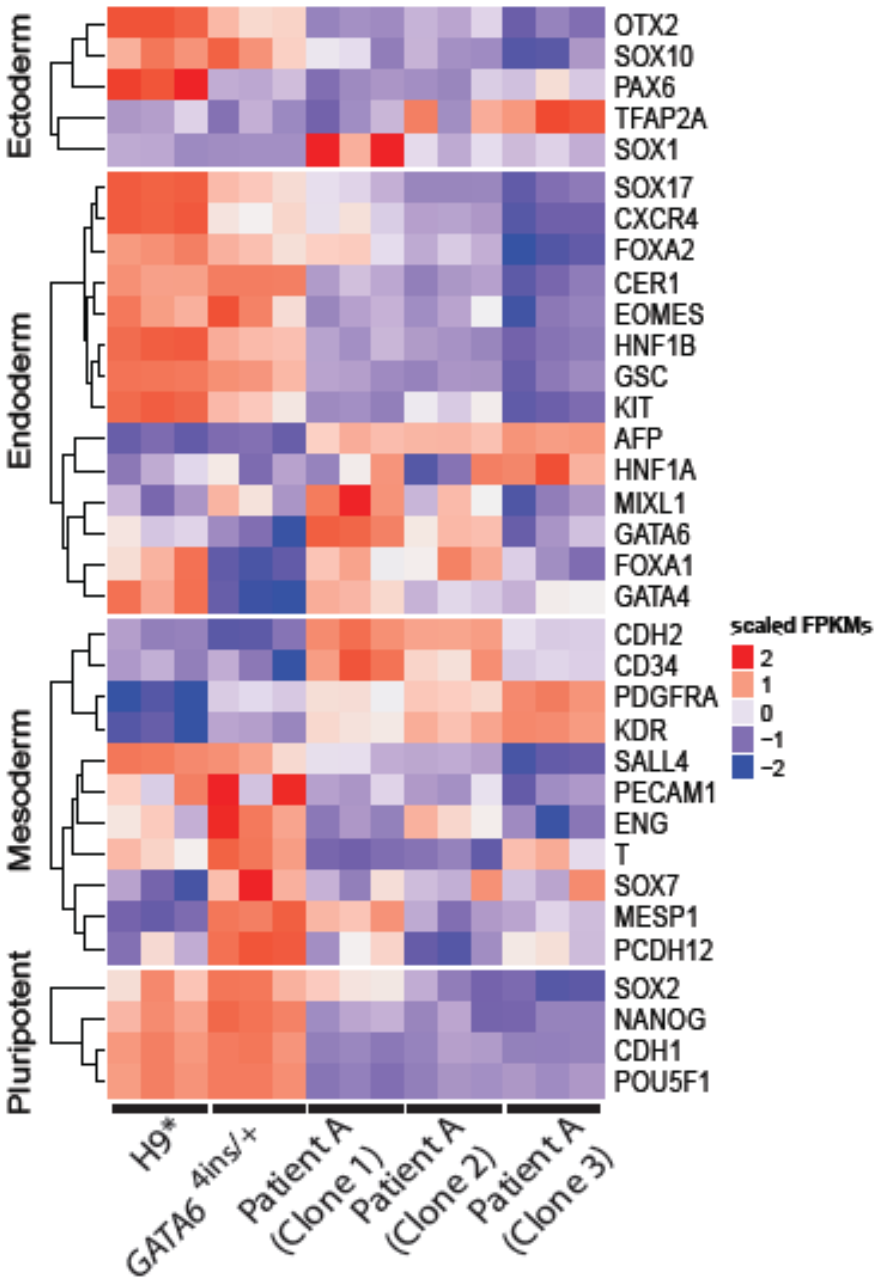


Figure 109. Heat map illustrating differential gene expression of key germ layer markers via RNA-seq between H9*cells and H9-derived *GATA6* $^{4ins/+}$ and clones 1-3 of Patient A mutant cells at the DE stage (day 3). *n* = 3 biological replicates for each cell line.

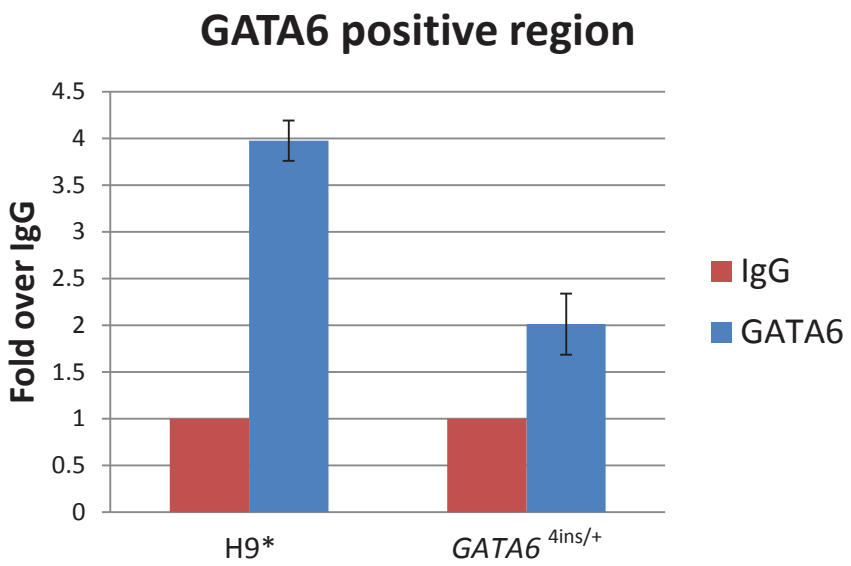


Figure 110. ChIP validation via qPCR using primers specific to a GATA6 positive binding region on day 3. Values of input samples were subtracted from the values of IgG or GATA6 samples and the graph shows fold over IgG (normalised to 1) of H9* and H9-derived *GATA6* $^{4ins/+}$ cells.

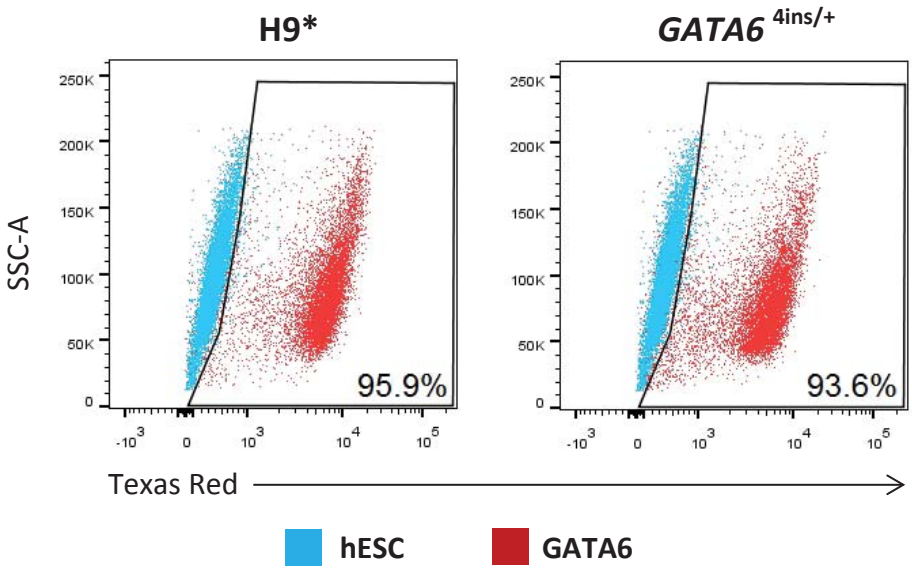


Figure 111. The number of GATA6+ cells is similar in H9* and H9-derived *GATA6* $^{4ins/+}$ cells differentiated via the lab protocol on day 3. Cells were fixed on day 3 and were stained for *GATA6*. Data show results of one experiment that is representative of at least two independent experiments.

Now that I had validated the ChIP, I proceeded to perform ChIP-seq on these samples with the aim of revealing the direct binding targets of GATA6. From the ChIP-seq data at the DE stage, 12,098 peaks were called (IDR ≤ 0.05; median width=417 bp) for H9* cells, of which 10,669 were associated to genes, 4,790 of them protein coding (Table S5). For *GATA6* $^{4ins/+}$ mutant cells, 2,220 peaks were called, of which 1,137 of them were protein coding (Table S6).

Comparing *GATA6* binding at the *GATA4* locus between H9* and *GATA6*$^{4ins/+}$ cells, I saw an enrichment of *GATA6* binding in H9* cells, suggesting that *GATA6* directly regulates *GATA4* during DE development (Figure 112). RNA-seq data showing *GATA4* expression levels in *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ mutant cells (Figure 112) are also consistent with qRT-PCR data (Figure 68), where *GATA4* is down-regulated in *GATA6* $^{4ins/+}$ cells by approximately 50% and almost completely abolished in *GATA6* $^{\Delta4/\Delta4}$ mutant cells, thus further validating the RNA-seq results.

To relate *GATA6* binding to global gene expression dynamics and investigate genes that were not only direct interacting partners of GATA6, but also differentially expressed, I compared the ChIP-seq analysis to the RNA-seq data set to identify the subset of *GATA6*-bound genes that were up- or down-regulated. In *GATA6* $^{\Delta4/\Delta4}$ mutant cells, I found that 1,120 protein coding genes were *GATA6*-bound and down-regulated including DE genes such as *CXCR4* and *SOX17*, and posterior foregut markers such as *HNF1B* and *HNF4A*, while 745 genes were *GATA6*-bound and up-regulated (Figure 113). 337 and 607 genes were *GATA6*-bound and down-regulated in *GATA6* $^{4ins/+}$ and Patient A cells respectively, while 254 and 616 genes were *GATA6*-bound and up-regulated in *GATA6* $^{4ins/+}$ and Patient A cells respectively (Figure 113). Overlapping the RNA-seq differentially expressed gene sets of *GATA6* $^{\Delta4/\Delta4}$, *GATA6* $^{4ins/+}$ and Patient A identified 143 genes commonly down-regulated and 104 genes commonly up-regulated in these three separate data sets (Figure 114). Key endoderm markers *CXCR4*, *SOX17*, *GATA4* were among the 143 genes commonly down-regulated, suggesting that *GATA6* is a direct regulator of DE development.

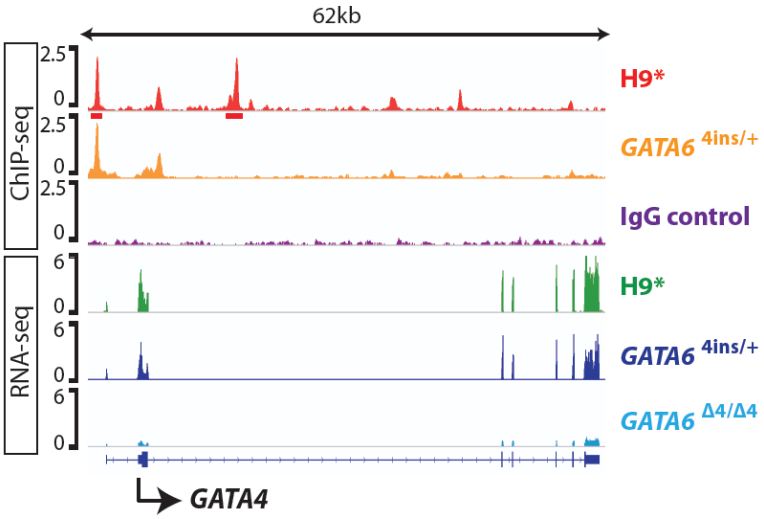
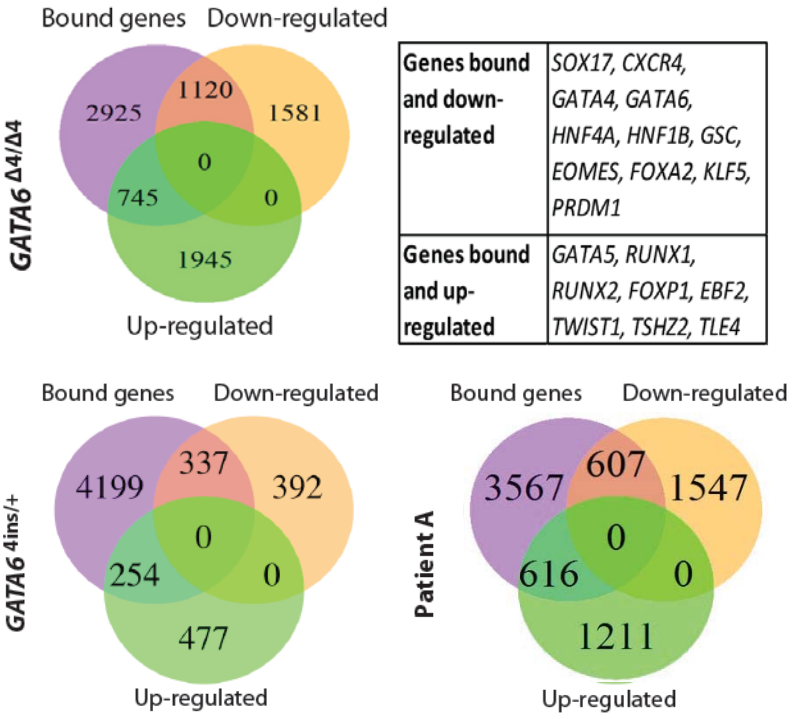


Figure 112. *GATA6* binding is enriched near the *GATA4* gene and *GATA4* is decreased in H9-derived *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ mutant cells. ChIP-seq binding profiles of H9* and GATA6 $^{4ins/+}$ and RNA-seq representation of *GATA4* locus showing gene expression in H9*cells and H9-derived *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ mutant cells at the DE stage (day 3). The input control profile (IgG control) is included for comparison. ChIP-seq binding profile is derived from merging two biological replicates.



Genes bound and down-regulated	*SOX17, CXCR4, GATA4, GATA6, HNF4A, HNF1B, GSC, EOMES, FOXA2, KLF5, PRDM1*
Genes bound and up-regulated	*GATA5, RUNX1, RUNX2, FOXP1, EBF2, TWIST1, TSHZ2, TLE4*

Figure 113. Venn diagrams indicating the overlap of GATA6-bound genes from ChIP-seq and differentially expressed genes from RNA-seq at the DE stage (day 3). Diagram shows GATA6-bound down- or up-regulated genes of H9-derived *GATA6* $^{\Delta4/\Delta4}$, *GATA6* $^{4ins/+}$ and Patient A mutant cells compared to H9* cells. Key bound genes up-or down-regulated for *GATA6* $^{\Delta4/\Delta4}$ are indicated in the table.

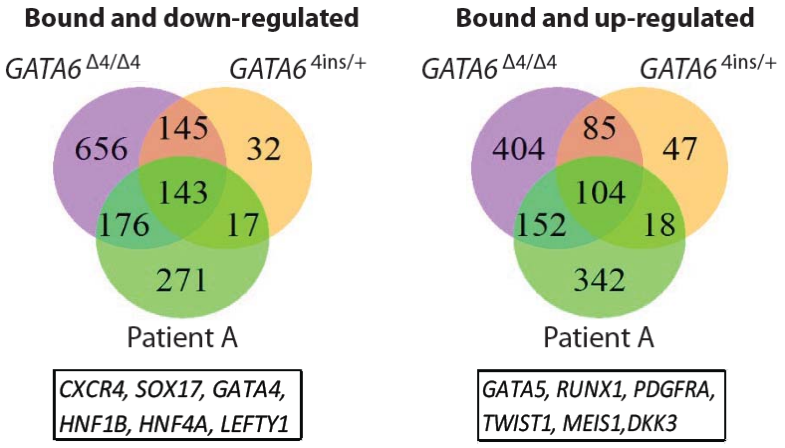
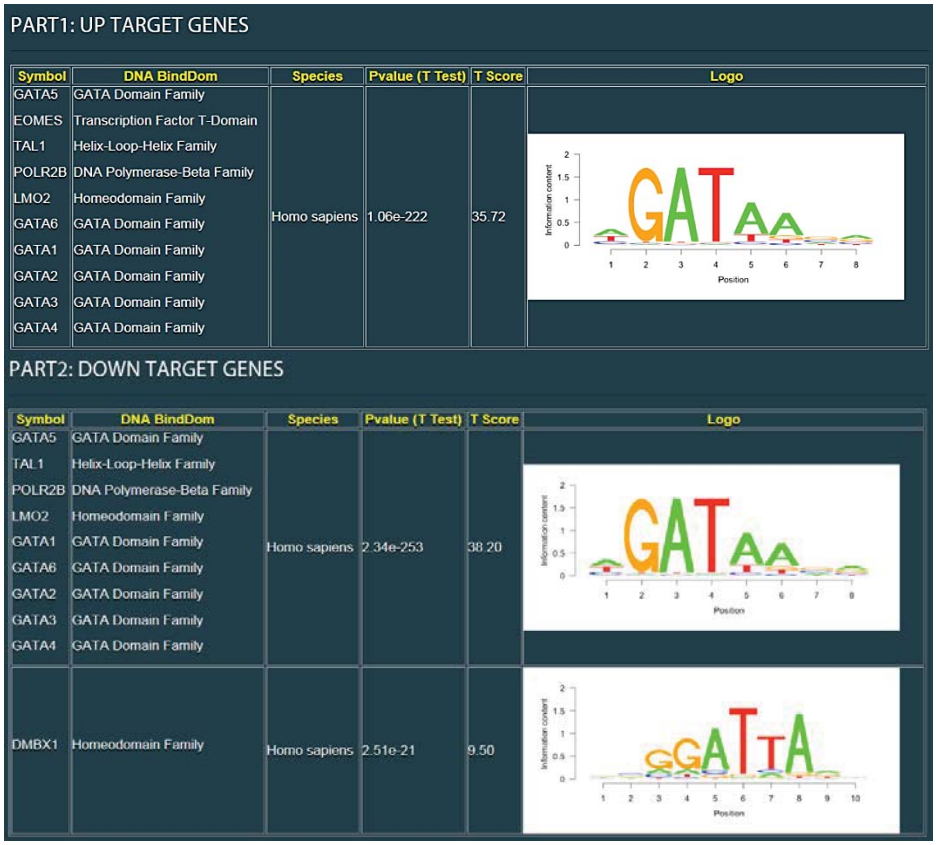


Figure 114. Venn diagrams indicating the triple overlap of *GATA6*-bound genes at the DE stage (day 3). Down- or up-regulated genes of H9-derived *GATA6* $^{\Delta4/\Delta4}$, *GATA6* $^{4ins/+}$ and Patient A mutant cells are represented. Key bound genes up-or down-regulated common in all three mutants are indicated in the respective boxes.

To infer genes that are directly targeted and regulated by GATA6 and with the help of Dr. Denil, I used a software package binding and expression target analysis (BETA) to integrate the ChIP-seq dataset of H9* with differential gene expression data of *GATA6* $^{\Delta4/\Delta4}$, *GATA6* $^{4ins/+}$ and Patient A from RNA-seq (Wang et al., 2013). Motif analyses generated by BETA showed the GATA motif as highly enriched in both up- and down-regulated target genes, thus further validating the ChIP (Figure 115). Results from activating/repressive function prediction did not show an activating or repressive function of GATA6 in *GATA6* $^{\Delta4/\Delta4}$, *GATA6* $^{4ins/+}$ and Patient A mutant cell lines (Figure 116). Next, I performed gene ontology (GO) analyses on the direct targets prediction of up- and down-regulated gene lists generated by BETA data set using the DAVID tool (Huang da et al., 2009b, Huang da et al., 2009a). I found that endoderm development is consistently up-regulated in H9* cells compared to *GATA6* $^{\Delta4/\Delta4}$, *GATA6* $^{4ins/+}$ and Patient A mutant cells, while mesoderm development is consistently up-regulated in all three mutant cell lines compared to H9* cells (Figure 117). Together, these results show a direct molecular function of *GATA6* in driving the development of the endoderm. Loss of one or two alleles of *GATA6* promotes mesodermal formation.



PART1: UP TARGET GENES

Symbol	DNA BindDom	Species	Pvalue (T Test)	T Score	Logo
GATA5	GATA Domain Family	Homo sapiens	1.06e-222	35.72	
EOMES	Transcription Factor T-Domain				
TAL1	Helix-Loop-Helix Family				
POLR2B	DNA Polymerase-Beta Family				
LMO2	Homeodomain Family				
GATA6	GATA Domain Family				
GATA1	GATA Domain Family				
GATA2	GATA Domain Family				
GATA3	GATA Domain Family				
GATA4	GATA Domain Family				

PART2: DOWN TARGET GENES

Symbol	DNA BindDom	Species	Pvalue (T Test)	T Score	Logo
GATA5	GATA Domain Family	Homo sapiens	2.34e-253	38.20	
TAL1	Helix-Loop-Helix Family				
POLR2B	DNA Polymerase-Beta Family				
LMO2	Homeodomain Family				
GATA1	GATA Domain Family				
GATA6	GATA Domain Family				
GATA2	GATA Domain Family				
GATA3	GATA Domain Family				
GATA4	GATA Domain Family				
DMBX1	Homeodomain Family	Homo sapiens	2.51e-21	9.50	

Figure 115. Screenshot of binding motif analysis on UP and DOWN target regions of GATA6 ChIP-seq on day 3 derived from BETA analysis. Similar motifs are grouped together, and the motif logo of the most significant factor in the group is provided in the last column. The motif symbol, DNA-binding domain and species are shown in the first three columns; the *t* score and the *P* value from the *t* test are shown in the middle two columns.

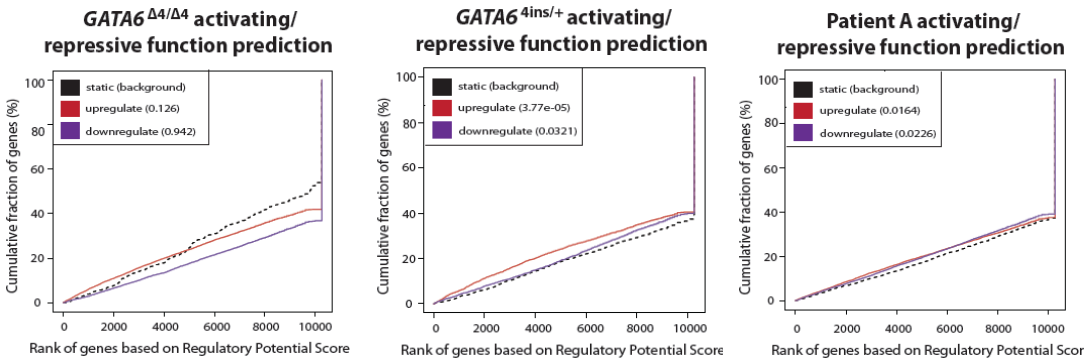


Figure 116. BETA output of activating/repressive function prediction of H9-derived *GATA6* $^{\Delta4/\Delta4}$, *GATA6* $^{4ins/+}$ and Patient A mutant cells on day 3. The red and the purple lines represent the up-regulated and down-regulated genes, respectively. The dashed black line indicates the non-differentially expressed genes as background. Genes are cumulated by the rank on the basis of the regulatory potential score from high to low. *P*-values that represent the significance of the UP or DOWN group distributions are compared with the NON group by the Kolmogorov-Smirnov test.

Gene Ontology of genes bound and differenitally expressed in *GATA6* $^{\Delta4/\Delta4}$

Category	P-value	Gene symbol
Up-regulated in WT		
Endoderm development	8.74E-04	GDF3, COL4A2, HNF1B, FGF8, NODAL, EOMES, SMAD2, VTN, MMP15, HMGA2, HSBP1, COL5A2, MIXL1, DUSP5, HHEX, DUSP1, GATA6
Pancreas development	0.03667	INSM1, HNF1B, FOXA2, WFS1, SOX4, SMAD2, BAD, ISL1, TCF7L2, HHEX, BAK1, ACVR2B, MEIS2, GATA6, EIF2AK3, BMP5, BMP6
Up-regulated in *GATA6* $^{\Delta4/\Delta4}$		
Skeletal system development	2.66E-06	THRA, NDST1, HEXB, EDN1, TGFB3, BCAN, GJA5, TGFB2, MBTD1, INSIG2, PAX7, TRIM45, CHST11, RAB23, PHOSPHO1, HHIP, ANO6, ALX1, COL10A1
Mesoderm formation	0.001386	FGFR2, SIX2, SMAD3, ITGA3, WLS, PAX2, SNAI1, WNT3, DKK1, HAND1, SFRP2, ITGA8, EPB41L5, FOXC1, TLX2, ACVR1

Gene Ontology of genes bound and differentially expressed in *GATA6* $^{4ins/+}$

Category	P-value	Gene symbol
Up-regulated in WT		
Endoderm development	6.78E-04	GDF3, COL4A2, NOG, HNF1B, NODAL, SMAD2, MMP15, HMGA2, HSBP1, DUSP5, HHEX, DUSP1, GATA6, ITGA7, COL11A1
Up-regulated in *GATA6* $^{4ins/+}$		
Mesoderm formation	3.31E-04	FGFR2, SIX2, ITGA3, WLS, SMAD1, ITGB1, SNAI1, WNT3, DKK1, HAND1, SFRP2, ITGA8, FOXC1, TLX2

Gene Ontology of genes bound and differentially expressed in Patient A

Category	P-value	Gene symbol
Up-regulated in WT		
Endoderm development	0.026782	GDF3, NANOG, HNF1B, ONECUT1, NODAL, EOMES, MMP15, KIF16B, HSBP1, ZFP36L1, HHEX, LHX1, ITGA7, COL6A1
Up-regulated in Patient A		
Mesoderm development	0.00119	FGFR2, NOG, FGF8, TP63, ITGB3, PAX2, WNT3, OSR1, HAND1, PPP2CA, YAP1, PALB2, TLX2, TBX3, SMAD4, SIX2, SMAD3, ITGA2, ITGA3, SMAD1, WLS, POGLUT1, ACVR2B, CTDNEP1, DKK1, ITGA8, EPB41L5, PRKAR1A, ACVR1

Figure 117. Enriched gene ontology derived from BETA analysis showing developmental pathways. Tables are derived from direct target genes differentially expressed between H9* (WT), H9-derived *GATA6* $^{\Delta4/\Delta4}$, *GATA6* $^{4ins/+}$ and Patient A mutant cells on day 3. *P*-values of the developmental pathways are indicated along with the respective genes.

Based on results from integrating the RNA-seq and ChIP-seq datasets, GATA6 seems to be a master regulator of DE development. This prompted me to investigate whether GATA6 interacts with other know DE regulators such as EOMES and SMAD2/3. With the help of Dr. Denil, I looked at the physical overlap of potential transcription factor binding sites from my *GATA6* ChIP-seq dataset on day 3 with previously published EOMES (Teo et al., 2011) and SMAD2/3 ChIP data (Brown et al., 2011) (Figure 118). The data were reprocessed as described in Chapter 2.11.2. This resulted in 16,303, 20,089 and 2,613 sites respectively. The overlap between *GATA6* and *EOMES* was 8,126/20,089 (40.5%); the overlap between *GATA6* and *SMAD2/3* was 950/2,613 (36.4%). Overlap of the three data sets is found in 858 locations. Of those sites upstream from a gene, the mean distance to the transcription start site was 831bp (median 0). Results from this experiment suggests that GATA6 co-operates with EOMES and SMAD2/3 to deploy the gene regulatory network governing human DE development.

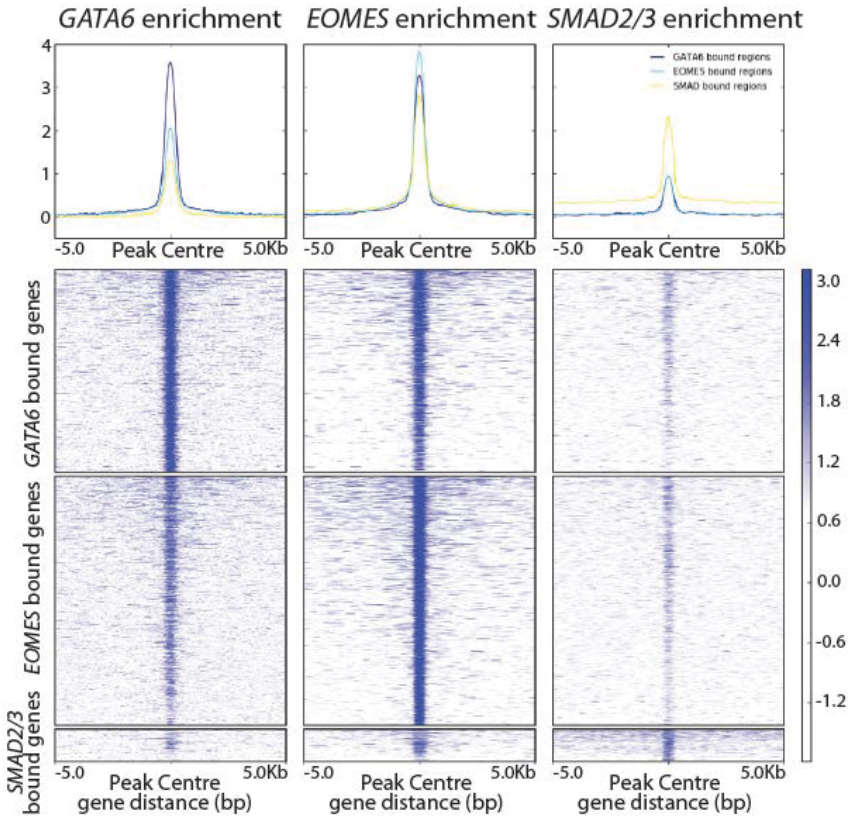


Figure 118. Density heat maps of *GATA6*-binding peak intensity at DE. Figure shows direct overlaps with known endodermal regulators including *SMAD2/3* and *EOMES* within a 10-kb window centered at the transcription start site (TSS).

3.5.2. GATA6 haploinsufficiency perturbs the pancreatic transcriptional network

Next, to investigate whether *GATA6* also plays a master regulator role at the pancreatic stage, I repeated the RNA-seq at the PE stage (day 12) for H9* cells, H9-derived *GATA6* $^{4ins/+}$ and Patient A mutant cells. H9* RNA-seq data closely recapitulated a previously profiled transcriptome using the same protocol (Spearman **ρ** =0.77 for in vitro, **ρ** =0.59 for in vivo, P < 2.2e-16), which was used for the analyses of *in vitro* MPCs described in Cebola et al. (Cebola et al., 2015).

I found 1,423 differentially expressed genes between H9* and *GATA6* $^{4ins/+}$ (Table S7), and 6,093 between clones 1-3 of Patient A mutant cells (Table S8). Of the 1,423 genes, 1,230 were protein coding genes; 899 down-regulated in *GATA6* $^{4ins/+}$ cells and 331 up-regulated in *GATA6* $^{4ins/+}$ cells. Of the 6,093 genes, 4,148 were protein coding genes; 2,424 down-regulated in Patient A cells and 1,724 up-regulated in Patient A cells. In a heat map derived from RNA-seq results at the PE stage focusing on genes involved in pancreatic development, I observed that consistent with qRT-PCR (Figure 97 and Figure 101) and FACS results (Figure 96) discussed earlier, *GATA6* $^{4ins/+}$ and Patient A mutant cells displayed decreased pancreatic formation (Figure 119).

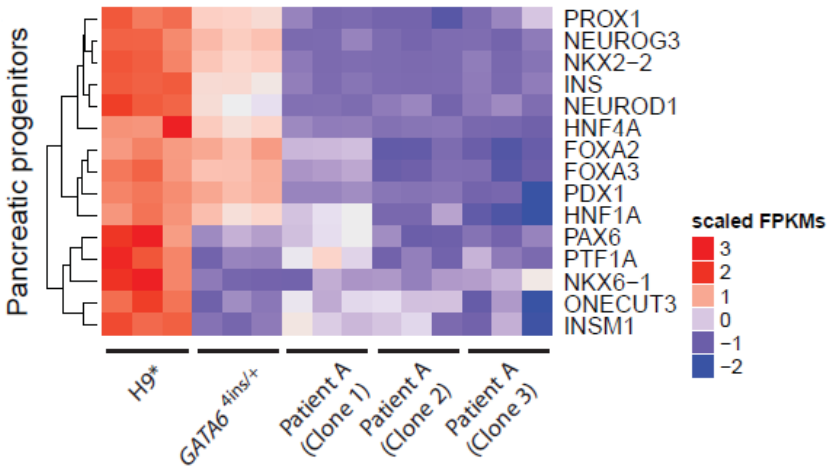


Figure 119. Heat map illustrating differential gene expression of key pancreatic markers via RNA-seq between H9*cells, H9-derived *GATA6* $^{4ins/+}$ and clones 1-3 of Patient A mutant cells at the PE stage. *n* = 3 biological replicates for each cell line.

Next, I performed ChIP on H9* cells and H9-derived *GATA6* $^{4ins/+}$ cells which were differentiated to the PE stage via the lab protocol to identify those genes bound directly to and interacting with GATA6 at the pancreatic stage. I validated the ChIP via qPCR using primers that would amplify a region on the human HNF4α P2 promoter, which acts as a GATA6 positive control. In addition, I also used a primer pair that amplifies a positive PDX1 binding location, again kindly provided to us by our collaborator Dr. Santi Rodriguez. Results from qPCR show an approximately 10-fold higher binding at the GATA6 positive region in H9* cells compared to *GATA6* $^{4ins/+}$ mutant cells and an approximately 7-fold higher binding at the PDX1 positive region in H9* cells compared to *GATA6* $^{4ins/+}$ mutant cells (Figure 120). To ensure that this was not a result of a decreased number of cells, I performed FACS on GATA6 at the PE stage and found that both H9* and *GATA6* $^{4ins/+}$ cells had similar numbers of GATA6+ cells of approximately 75% GATA6+ cells (Figure 121), indicating that the decreased DNA binding observed in *GATA6* $^{4ins/+}$ mutant cells is indeed a consequence of heterozygous loss of *GATA6*.

Unfortunately, when I proceeded to perform ChIP-seq on H9 WT cells on day 12, the ChIP-seq data retrieved was of low quality, which resulted in only 171 peaks called. This was in contrast to GATA6 ChIP-seq data published by Cebola et al., where using the same parameters, 2,060 peaks were called (Cebola et al., 2015). The ChIP experiment was repeated two more times and ChIP-seq was performed on samples derived from these two independent experiments, but the data quality remained low.

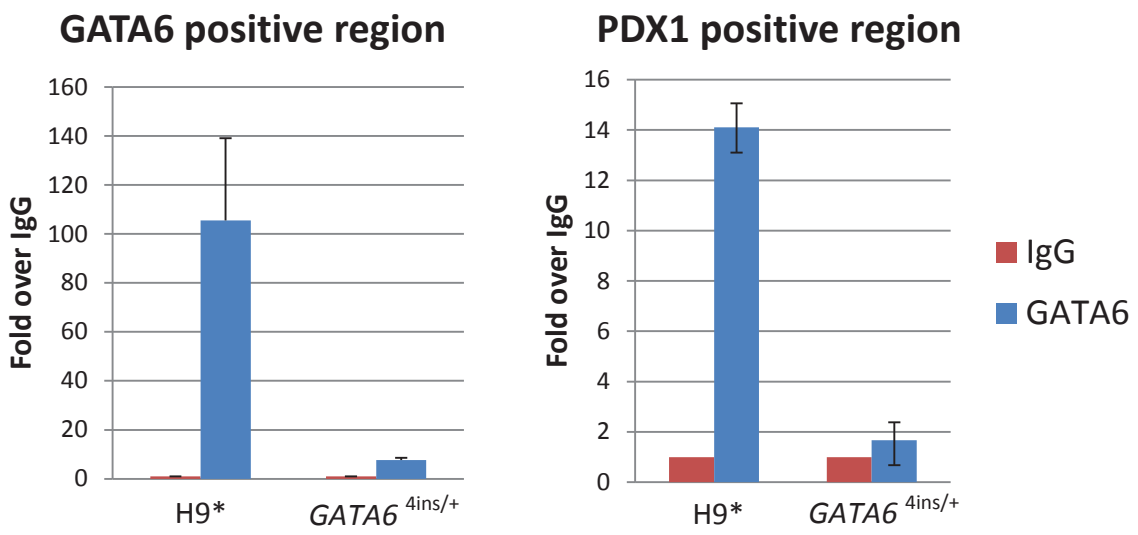


Figure 120. ChIP validation via qPCR using primers specific to a GATA6 and PDX1 positive binding region on day 12 for H9* and H9-derived *GATA6* $^{4ins/+}$ mutant cells. Values of input samples were subtracted from the values of IgG or *GATA6* samples and the graph shows fold over IgG (normalised to 1) of H9* and H9-derived *GATA6* $^{4ins/+}$ cells.

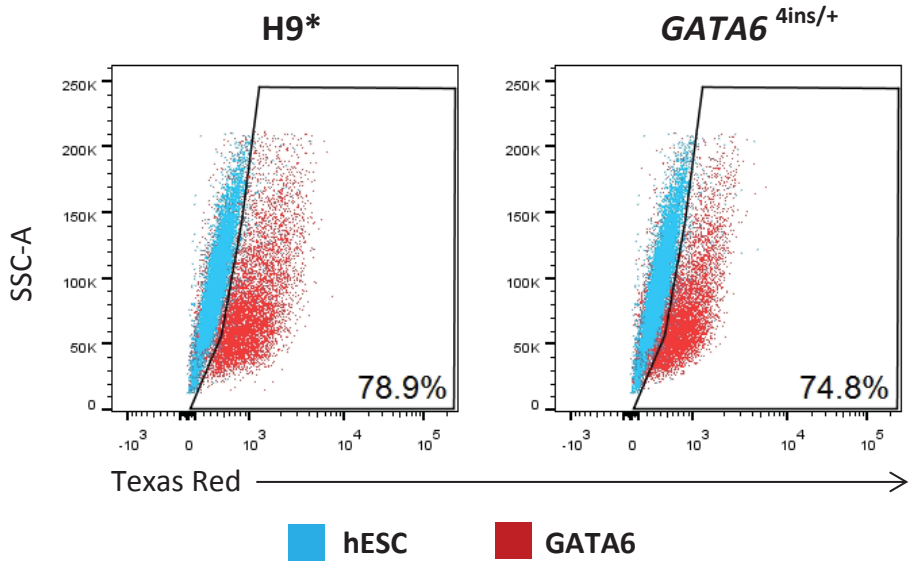


Figure 121. Number of GATA6+ cells is similar in H9* and H9-derived *GATA6* $^{4ins/+}$ cells differentiated via the lab protocol at day 12. Cells were fixed at day 12 and were stained for GATA6. Data show results of one experiment that is representative of at least two independent experiments.

As a result, I was prompted to use FSPS13.B cell line and its respective heterozygous mutant cell lines for ChIP and ChIP-seq experiments. I hence performed ChIP on FSPS13.B* cells and FSPS13.B -derived *GATA6* $^{\Delta 14/+}$ and *GATA6* $^{GFP/+}$ cells which were differentiated to the PE stage via the lab protocol. Validation of the ChIP via qRT-PCR showed an approximately 7-fold and 4-fold higher binding at the GATA6 and PDX1 positive binding regions respectively in FSPS13.B* cells compared to *GATA6* $^{\Delta 14/+}$ mutant cells (Figure 122). On the other hand, an approximately 2-fold higher binding at the GATA6 and PDX1 positive binding regions in FSPS13.B* cells compared to *GATA6* $^{GFP/+}$ mutant cells was observed (Figure 122). From FACS data of GATA6 in FSPS13.B*, *GATA6* $^{\Delta 14/+}$ and *GATA6* $^{GFP/+}$ mutant cells, FSPS13.B* and *GATA6* $^{GFP/+}$ mutant cells had similar levels of GATA6+ cells at the PE stage, whereas *GATA6* $^{\Delta 14/+}$ mutant cells had an approximately 30% decrease of GATA6+ cells (Figure 123). Although a decrease was seen in the number of GATA6-expressing cells in *GATA6* $^{\Delta 14/+}$ mutant cells, the decrease in binding was more so for ChIP, suggesting that the decrease in binding could be a combined consequence of fewer cells and the loss of one allele of GATA6.

Sequencing was next performed on these ChIP samples. From the ChIP-seq data at the PE stage, 2,306 peaks were called for FSPS13.B* cells, of which 1,096 of them protein coding (Table S9). For *GATA6* $^{\Delta 14/+}$ mutant cells, 543 peaks were called, of which 234 of them were protein coding (Table S10). For *GATA6* $^{GFP/+}$ mutant cells, 2,376 peaks were called, of which 1,157 of them were protein coding (Table S11). Notably, PDX1 was found to bind GATA6 in both FSPS13.B* and *GATA6* $^{GFP/+}$ mutant cells, but not in *GATA6* $^{\Delta 14/+}$ mutant cells.

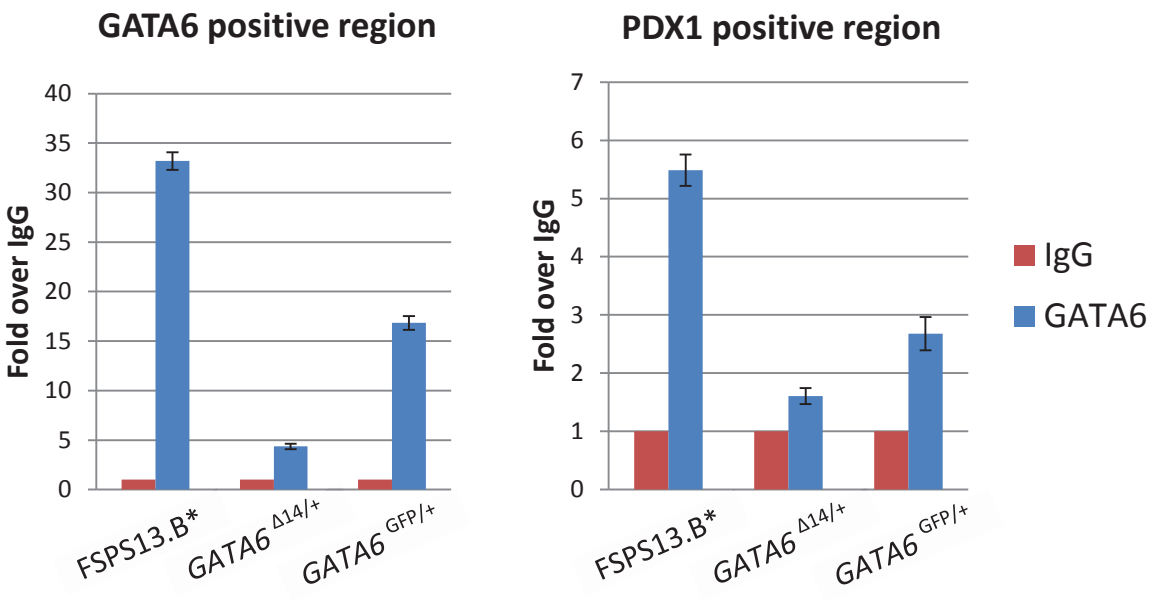


Figure 122. ChIP validation via qPCR using primers specific to a GATA6 and PDX1 positive binding region on day 12 for FSPS13.B* and FSPS13.B-derived *GATA6* $^{\Delta 14/+}$ and *GATA6* $^{GFP/+}$ mutant cells. Values of input samples were subtracted from the values of IgG or GATA6 samples and the graph shows fold over IgG (normalised to 1) of FSPS13.B* and FSPS13.B-derived *GATA6* $^{\Delta 14/+}$ and *GATA6* $^{GFP/+}$ cells.

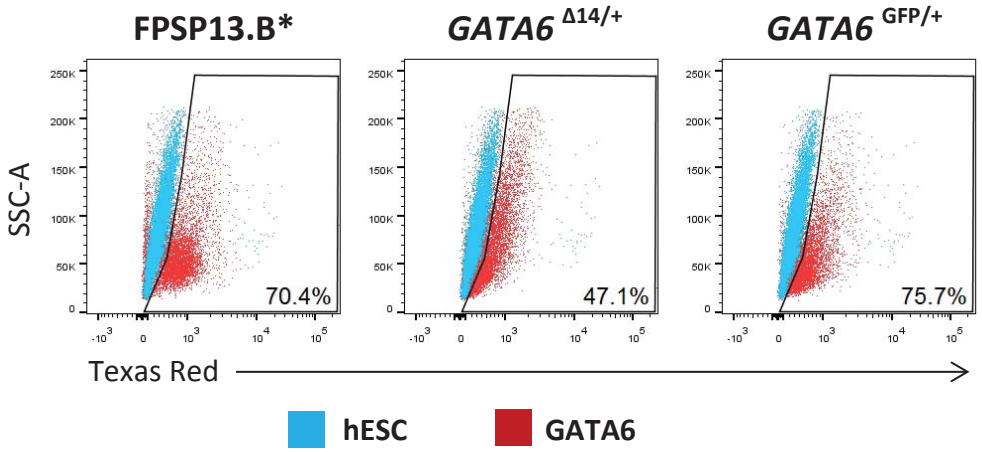
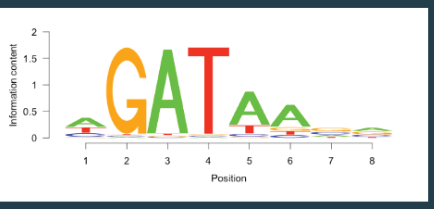


Figure 123. The number of GATA6+ cells is similar in FSPS13.B* and FSPS13.B-derived *GATA6* $^{GFP/+}$ cells but down-regulated in *GATA6* $^{\Delta 14/+}$ cells differentiated via the lab protocol on day 12. Cells were fixed on day 12 and were stained for GATA6. Data show results of one experiment that is representative of at least two independent experiments.

Next, Dr. Denil helped to repeat the BETA analyses to integrate the ChIP-seq dataset of FSPS13.B* with differential gene expression data of H9-derived *GATA6* $^{4ins/+}$ and Patient A mutant cells on day 12. Motif analyses generated by BETA showed the GATA motif as highly enriched in both up- and down-regulated target genes, thus further validating the ChIP (Figure 124). Results from activating/repressive function prediction did not show an activating or repressive function of GATA6 in *GATA6* $^{\Delta 14/+}$ and Patient A mutant cell lines at the PE stage (Figure 125).

Next, I performed gene ontology (GO) using DAVID on the up- and down-regulated gene lists generated by BETA and found that pancreas-related terms such as regulation of insulin secretion, endocrine and pancreas development are up-regulated in FSPS13.B* cells compared to *GATA6* $^{\Delta 14/+}$ and Patient A mutant cells (Figure 126). Interestingly, skeletal and nervous system developments, which are of mesodermal and ectodermal origins respectively, were observed to be up-regulated in both *GATA6* $^{\Delta 14/+}$ and Patient A mutant cells (Figure 126). Together, these results show a direct molecular function of *GATA6* in driving the specification of the human pancreas.

PART1: UP TARGET GENES

Symbol	DNA BindDom	Species	Pvalue (T Test)	T Score	Logo
GATA4	GATA Domain Family	Homo sapiens	1.66e-45	16.77	
TAL1	Helix-Loop-Helix Family				
GATA6	GATA Domain Family				
POLR2B	DNA Polymerase-Beta Family				
LMO2	Homeodomain Family				
GATA3	GATA Domain Family				
GATA2	GATA Domain Family				
GATA1	GATA Domain Family				
GATA5	GATA Domain Family				
DPRX	Homeodomain Family	Homo sapiens	5.36e-04	3.31	

PART2: DOWN TARGET GENES

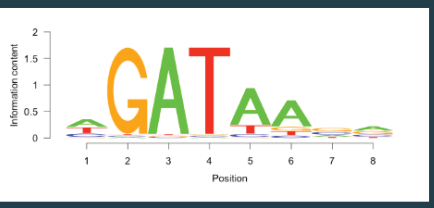
Symbol	DNA BindDom	Species	Pvalue (T Test)	T Score	Logo
GATA5	GATA Domain Family	Homo sapiens	2.02e-17	9.26	
DPRX	Homeodomain Family				
LMO2	Homeodomain Family				
POLR2B	DNA Polymerase-Beta Family				
GATA3	GATA Domain Family				
GATA6	GATA Domain Family				
GATA2	GATA Domain Family				
GATA1	GATA Domain Family				
GATA4	GATA Domain Family				
RHOXF1	Homeodomain Family	Homo sapiens	1.58e-04	3.67	

Figure 124. Screenshot of binding motif analysis on UP and DOWN target regions of GATA6 ChIP-seq on day 12 derived from BETA analysis. Similar motifs are grouped together, and the motif logo of the most significant factor in the group is provided in the last column. The motif symbol, DNA-binding domain and species are shown in the first three columns; the *t* score and the *P* value from the *t* test are shown in the middle two columns.

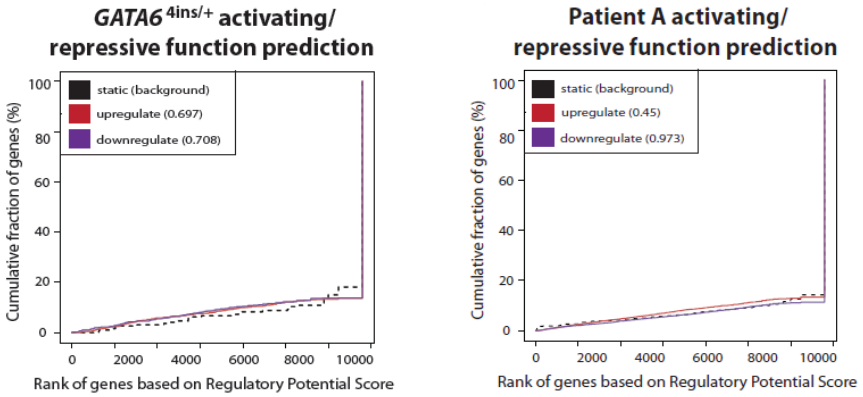


Figure 125. BETA output of activating/repressive function prediction of H9-derived *GATA6* $^{4ins/+}$ and Patient A mutant cells on day 12. The red and the purple lines represent the up-regulated and down-regulated genes, respectively. The dashed black line indicates the non-differentially expressed genes as background. Genes are cumulated by the rank on the basis of the regulatory potential score from high to low. *P*-values that represent the significance of the UP or DOWN group distributions are compared with the NON group by the Kolmogorov-Smirnov test.

Gene Ontology of genes bound and differenitally expressed in *GATA6* $^{4ins/+}$

Category	P-value	Gene symbol
Up-regulated in WT		
Regulation of insulin secretion	0.05181	STX1A, HNF4A, BRSK2, PDX1, C2CD2L
Regulation of heart contraction	5.70E-04	TNNT2, MYL4, SCN1B, ATP2B4, ADM, MYL3, NPPA, KCNK3, CACNA1B
Up-regulated in *GATA6* $^{4ins/+}$		
Skeletal system development	1.43E-09	LUM, TBX1, GLI2, FOXP1, HOXB4, HOXB1, DHRS3, OSR2, HOXB7, HOXB8, HOXB5, HOXC4, HOXB6, HOXC5, HOXB9, RARB, IGFBP4
Nervous system development	2.04E-06	LUM, GLI2, EPHB2, OVOL2, DYNLL2, HLX, PAFAH1B3, RANBP1, H2AFX, RARB, NR2F1, GBA, CENPF, ARID1A, TBX1, SHANK1, SHANK2, FOXP1, ATF5, HOXB1, DHFR, HOXB8, CLIC5, MNX1, EFNA5, KDM4A, LRP2, CALM1

Gene Ontology of genes bound and differenitally expressed in Patient A

Category	P-value	Gene symbol
Up-regulated in WT		
Endocrine system development	0.001505	HES1, GATA2, HNF1B, FGF8, APOA1, FOXA2, SALL1, MNX1, PAX6, DLL1, TBX1, PDX1, INSR
Endocrine pancreas development	0.003831	HES1, HNF1B, FOXA2, MNX1, PAX6, DLL1, PDX1
Up-regulated in Patient A		
Skeletal system development	4.45E-05	RBP4, HOXA13, LUM, PTH1R, SLC38A10, GLI2, FUZ, HOXC6, HOXC8, HOXC9, OSR2, CD44, HOXA5, HOXC4, HOXA6, RARB, GHR, HAPLN1
Nervous system development	3.24E-06	STIL, CCDC85C, STAR, PPP2R5D, HOXC8, PACSIN1, SMARCD3, GATA3, SMARCD1, S1PR5, DYNC2H1, H2AFX, ROBO2, RARB, MATN2, SCRT2

Figure 126. Enriched GO derived from BETA analysis showing developmental pathways. Tables are derived from direct target genes differentially expressed between H9* (WT), H9-derived *GATA6* $^{4ins/+}$ and Patient A mutant cells on day 12. *P*-values of the developmental pathways are indicated along with the respective genes.

CHAPTER 4 DISCUSSION

4.1. DE and pancreatic differentiation of hPSCs *in vitro* mimics developmental events during pancreatic formation in humans

Human PSCs offer a unique opportunity to study disease phenotypes not reproduced in model organisms such as the mouse. This is particularly relevant for my project where a discrepancy in genetics and the subsequent disease phenotype has been observed between mice and human. In this aspect, efficient generation of pancreatic cell types *in vitro* using hPSCs presents the first step toward successful disease modelling to potentially provide insights into the molecular mechanisms underlying pancreatic agenesis. Indeed, my results show that hPSCs can be efficiently differentiated into near homogenous populations of DE and pancreatic progenitor cells using several established defined culture systems. Importantly, the DE and pancreatic progenitor cells generated across these protocols follow a normal path of development, with the initial down-regulation of pluripotency genes such as *NANOG*, *OCT4* and *SOX2* followed by the up-regulation of DE markers *CXCR4* and *SOX17*, and the subsequent up-regulation of key pancreatic-specific genes such as *PDX1* and *NKX6-1*. In addition, GO analyses derived from RNA-seq show endoderm (Figure 106) and pancreatic (Figure 126) development among the top enriched pathways, further reinforcing these observations.

It has been well documented that hESCs and hiPSCs share many similar properties such as morphology, proliferation, gene expression, and the ability to differentiate into various cell types etc. (Takahashi et al., 2007, Takahashi and Yamanaka, 2006, Evans, 2011). However, variation in the efficiencies of differentiation has been reported between different hPSC lines (Osafune et al., 2008, Chin et al., 2009). Hence, it was not surprising to observe a difference in β-cell specification efficiency between the H9 and FSPS13.B cell lines despite their similar pancreatic progenitor specification efficiencies. The process of β-cell differentiation is controlled by a complex network involving tight regulation of genes required for the development of the pancreas. Naturally, culture conditions play a critical role in

determining the successful generation of pancreatic insulin-secreting β cells from hPSCs (Rostovskaya et al., 2015). Since the difference in β-cell specification efficiency between the H9 and FSPS13.B cell lines was observed using the same lab-derived protocol, variation in the efficiencies of differentiation is most likely explained by the different genetic backgrounds of the H9 and FSPS13.B cell lines.

The goal of deriving functional β-cells from hPSCs still remains a major challenge. Although substantial improvement has been made to differentiate hPSCs toward functional pancreatic β-cells, existing protocols for *in vitro* differentiation produce immature pancreatic β-cells that are not highly responsive to glucose stimulation. Pancreatic β-cell maturation is characterised by the ability of the differentiated β-cells to perform glucose-stimulated insulin secretion (GSIS). This challenge presents a hindrance to the use of hPSCs in applications such as disease modelling, where differentiation of hPSCs into mature, glucose responsive β-cells is required for establishing the disease phenotypes *in vitro*, and to understanding the molecular mechanisms underlying different forms of diabetes. While previously published protocols have shown an improvement in producing glucose-responsive insulin-secreting β-cells *in vivo*, the GSIS of the β-cells *in vitro* still remains limited, indicating an immature nature of these cells (Maehr et al., 2009, Zhang et al., 2009, Jiang et al., 2007a, D'Amour et al., 2006). Furthermore, a recent study has reported that insulin-secreting β cells differentiated from hPSCs are highly similar to human fetal pancreatic β-cells and do not resemble adult β-cells (Hrvatin et al., 2014). This could be one of the factors responsible for the inverse GSIS response that I observed with the H9 cells, where the immature nature of the cells impedes the cells from proper function.

Another factor that could have hindered the success of eliciting a GSIS response from β-cells *in vitro* is the low efficiency in producing insulin-secreting β-cells. With only approximately 10% and 6% of insulin-secreting β-cells generated from H9 or FSPS13.B cells respectively, the current lab-derived protocol is most likely lacking critical signals required for efficient generation of insulin-secreting β cells. Perhaps one solution to circumvent these problems is to adopt more recently

published protocols that report an increased efficiency in generating an average of 30-50% of insulin-producing cells (Rezania et al., 2014, Pagliuca et al., 2014).

4.2. *GATA6* and *GATA4* expression patterns during human pancreatic development

The expression pattern of *GATA6* and *GATA4* during human pancreas development has not been well characterised to date. My work has revealed the precise expression kinetics of *GATA6* and *GATA4* during *in vitro* differentiation into the pancreatic lineage.

That *GATA6* is expressed from the DE stage and remains expressed throughout pancreatic development suggests an important role of this transcription factor during human pancreas development. Interestingly, the expression pattern of *GATA4* is highly similar to that of *GATA6* in that it is not expressed in pluripotent cells and its expression is first observed in the DE stage and remains expressed throughout pancreatic development. This indicates a similar and possibly redundant role of both transcription factors. These findings confirm previous studies where both *GATA6* and *GATA4* have been reported to be expressed in DE cells in hPSC differentiation cultures (McLean et al., 2007, Vallier et al., 2009b). Similarly in mice, *Gata6* and *Gata4* are both expressed in the DE and its derivatives, including the pancreas (Decker et al., 2006, Watt et al., 2007).

My results show that *GATA4* levels are consistently more highly expressed than *GATA6* at the later stages of pancreatic development (D9 onwards), suggesting a critical role of *GATA4* in the development of the human pancreas. This is consistent with a previous report describing GATA4 expression at the onset of pancreatic development in human embryos, although it was unknown from the study whether GATA4 is also expressed in DE cells prior to pancreas formation (Jennings et al., 2013). It has also been reported that *GATA4* mutations are a rare cause of NDM and pancreatic agenesis in five patients harbouring *GATA4* mutations, confirming a role for *GATA4* in the development of the human pancreas (Shaw-Smith et al., 2014).

4.3. TALEN as a genome editing tool for disease modelling

The successful generation of TALEN-derived *GATA6* mutant hESC and hiPSC lines via both NHEJ and HR shows the versatility of TALENs as a genome editing tool for disease modelling. Interestingly, using the same TALEN cut sites, the cutting efficiency in H9 cells was observed to be higher than in FSPS13.B cells, possibly due to differences in nuclease cleavage efficiencies and/or intrinsic differences in activities of DNA repair pathways.

In addition to generating mutant lines, TALENs can also be used to correct mutations via homologous recombination. In the context of my project, it would be useful to correct the missense *GATA6* mutation in Patient A and Patient B to derive isogenic wild-type control cell lines as this would eliminate differences arising from different genetic backgrounds. However, this was not performed due to time constraints and was not prioritised because the patient phenotypes were similar to the TALEN-generated mutants which indicated the suitability of the TALEN-generated mutants as a disease modelling platform.

The similar differentiation efficiencies between the untargeted hPSCs (H9 and FPSP13.B) and their respective isogenic controls, which are targeted hPSCs but harbour no observable mutation around the target site (H9* and FPSP13.B*), indicated that off-target effects (if any) did not affect pancreatic specification. The normal karyotype displayed by all the targeted wild-type and mutant hPSC lines over many passages also suggested that the TALEN targeting did not introduce any gross chromosomal rearrangements and abnormalities such as deletions, inversions or translocations in addition to local mutations that can occur when imprecise repair of on- and off-target DNA cleavages take place (Lee et al., 2010, Lee et al., 2012, Park et al., 2014, Brunet et al., 2009, Cho et al., 2014).

The ability of the TALEN-derived mutant hPSC lines to retain pluripotency similarly to wild-type and untargeted hPSCs indicate that genome editing via TALENs did not affect the pluripotency status of hPSCs. As *GATA6* is not expressed in

undifferentiated cells, it is unlikely that this transcription factor plays an important role in pluripotency, which is evident from the non-effect that loss of *GATA6* has on pluripotency.

4.4. *GATA6* is required for DE specification

Using the TALEN-generated *GATA6* homozygous knockout H9 (*GATA6* $^{\Delta4/\Delta4}$ and *GATA6* $^{GFP/GFP}$) and FSPS13.B (*GATA6* $^{\Delta14/\Delta11}$) mutant cell lines, I have shown that *GATA6* is essential for the formation of the DE in humans. Although truncated proteins were detected in H9-*GATA6* $^{\Delta4/\Delta4}$ and FSPS13.B-*GATA6* $^{\Delta14/\Delta11}$ mutant cell lines, their phenotypes were indistinguishable from the H9-*GATA6* $^{GFP/GFP}$ mutant cell line that had no detectable GATA6 protein. This indicates that, consistent with its known biochemical characteristics (Bates et al., 2008, Molkentin, 2000), the truncated GATA6 protein lacking the C-terminal DNA-binding zinc-finger domains is non-functional. Thus, it can be inferred that in the absence of functional GATA6 proteins, a human embryo would likely fail to form the pancreas as a consequence to a primary defect in definitive endoderm formation. This finding, coupled with the rapid and strong up-regulation of *GATA6* at the DE stage, as well as its co-localisation with key DE marker *SOX17* strongly suggests that *GATA6* does indeed play an important role in DE formation.

Using the TALEN-generated *GATA6* heterozygous knockout H9 (*GATA6* $^{4ins/+}$ and *GATA6* $^{GFP/+}$) and FSPS13.B (*GATA6* $^{1ins/+}$, *GATA6* $^{\Delta14/+}$, *GATA6* $^{\Delta21\_8ins/+}$ and *GATA6* $^{GFP/+}$) mutant cell lines, as well as the patient-derived *GATA6* heterozygous mutant lines Patient A and Patient B, an impairment in DE formation was observed using the lab-derived protocol and STEMdiff pancreatic progenitor kit from SCT, with FSPS13.B-*GATA6* $^{GFP/+}$ mutant cells being the exception. However, a discrepancy was observed in cells differentiated via the PSC Definitive Endoderm Induction Kit from Life Technologies, where the DE formed as efficiently as the wild-type controls. Hence, it is possible that the DE phenotypes observed in the heterozygous mutants differentiated via the lab-derived protocol and STEMdiff pancreatic progenitor kit from SCT are a consequence of the DE differentiation protocol used. Unfortunately, I

am unable to verify this hypothesis through detailed comparison of the protocols by eliminating/adding certain growth factors due to the restrictions imposed by Life Technologies, where I was unable to obtain more information on the media formulation.

Furthermore, the fundamental differences in the differentiation protocols may underlie (or contribute to) the results I obtained and those recently published by Shi et al. and Tiyaboonchai et al. (Shi et al., 2017, Tiyaboonchai et al., 2017). For example, the growth factor and small molecule components as well as medium formulations differ substantially for the first three days of DE differentiation among the three studies. This was further evidenced in the study led by Tiyaboonchai where the group showed that a *GATA6* heterozygous iPSC line derived from an agenesis patient unexpectedly produced beta-like cells *in vitro* by simply reducing the concentration of retinoic acid 80-fold (Tiyaboonchai et al., 2017). This change led to statistically significantly fewer $PDX1^{+}$ cells from the patient line when compared to a wild-type iPSC line that showed negligible sensitivity to the same culture regime.

Comparing across the patient-derived and TALEN-derived H9 and FSPS13.B heterozygous mutants, a spectrum of DE phenotypes was observed using the lab-derived and SCT protocols; FSPS13.B-*GATA6* $^{GFP/+}$ mutant cells displayed no defect in DE formation, and FSPS13.B NHEJ-generated mutants as well as Patients A and B displayed a similar but a slightly more severe DE defected compared to H9-*GATA6* $^{4ins/+}$ and *GATA6* $^{GFP/+}$ mutant cells. This observation, however, was unsurprising as recent studies have reported that *GATA6* mutations can cause diverse diabetic phenotypes, ranging from pancreatic agenesis to adult-onset diabetes where most, but not all, patients display exocrine insufficiency requiring insulin treatment and enzyme replacement therapy, and other extrapancreatic features (De Franco et al., 2013, Bonnefond et al., 2012).

In the report by Bonnefond et al., two French sisters were described with the same *GATA6* allele (c.1504_1505del; p.Lys502Aspfs*5) but presented strikingly different clinical manifestations—one with permanent neonatal diabetes and the other without (Bonnefond et al., 2012). Similarly, Shi et al. (2017) engineered using CRISPR/Cas9 the common *GATA6* agenesis mutation c.1366C>T (p.Arg456Cys) in HUES8 cells—the same allele present in the patient A-derived iPSC line (*GATA6*$^{R456C/+}$)—and observe no heterozygous phenotype at the DE or pancreatic progenitor (PDX1^{+}) stages (Shi et al., 2017), whereas I do, both at the DE stage and beyond. A more recent publication by Yau et al. (2017) describes the inheritance of a novel *GATA6* frame-shift mutation (c.635_660del; p.Pro212fs) in three children from a *GATA6* mosaic mother, and each child (one is an abortus) presents a different phenotype (Yau et al., 2017).

By analogy, it is entirely possible that hiPSC derived from the patients in the Yau et al. (2017) and Bonnefond et al. (2012) studies would each behave entirely differently when differentiated *in vitro*. The simplest explanation for the existence of such "resilient individuals" who are not impacted by deleterious *GATA6* alleles is the influence of modifier genes and rare variants attributable to individual genetic backgrounds (Lek et al., 2016, Chen et al., 2016). *GATA4* is an obvious choice for a genetic modifier, given its expression in the DE, genetic interaction with *Gata6* in mice, the identification of rare *GATA4* heterozygous patients with pancreatic agenesis as well as our finding that *GATA4* is bound and regulated by GATA6 *in vitro* (Figure 113) (Morrisey et al., 1996b, Freyer et al., 2015, Shaw-Smith et al., 2014, D'Amato et al., 2010). Indeed, Shi et al. (2017) elegantly show dosage-dependent effects of *GATA4* alleles on phenotypes associated with *GATA6* heterozygosity during *in vitro* differentiation.

The observation of extrapancreatic abnormalities in *GATA6* patients, which include malformations in endodermal-derived organs such as congenital heart defects (Kodo et al., 2009, Lin et al., 2010), hepatobiliary malformations, gall bladder agenesis, and gut herniation (Allen et al., 2012), further provide evidence that *GATA6* plays an important role in the development of the DE. In addition, the two patients in the family studied in Yau et al. (2017) present defects in a number of endoderm-derived organs, further supporting that diminished GATA6 levels during DE formation underlie a constellation of clinical endodermal phenotypes (Yau et al., 2017). Indeed, when I ran pilot differentiations to specify H9* and H9-*GATA6* $^{4ins/+}$ mutant cells into the hepatic lineage, I observed decreased differentiation efficiencies for H9-*GATA6* $^{4ins/+}$ mutant cells to differentiate into hepatic progenitors, the precursors of hepatic cells (data not shown). Unfortunately, I was unable to successfully differentiate the wild-type H9 cells into mature hepatocytes as the hepatic differentiation protocol was still being optimised in the lab when the experiments were performed. As such, I was unable to perform phenotypic comparisons between H9* and H9-*GATA6* $^{4ins/+}$ mutant cells at a later hepatic development stage.

4.5. *GATA6* is required for pancreatic progenitor specification

My work has demonstrated that *GATA6* is required for pancreatic specification. Surprisingly, despite the broad spectrum of phenotypes observed at the DE stage, a less variable phenotype of 50-80% down-regulation of *PDX1* across all heterozygous mutants was seen at the pancreatic stage and this was consistently observed using all three DE specification protocols. This finding strongly indicates that the pancreatic phenotype seen in the *GATA6* heterozygous mutant cell lines is most likely a true effect of *GATA6* haploinsufficiency, thus establishing a human *in vitro* hPSC model system to study the role of *GATA6* in the development of the human pancreas. However, this present system has its limitations. As heterozygous *GATA6* mutations have been reported to have incomplete penetrance as displayed by different phenotypes in family members having identical mutations (Bonnefond et al., 2012), it is possible that my *in vitro* PSC model system is lacking intrinsic signalling pathways or factors present *in vivo* that may mitigate the negative effects of *GATA6* haploinsufficiency, thus driving a more severe phenotype.

The discordant phenotypes between mice and human models, especially for haploinsufficient disease genes, have been observed and widely discussed (Seidman and Seidman, 2002). In my hPSC model system, deleting one allele of *GATA6* impaired pancreatic formation as seen from the reduction of PDX1+ cells across all genetic backgrounds of the hPSCs used in this study. This finding demonstrates phenotypes not previously reported in mice (Morrisey et al., 1998, Carrasco et al., 2012, Xuan et al., 2012). For instance, pancreatic defects were not observed in *Gata6*$^{-/+}$ or *Gata4*$^{-/+}$ mouse embryos or adults. This suggests distinct gene dosage sensitivities between both species. Furthermore, the genetic background of *GATA6* patients is much more diverse than the inbred mouse strains. Thus, extreme phenotypes such as pancreatic agenesis may be seen in some, but not all, *GATA6* heterozygous patients. Lastly, the different timing where *GATA6* is deleted between the mice and human model systems could also contribute toward the discordant phenotypes between both systems. Due to early embryonic lethality of *Gata6* and *Gata4* embryos caused by extraembryonic defects (Morrisey et al., 1998), it was

necessary for *Gata6/4* to be conditionally inactivated using the Cre-LoxP system in the early pancreatic progenitors or the gut endoderm stages prior to pancreatic specification (Carrasco et al., 2012, Xuan et al., 2012). In contrast, in my *in vitro* hPSC model system, *GATA6* is deleted in pluripotent cells, before the initiation of differentiation. An inducible knockout system, such as a tetracycline-inducible shRNA or doxycycline-inducible CRISPR interference system, where *GATA6* can be inactivated at specific stages such as the gut endoderm could be a suitable method to replicate the mice model more closely (Bertero et al., 2016, Mandegar et al., 2016). Despite these differences, *GATA6/4* interactions are observed in both mice and hPSC model systems, supporting the use of both systems for investigating genetic and environmental modifiers in *GATA6*-linked human disease.

4.6. *GATA6* is a key regulator of DE and pancreatic progenitor specification

The molecular mechanism by which *GATA6* controls DE and pancreatic specification in humans was not known prior to my study. My results using genome-wide transcriptional analyses from RNA-seq revealed that *GATA6* transcriptionally activates the expression of endoderm markers in human. Interestingly, loss of both alleles of *GATA6* leads to an increase in mesoderm development. This suggests a possible role of *GATA6* not only to enable endoderm formation, but also to suppress mesoderm formation.

For the first time, results from ChIP-seq data suggest a direct molecular mechanism whereby GATA6 directly controls the gene expression of endoderm markers such as SOX17 and CXCR4, placing it centrally in the regulation of endoderm specification. The direct binding of GATA6 to GATA4 at both the DE and pancreatic progenitor stages also indicates that GATA6 and GATA4 are interacting partners, a finding that has also been previously reported in the developing and postnatal myocardium (Charron et al., 1999). Thus, the down-regulation of *GATA4* expression in the *GATA6* mutants suggests that *GATA6* is directly responsible for this observation. Interestingly, results from my ChIP-seq data also shows PDX1 as a binding partner of GATA6 (data not shown), a finding that was previously not shown in an earlier study (Cebola et al., 2015). This suggests that *GATA6* also plays a direct role in pancreatic specification.

Performing ChIP and ChIP-seq on Patients A and B could further identify important direct binding partners of GATA6. Since *GATA6* mutations in Patients A and B did not lead to nonsense-mediated decay of the GATA6 protein (Figure 52), it would be possible to perform ChIP on these samples. Thus, subsequent bioinformatics analyses comparing Patient A and B ChIP-seq datasets to their respective isogenic corrected wild-type control cell lines could further elucidate the molecular mechanisms of GATA6. However, this was not done due to time constraints.

Overlapping the GATA6, EOMES and SMAD2/3 ChIP-seq datasets at the DE level suggests a fundamental role of GATA6 at the DE stage, and that EOMES is required in the interaction of GATA6-SMAD2/3. Thus, as EOMES limits the expression of mesodermal markers, it can be speculated that GATA6 mutations may impede DNA binding of EOMES-SMAD2/3 linked to endoderm formation (Figure 127).

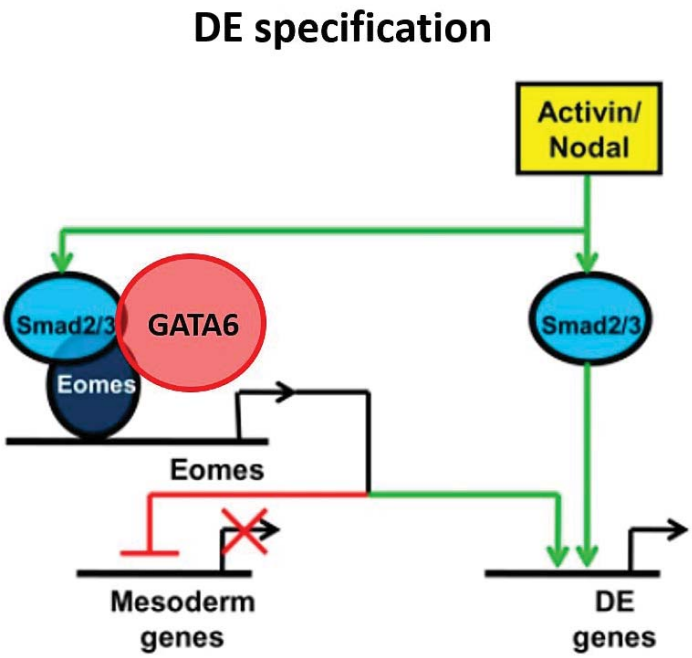


Figure 127. Model depicting the molecular mechanism of action for GATA6 in the formation of the DE. EOMES, SMAD2/3 and GATA6 interacts to initiate DE differentiation while repressing mesoderm genes. Illustration adapted and modified from Teo et al., 2011**.**

CHAPTER 5 FUTURE EXPERIMENTS

5.1. Unknown effects on penetrance of *GATA6* heterozygous mutants using various DE or pancreatic specification protocols

The discrepancy observed at the DE stage between different protocols suggests that the specification protocols used to differentiate the cells can have a significant impact on the DE phenotype observed. Hence, future work involving the identification of growth factors and/or signalling pathways that cause this discrepancy can be useful in further investigating the role of *GATA6* in the formation of the DE.

Moreover, it must be acknowledged that adherent differentiation fails to achieve the 3D complexity of human endoderm formation *in vivo*. Thus, future studies involving specification of pancreatic progenitors into functional and mature β-cells will likely benefit from 3D organoid systems which will more closely represent the *in vivo* environment of developing organs in humans, enabling the interactions between different pancreatic cell types and the interplay with possible niche signals.

With the availability of robust commercially available DE and pancreatic progenitor differentiation kits, perhaps studies of early pancreatic lineage commitment can be standardized intra- and inter-laboratory in an effort to minimize line-to-line and protocol-to-protocol differences.

5.2. Unexplored role of *GATA6* in other endoderm-derived organs

Although my work is focused on the pancreatic lineage, it is likely that *GATA6* haploinsufficiency plays a role in the development of other endoderm-derived organs such as the gall bladder, intestine and liver given that results from my study have established a critical role of *GATA6* in early endoderm formation. Indeed, preliminary results from my study where *GATA6* heterozygous mutants were specified toward the liver lineage have indicated that *GATA6* haploinsufficiency gives rise to defects in liver formation. The role of *GATA6* in these organs can be studied more closely using an inducible knockout system, such as a tetracycline-inducible system, where gene knockdown in hPSCs and even in differentiated cells can be rapid and tightly controlled, thus providing a unique opportunity for functional analyses in multiple cell types relevant for the study of human development.

5.3. Other possible roles of *GATA6*

Additional roles of *GATA6* such as whether *GATA6* haploinsufficiency can impair the proliferation and maintenance of pancreatic progenitors during their maturation into β-cells, or whether *GATA6* dosage may influence β-cell mass and function remain to be investigated. Utilising xenograft models such as grafting or transplanting β-like cells under the kidney capsules of mice can be used to further address these questions.

FINAL CONCLUSIONS

The transcription factor *GATA6* has recently been identified as the most common cause of pancreatic agenesis in humans. My work has revealed dosage-dependent requirements for *GATA6* in lineage specification leading to the formation of pancreatic progenitors and immature β-cells using hPSCs as an *in vitro* system for disease modelling. The similarities in DE and pancreatic phenotypes observed between TALEN-derived lines and patient-derived lines indicate the success in disease modelling using genome editing tools coupled with hPSCs, and establish the suitability of using genome editing tools such as TALENs in the study of human diseases. On the molecular level, *GATA6* directly regulates the development of the DE and pancreatic progenitors. Thus, this work provides evidence that *GATA6* is involved in the development of the human definitive endoderm and pancreas as well as the molecular mechanisms by which it regulates this developmental process.

BIBLIOGRAPHY

AHLGREN, U., JONSSON, J. & EDLUND, H. 1996. The morphogenesis of the pancreatic mesenchyme is uncoupled from that of the pancreatic epithelium in IPF1/PDX1-deficient mice. *Development,* 122**,** 1409-16.

ALLEN, H. L., FLANAGAN, S. E., SHAW-SMITH, C., DE FRANCO, E., AKERMAN, I., CASWELL, R., FERRER, J., HATTERSLEY, A. T. & ELLARD, S. 2012. GATA6 haploinsufficiency causes pancreatic agenesis in humans. *Nat Genet,* 44**,** 20-22.

ANDERSON, M. S. & BLUESTONE, J. A. 2005. The NOD mouse: a model of immune dysregulation. *Annu Rev Immunol,* 23**,** 447-85.

ANG, S. L., WIERDA, A., WONG, D., STEVENS, K. A., CASCIO, S., ROSSANT, J. & ZARET, K. S. 1993. The formation and maintenance of the definitive endoderm lineage in the mouse: involvement of HNF3/forkhead proteins. *Development,* 119**,** 1301-15.

ANTONARAKIS, S. E. & BECKMANN, J. S. 2006. Mendelian disorders deserve more attention. *Nat Rev Genet,* 7**,** 277-282.

APELQVIST, Å., AHLGREN, U. & EDLUND, H. 1997. Sonic hedgehog directs specialised mesoderm differentiation in the intestine and pancreas. *Current Biology,* 7**,** 801-804.

ARCECI, R. J., KING, A. A., SIMON, M. C., ORKIN, S. H. & WILSON, D. B. 1993. Mouse GATA-4: a retinoic acid-inducible GATA-binding transcription factor expressed in endodermally derived tissues and heart. *Mol Cell Biol,* 13**,** 2235-46.

ATKINSON, M. A. & EISENBARTH, G. S. 2001. Type 1 diabetes: new perspectives on disease pathogenesis and treatment. *Lancet,* 358**,** 221-9.

BAILEY, T., KRAJEWSKI, P., LADUNGA, I., LEFEBVRE, C., LI, Q., LIU, T., MADRIGAL, P., TASLIM, C. & ZHANG, J. 2013. Practical guidelines for the comprehensive analysis of ChIP-seq data. *PLoS Comput Biol,* 9**,** e1003326.

BATES, D. L., CHEN, Y., KIM, G., GUO, L. & CHEN, L. 2008. Crystal structures of multiple GATA zinc fingers bound to DNA reveal new insights into DNA recognition and self-association by GATA. *J Mol Biol,* 381**,** 1292-306.

BEDELL, V. M., WANG, Y., CAMPBELL, J. M., POSHUSTA, T. L., STARKER, C. G., KRUG II, R. G., TAN, W., PENHEITER, S. G., MA, A. C., LEUNG, A. Y. H., FAHRENKRUG, S. C., CARLSON, D. F., VOYTAS, D. F., CLARK, K. J., ESSNER, J. J. & EKKER, S. C. 2012. In vivo genome editing using a high-efficiency TALEN system. *Nature,* 491**,** 114-118.

BERNARDO, A. S., CHO, C. H., MASON, S., DOCHERTY, H. M., PEDERSEN, R. A., VALLIER, L. & DOCHERTY, K. 2009. Biphasic induction of Pdx1 in mouse and human embryonic stem cells can mimic development of pancreatic beta-cells. *Stem Cells,* 27**,** 341-51.

BERTERO, A., PAWLOWSKI, M., ORTMANN, D., SNIJDERS, K., YIANGOU, L., CARDOSO DE BRITO, M., BROWN, S., BERNARD, W. G., COOPER, J. D., GIACOMELLI, E., GAMBARDELLA, L., HANNAN, N. R., IYER, D., SAMPAZIOTIS, F., SERRANO, F., ZONNEVELD, M. C., SINHA, S., KOTTER, M. & VALLIER, L. 2016. Optimized inducible shRNA and CRISPR/Cas9 platforms for in vitro studies of human development using hPSCs. *Development,* 143**,** 4405-4418.

BIBIKOVA, M., BEUMER, K., TRAUTMAN, J. K. & CARROLL, D. 2003. Enhancing gene targeting with designed zinc finger nucleases. *Science,* 300**,** 764.

BIBIKOVA, M., CARROLL, D., SEGAL, D. J., TRAUTMAN, J. K., SMITH, J., KIM, Y. G. & CHANDRASEGARAN, S. 2001. Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol Cell Biol,* 21**,** 289-97.

BIBIKOVA, M., GOLIC, M., GOLIC, K. G. & CARROLL, D. 2002. Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. *Genetics,* 161**,** 1169-75.

BINGHAM, C., ELLARD, S., ALLEN, L., BULMAN, M., SHEPHERD, M., FRAYLING, T., BERRY, P. J., CLARK, P. M., LINDNER, T., BELL, G. I., RYFFEL, G. U., NICHOLLS, A. J. & HATTERSLEY, A. T. 2000. Abnormal nephron development associated with a frameshift mutation in the transcription factor hepatocyte nuclear factor-1 beta. *Kidney Int,* 57**,** 898-907.

BOCH, J. & BONAS, U. 2010. Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annu Rev Phytopathol,* 48**,** 419-36.

BOCH, J., SCHOLZE, H., SCHORNACK, S., LANDGRAF, A., HAHN, S., KAY, S., LAHAYE, T., NICKSTADT, A. & BONAS, U. 2009. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science,* 326**,** 1509-12.

BONAL, C. & HERRERA, P. L. 2008. Genes controlling pancreas ontogeny. *Int J Dev Biol,* 52**,** 823-35.

BONNEFOND, A., SAND, O., GUERIN, B., DURAND, E., DE GRAEVE, F., HUYVAERT, M., RACHDI, L., KERR-CONTE, J., PATTOU, F., VAXILLAIRE, M., POLAK, M., SCHARFMANN, R., CZERNICHOW, P. & FROGUEL, P. 2012. GATA6 inactivating mutations are associated with heart defects and, inconsistently, with pancreatic agenesis and diabetes. *Diabetologia,* 55**,** 2845-7.

BOSSARD, P. & ZARET, K. S. 1998. GATA transcription factors as potentiators of gut endoderm differentiation. *Development,* 125**,** 4909-17.

BOWLES, J., SCHEPERS, G. & KOOPMAN, P. 2000. Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. *Dev Biol,* 227**,** 239-55.

BREWER, A., GOVE, C., DAVIES, A., MCNULTY, C., BARROW, D., KOUTSOURAKIS, M., FARZANEH, F., PIZZEY, J., BOMFORD, A. & PATIENT, R. 1999. The Human and Mouse GATA-6 Genes Utilize Two Promoters and Two Initiation Codons. *Journal of Biological Chemistry,* 274**,** 38004-38016.

BRIGGS, A. W., RIOS, X., CHARI, R., YANG, L., ZHANG, F., MALI, P. & CHURCH, G. M. 2012. Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. *Nucleic Acids Res,* 40**,** e117.

BRISSOVA, M., FOWLER, M. J., NICHOLSON, W. E., CHU, A., HIRSHBERG, B., HARLAN, D. M. & POWERS, A. C. 2005. Assessment of human pancreatic islet architecture and composition by laser scanning confocal microscopy. *J Histochem Cytochem,* 53**,** 1087-97.

BROWN, S., TEO, A., PAUKLIN, S., HANNAN, N., CHO, C. H., LIM, B., VARDY, L., DUNN, N. R., TROTTER, M., PEDERSEN, R. & VALLIER, L. 2011. Activin/Nodal signaling controls divergent transcriptional networks in human embryonic stem cells and in endoderm progenitors. *Stem Cells,* 29**,** 1176-85.

BRUNET, E., SIMSEK, D., TOMISHIMA, M., DEKELVER, R., CHOI, V. M., GREGORY, P., URNOV, F., WEINSTOCK, D. M. & JASIN, M. 2009. Chromosomal translocations induced at specified loci in human stem cells. *Proc Natl Acad Sci U S A,* 106**,** 10620-5.

CABRERA, O., BERMAN, D. M., KENYON, N. S., RICORDI, C., BERGGREN, P. O. & CAICEDO, A. 2006. The unique cytoarchitecture of human pancreatic islets has implications for islet cell function. *Proc Natl Acad Sci U S A,* 103**,** 2334-9.

CAI, J., YU, C., LIU, Y., CHEN, S., GUO, Y., YONG, J., LU, W., DING, M. & DENG, H. 2010. Generation of homogeneous PDX1(+) pancreatic progenitors from human ES cell-derived endoderm cells. *J Mol Cell Biol,* 2**,** 50-60.

CANO, D. A., SORIA, B., MARTIN, F. & ROJAS, A. 2014. Transcriptional control of mammalian pancreas organogenesis. *Cell Mol Life Sci,* 71**,** 2383-402.

CARBON, S., IRELAND, A., MUNGALL, C. J., SHU, S., MARSHALL, B., LEWIS, S., THE AMI, G. O. H. & THE WEB PRESENCE WORKING, G. 2009. AmiGO: online access to ontology and annotation data. *Bioinformatics,* 25**,** 288-289.

CARRASCO, M., DELGADO, I., SORIA, B., MARTIN, F. & ROJAS, A. 2012. GATA4 and GATA6 control mouse pancreas organogenesis. *J Clin Invest,* 122**,** 3504-15.

CASTAING, M., DUVILLIE, B., QUEMENEUR, E., BASMACIOGULLARI, A. & SCHARFMANN, R. 2005. Ex vivo analysis of acinar and endocrine cell development in the human embryonic pancreas. *Dev Dyn,* 234**,** 339-45.

CAVENER, D. R. & RAY, S. C. 1991. Eukaryotic start and stop translation sites. *Nucleic Acids Research,* 19**,** 3185-3192.

CEBOLA, I., RODRIGUEZ-SEGUI, S. A., CHO, C. H., BESSA, J., ROVIRA, M., LUENGO, M., CHHATRIWALA, M., BERRY, A., PONSA-COBAS, J., MAESTRO, M. A., JENNINGS, R. E., PASQUALI, L., MORAN, I., CASTRO, N., HANLEY, N. A., GOMEZ-SKARMETA, J. L., VALLIER, L. & FERRER, J. 2015. TEAD and YAP regulate the enhancer network of human embryonic pancreatic progenitors. *Nat Cell Biol,* 17**,** 615-26.

CERMAK, T., DOYLE, E. L., CHRISTIAN, M., WANG, L., ZHANG, Y., SCHMIDT, C., BALLER, J. A., SOMIA, N. V., BOGDANOVE, A. J. & VOYTAS, D. F. 2011. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Research,* 39**,** e82-e82.

CHAMPERIS TSANIRAS, S. & JONES, P. M. 2010. Generating pancreatic beta-cells from embryonic stem cells by manipulating signaling pathways. *J Endocrinol,* 206**,** 13-26.

CHARRON, F., PARADIS, P., BRONCHAIN, O., NEMER, G. & NEMER, M. 1999. Cooperative interaction between GATA-4 and GATA-6 regulates myocardial gene expression. *Mol Cell Biol,* 19**,** 4355-65.

CHEN, R., SHI, L., HAKENBERG, J., NAUGHTON, B., SKLAR, P., ZHANG, J., ZHOU, H., TIAN, L., PRAKASH, O., LEMIRE, M., SLEIMAN, P., CHENG, W. Y., CHEN, W., SHAH, H., SHEN, Y., FROMER, M., OMBERG, L., DEARDORFF, M. A., ZACKAI, E., BOBE, J. R., LEVIN, E., HUDSON, T. J., GROOP, L., WANG, J., HAKONARSON, H., WOJCICKI, A., DIAZ, G. A., EDELMANN, L., SCHADT, E. E. & FRIEND, S. H. 2016. Analysis of 589,306 genomes identifies individuals resilient to severe Mendelian childhood diseases. *Nat Biotechnol,* 34**,** 531-8.

CHIN, M. H., MASON, M. J., XIE, W., VOLINIA, S., SINGER, M., PETERSON, C., AMBARTSUMYAN, G., AIMIUWU, O., RICHTER, L., ZHANG, J., KHVOROSTOV, I., OTT, V., GRUNSTEIN, M., LAVON, N., BENVENISTY, N., CROCE, C. M., CLARK, A. T., BAXTER, T., PYLE, A. D., TEITELL, M. A., PELEGRINI, M., PLATH, K. & LOWRY, W. E. 2009. Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell,* 5**,** 111-23.

CHO, C. H., HANNAN, N. R., DOCHERTY, F. M., DOCHERTY, H. M., JOAO LIMA, M., TROTTER, M. W., DOCHERTY, K. & VALLIER, L. 2012. Inhibition of activin/nodal signalling is necessary for pancreatic differentiation of human pluripotent stem cells. *Diabetologia,* 55**,** 3284-95.

CHO, S. W., KIM, S., KIM, Y., KWEON, J., KIM, H. S., BAE, S. & KIM, J. S. 2014. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res,* 24**,** 132-41.

CHOULIKA, A., PERRIN, A., DUJON, B. & NICOLAS, J. F. 1995. Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae. *Mol Cell Biol,* 15**,** 1968-73.

CHRISTIAN, M., CERMAK, T., DOYLE, E. L., SCHMIDT, C., ZHANG, F., HUMMEL, A., BOGDANOVE, A. J. & VOYTAS, D. F. 2010. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics,* 186**,** 757-61.

CIRILLO, L. A., LIN, F. R., CUESTA, I., FRIEDMAN, D., JARNIK, M. & ZARET, K. S. 2002. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell,* 9**,** 279-89.

CLEMENTS, D. & WOODLAND, H. R. 2000. Changes in embryonic cell fate produced by expression of an endodermal transcription factor, Xsox17. *Mech Dev,* 99**,** 65-70.

COFFINIER, C., THEPOT, D., BABINET, C., YANIV, M. & BARRA, J. 1999. Essential role for the homeoprotein vHNF1/HNF1beta in visceral endoderm differentiation. *Development,* 126**,** 4785-94.

CONG, L., RAN, F. A., COX, D., LIN, S., BARRETTO, R., HABIB, N., HSU, P. D., WU, X., JIANG, W., MARRAFFINI, L. A. & ZHANG, F. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science,* 339**,** 819-23.

CONG, L., ZHOU, R., KUO, Y. C., CUNNIFF, M. & ZHANG, F. 2012. Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat Commun,* 3**,** 968.

D'AMATO, E., GIACOPELLI, F., GIANNATTASIO, A., D'ANNUNZIO, G., BOCCIARDI, R., MUSSO, M., LORINI, R. & RAVAZZOLO, R. 2010. Genetic investigation in an Italian child with an unusual association of atrial septal defect, attributable to a new familial GATA4 gene mutation, and neonatal diabetes due to pancreatic agenesis. *Diabet Med,* 27**,** 1195-200.

D'AMOUR, K. A., AGULNICK, A. D., ELIAZER, S., KELLY, O. G., KROON, E. & BAETGE, E. E. 2005. Efficient differentiation of human embryonic stem cells to definitive endoderm. *Nat Biotechnol,* 23**,** 1534-41.

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. & BAETGE, E. E. 2006. Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells. *Nat Biotechnol,* 24**,** 1392-401.

DE FRANCO, E., FLANAGAN, S. E., HOUGHTON, J. A., LANGO ALLEN, H., MACKAY, D. J., TEMPLE, I. K., ELLARD, S. & HATTERSLEY, A. T. 2015. The effect of early, comprehensive genomic testing on clinical care in neonatal diabetes: an international cohort study. *Lancet,* 386**,** 957-63.

DE FRANCO, E., SHAW-SMITH, C., FLANAGAN, S. E., SHEPHERD, M. H., HATTERSLEY, A. T. & ELLARD, S. 2013. GATA6 mutations cause a broad phenotypic spectrum of diabetes from pancreatic agenesis to adult-onset diabetes without exocrine insufficiency. *Diabetes,* 62**,** 993-7.

DECKER, K., GOLDMAN, D. C., GRASCH, C. L. & SUSSEL, L. 2006. Gata6 is an important regulator of mouse pancreas development. *Dev Biol,* 298**,** 415-29.

DOBIN, A., DAVIS, C. A., SCHLESINGER, F., DRENKOW, J., ZALESKI, C., JHA, S., BATUT, P., CHAISSON, M. & GINGERAS, T. R. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics,* 29**,** 15-21.

DORFMAN, D. M., WILSON, D. B., BRUNS, G. A. & ORKIN, S. H. 1992. Human transcription factor GATA-2. Evidence for regulation of preproendothelin-1 gene expression in endothelial cells. *J Biol Chem,* 267**,** 1279-85.

DOYLE, E. L., BOOHER, N. J., STANDAGE, D. S., VOYTAS, D. F., BRENDEL, V. P., VANDYK, J. K. & BOGDANOVE, A. J. 2012. TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic Acids Research,* 40**,** W117-W122.

DUFORT, D., SCHWARTZ, L., HARPAL, K. & ROSSANT, J. 1998. The transcription factor HNF3beta is required in visceral endoderm for normal primitive streak morphogenesis. *Development,* 125**,** 3015-25.

EDGHILL, E. L., DIX, R. J., FLANAGAN, S. E., BINGLEY, P. J., HATTERSLEY, A. T., ELLARD, S. & GILLESPIE, K. M. 2006. HLA genotyping supports a nonautoimmune etiology in patients diagnosed with diabetes under the age of 6 months. *Diabetes,* 55**,** 1895-8.

ENGLER, C., GRUETZNER, R., KANDZIA, R. & MARILLONNET, S. 2009. Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS One,* 4**,** e5553.

ENGLER, C., KANDZIA, R. & MARILLONNET, S. 2008. A one pot, one step, precision cloning method with high throughput capability. *PLoS One,* 3**,** e3647.

ESNI, F., GHOSH, B., BIANKIN, A. V., LIN, J. W., ALBERT, M. A., YU, X., MACDONALD, R. J., CIVIN, C. I., REAL, F. X., PACK, M. A., BALL, D. W. & LEACH, S. D. 2004. Notch inhibits Ptf1 function and acinar cell differentiation in developing mouse and zebrafish pancreas. *Development,* 131**,** 4213.

EVANS, M. 2011. Discovering pluripotency: 30 years of mouse embryonic stem cells. *Nat Rev Mol Cell Biol,* 12**,** 680-6.

EVANS, M. J. & KAUFMAN, M. H. 1981. Establishment in culture of pluripotential cells from mouse embryos. *Nature,* 292**,** 154-156.

EVANS, T. & FELSENFELD, G. 1989. The erythroid-specific transcription factor Eryf1: a new finger protein. *Cell,* 58**,** 877-85.

FAJANS, S. S., BELL, G. I. & POLONSKY, K. S. 2001. Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young. *N Engl J Med,* 345**,** 971-80.

FRAYLING, T. M., EVANS, J. C., BULMAN, M. P., PEARSON, E., ALLEN, L., OWEN, K., BINGHAM, C., HANNEMANN, M., SHEPHERD, M., ELLARD, S. & HATTERSLEY, A. T. 2001. beta-cell genes and diabetes: molecular and clinical characterization of mutations in transcription factors. *Diabetes,* 50 Suppl 1**,** S94-100.

FREYER, L., SCHROTER, C., SAIZ, N., SCHRODE, N., NOWOTSCHIN, S., MARTINEZ-ARIAS, A. & HADJANTONAKIS, A. K. 2015. A loss-of-function and H2B-Venus transcriptional reporter allele for Gata6 in mice. *BMC Dev Biol,* 15**,** 38.

GAO, N., LELAY, J., VATAMANIUK, M. Z., RIECK, S., FRIEDMAN, J. R. & KAESTNER, K. H. 2008. Dynamic regulation of Pdx1 enhancers by Foxa1 and Foxa2 is essential for pancreas development. *Genes Dev,* 22**,** 3435-48.

GARG, V., KATHIRIYA, I. S., BARNES, R., SCHLUTERMAN, M. K., KING, I. N., BUTLER, C. A., ROTHROCK, C. R., EAPEN, R. S., HIRAYAMA-YAMADA, K., JOO, K., MATSUOKA, R., COHEN, J. C. & SRIVASTAVA, D. 2003. GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature,* 424**,** 443-7.

GRADWOHL, G., DIERICH, A., LEMEUR, M. & GUILLEMOT, F. 2000. neurogenin3 is required for the development of the four endocrine cell lineages of the pancreas. *Proceedings of the National Academy of Sciences of the United States of America,* 97**,** 1607-1611.

GU, G., DUBAUSKAITE, J. & MELTON, D. A. 2002. Direct evidence for the pancreatic lineage: NGN3+ cells are islet progenitors and are distinct from duct progenitors. *Development,* 129**,** 2447-57.

GUO, S., DAI, C., GUO, M., TAYLOR, B., HARMON, J. S., SANDER, M., ROBERTSON, R. P., POWERS, A. C. & STEIN, R. 2013. Inactivation of specific beta cell transcription factors in type 2 diabetes. *J Clin Invest,* 123**,** 3305-16.

GURDON, J. B. & MELTON, D. A. 2008. Nuclear reprogramming in cells. *Science,* 322**,** 1811-5.

GUZ, Y., MONTMINY, M. R., STEIN, R., LEONARD, J., GAMER, L. W., WRIGHT, C. V. & 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**,** 11-8.

HAUMAITRE, C., BARBACCI, E., JENNY, M., OTT, M. O., GRADWOHL, G. & CEREGHINI, S. 2005. Lack of TCF2/vHNF1 in mice leads to pancreas agenesis. *Proc Natl Acad Sci U S A,* 102**,** 1490-5.

HEBROK, M., KIM, S. K. & MELTON, D. A. 1998. Notochord repression of endodermal Sonic hedgehog permits pancreas development. *Genes & Development,* 12**,** 1705-1713.

HEBROK, M., KIM, S. K., ST JACQUES, B., MCMAHON, A. P. & MELTON, D. A. 2000. Regulation of pancreas development by hedgehog signaling. *Development,* 127**,** 4905-4913.

HENSELEIT, K. D., NELSON, S. B., KUHLBRODT, K., HENNINGS, J. C., ERICSON, J. & SANDER, M. 2005. NKX6 transcription factor activity is required for alpha- and beta-cell development in the pancreas. *Development,* 132**,** 3139-49.

HIROYUKI, S. & SUSUMU, K. 1981. New restriction endonucleases from Flavobacterium okeanokoites (FokI) and Micrococcus luteus (MluI). *Gene,* 16**,** 73-78.

HO, I. C., VORHEES, P., MARIN, N., OAKLEY, B. K., TSAI, S. F., ORKIN, S. H. & LEIDEN, J. M. 1991. Human GATA-3: a lineage-restricted transcription factor that regulates the expression of the T cell receptor alpha gene. *EMBO J,* 10**,** 1187-92.

HORIKAWA, Y., IWASAKI, N., HARA, M., FURUTA, H., HINOKIO, Y., COCKBURN, B. N., LINDNER, T., YAMAGATA, K., OGATA, M., TOMONAGA, O., KUROKI, H., KASAHARA, T., IWAMOTO, Y. & BELL, G. I. 1997. Mutation in hepatocyte nuclear factor-1 beta gene (TCF2) associated with MODY. *Nat Genet,* 17**,** 384-5.

HRVATIN, S., O'DONNELL, C. W., DENG, F., MILLMAN, J. R., PAGLIUCA, F. W., DIIORIO, P., REZANIA, A., GIFFORD, D. K. & MELTON, D. A. 2014. Differentiated human stem cells resemble fetal, not adult, beta cells. *Proc Natl Acad Sci U S A,* 111**,** 3038-43.

HUANG DA, W., SHERMAN, B. T. & LEMPICKI, R. A. 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res,* 37**,** 1-13.

HUANG DA, W., SHERMAN, B. T. & LEMPICKI, R. A. 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc,* 4**,** 44-57.

HUANG, P., XIAO, A., ZHOU, M., ZHU, Z., LIN, S. & ZHANG, B. 2011. Heritable gene targeting in zebrafish using customized TALENs. *Nat Biotech,* 29**,** 699-700.

HUDSON, C., CLEMENTS, D., FRIDAY, R. V., STOTT, D. & WOODLAND, H. R. 1997. Xsox17alpha and -beta mediate endoderm formation in Xenopus. *Cell,* 91**,** 397-405.

HUGGON, I. C., DAVIES, A., GOVE, C., MOSCOSO, G., MONIZ, C., FOSS, Y., FARZANEH, F. & TOWNER, P. 1997. Molecular cloning of human GATA-6 DNA binding protein: high levels of expression in heart and gut. *Biochim Biophys Acta,* 1353**,** 98-102.

IAFUSCO, D., STAZI, M. A., COTICHINI, R., COTELLESSA, M., MARTINUCCI, M. E., MAZZELLA, M., CHERUBINI, V., BARBETTI, F., MARTINETTI, M., CERUTTI, F. & PRISCO, F. 2002. Permanent diabetes mellitus in the first year of life. *Diabetologia,* 45**,** 798-804.

JACKSON, D. A., SYMONS, R. H. & BERG, P. 1972. Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of Escherichia coli. *Proc Natl Acad Sci U S A,* 69**,** 2904-9.

JACKSON, S. P. & BARTEK, J. 2009. The DNA-damage response in human biology and disease. *Nature,* 461**,** 1071-8.

JENNINGS, R. E., BERRY, A. A., KIRKWOOD-WILSON, R., ROBERTS, N. A., HEARN, T., SALISBURY, R. J., BLAYLOCK, J., PIPER HANLEY, K. & HANLEY, N. A. 2013. Development of the Human Pancreas From Foregut to Endocrine Commitment. *Diabetes,* 62**,** 3514-3522.

JENNINGS, R. E., BERRY, A. A., STRUTT, J. P., GERRARD, D. T. & HANLEY, N. A. 2015. Human pancreas development. *Development,* 142**,** 3126-3137.

JEON, J., CORREA-MEDINA, M., RICORDI, C., EDLUND, H. & DIEZ, J. A. 2009. Endocrine cell clustering during human pancreas development. *J Histochem Cytochem,* 57**,** 811-24.

JIANG, J., AU, M., LU, K., ESHPETER, A., KORBUTT, G., FISK, G. & MAJUMDAR, A. S. 2007a. Generation of insulin-producing islet-like clusters from human embryonic stem cells. *Stem Cells,* 25**,** 1940-53.

JIANG, W., SHI, Y., ZHAO, D., CHEN, S., YONG, J., ZHANG, J., QING, T., SUN, X., ZHANG, P., DING, M., LI, D. & DENG, H. 2007b. In vitro derivation of functional insulin-producing cells from human embryonic stem cells. *Cell Res,* 17**,** 333-44.

JONSSON, J., CARLSSON, L., EDLUND, T. & EDLUND, H. 1994. Insulin-promoter-factor 1 is required for pancreas development in mice. *Nature,* 371**,** 606-9.

JOULIN, V., BORIES, D., ELÉOUET, J. F., LABASTIE, M. C., CHRÉTIEN, S., MATTÉI, M. G. & ROMÉO, P. H. 1991. A T-cell specific TCR delta DNA binding protein is a member of the human GATA family. *The EMBO Journal,* 10**,** 1809-1816.

JOUNG, J. K. & SANDER, J. D. 2013. TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol,* 14**,** 49-55.

KANAI-AZUMA, M., KANAI, Y., GAD, J. M., TAJIMA, Y., TAYA, C., KUROHMARU, M., SANAI, Y., YONEKAWA, H., YAZAKI, K., TAM, P. P. & HAYASHI, Y. 2002. Depletion of definitive gut endoderm in Sox17-null mutant mice. *Development,* 129**,** 2367-79.

KANAKATTI SHANKAR, R., PIHOKER, C., DOLAN, L. M., STANDIFORD, D., BADARU, A., DABELEA, D., RODRIGUEZ, B., BLACK, M. H., IMPERATORE, G., HATTERSLEY, A., ELLARD, S. & GILLIAM, L. K. 2013. Permanent neonatal diabetes mellitus: prevalence and genetic diagnosis in the SEARCH for Diabetes in Youth Study. *Pediatr Diabetes,* 14**,** 174-80.

KAWAGUCHI, Y., COOPER, B., GANNON, M., RAY, M., MACDONALD, R. J. & WRIGHT, C. V. 2002. The role of the transcriptional regulator Ptf1a in converting intestinal to pancreatic progenitors. *Nat Genet,* 32**,** 128-34.

KELLEY, C., BLUMBERG, H., ZON, L. I. & EVANS, T. 1993. GATA-4 is a novel transcription factor expressed in endocardium of the developing heart. *Development,* 118**,** 817-27.

KETOLA, I., OTONKOSKI, T., PULKKINEN, M. A., NIEMI, H., PALGI, J., JACOBSEN, C. M., WILSON, D. B. & HEIKINHEIMO, M. 2004. Transcription factor GATA-6 is expressed in the endocrine and GATA-4 in the exocrine pancreas. *Mol Cell Endocrinol,* 226**,** 51-7.

KIM, D., PERTEA, G., TRAPNELL, C., PIMENTEL, H., KELLEY, R. & SALZBERG, S. L. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol,* 14**,** R36.

KIM, S. K. & MELTON, D. A. 1998. Pancreas development is promoted by cyclopamine, a Hedgehog signaling inhibitor. *Proceedings of the National Academy of Sciences of the United States of America,* 95**,** 13036-13041.

KIM, Y. G., CHA, J. & CHANDRASEGARAN, S. 1996. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A,* 93**,** 1156-60.

KO, L. J. & ENGEL, J. D. 1993. DNA-binding specificities of the GATA transcription factor family. *Molecular and Cellular Biology,* 13**,** 4011-4022.

KO, L. J., YAMAMOTO, M., LEONARD, M. W., GEORGE, K. M., TING, P. & ENGEL, J. D. 1991. Murine and human T-lymphocyte GATA-3 factors mediate transcription through a cis-regulatory element within the human T-cell receptor delta gene enhancer. *Mol Cell Biol,* 11**,** 2778-84.

KODO, K., NISHIZAWA, T., FURUTANI, M., ARAI, S., YAMAMURA, E., JOO, K., TAKAHASHI, T., MATSUOKA, R. & YAMAGISHI, H. 2009. GATA6 mutations cause human cardiac outflow tract defects by disrupting semaphorin-plexin signaling. *Proc Natl Acad Sci U S A,* 106**,** 13933-8.

KOUTSOURAKIS, M., LANGEVELD, A., PATIENT, R., BEDDINGTON, R. & GROSVELD, F. 1999. The transcription factor GATA6 is essential for early extraembryonic development. *Development,* 126**,** 723-32.

KOZAK, M. 1981. Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. *Nucleic Acids Research,* 9**,** 5233-5252.

KOZAK, M. 1997. Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 and +6. *The EMBO Journal,* 16**,** 2482.

KRAPP, A., KNOFLER, M., LEDERMANN, B., BURKI, K., BERNEY, C., ZOERKLER, N., HAGENBUCHLE, O. & 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**,** 3752-63.

KROON, E., MARTINSON, L. A., KADOYA, K., BANG, A. G., KELLY, O. G., ELIAZER, S., YOUNG, H., RICHARDSON, M., SMART, N. G., CUNNINGHAM, J., AGULNICK, A. D., D'AMOUR, K. A., CARPENTER, M. K. & BAETGE, E. E. 2008. Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells in vivo. *Nat Biotechnol,* 26**,** 443-52.

KUBO, A., SHINOZAKI, K., SHANNON, J. M., KOUSKOFF, V., KENNEDY, M., WOO, S., FEHLING, H. J. & KELLER, G. 2004. Development of definitive endoderm from embryonic stem cells in culture. *Development,* 131**,** 1651-62.

LAMMERT, E., CLEAVER, O. & MELTON, D. 2001. Induction of Pancreatic Differentiation by Signals from Blood Vessels. *Science,* 294**,** 564-567.

LANDT, S. G., MARINOV, G. K., KUNDAJE, A., KHERADPOUR, P., PAULI, F., BATZOGLOU, S., BERNSTEIN, B. E., BICKEL, P., BROWN, J. B., CAYTING, P., CHEN, Y., DESALVO, G., EPSTEIN, C., FISHER-AYLOR, K. I., EUSKIRCHEN, G., GERSTEIN, M., GERTZ, J., HARTEMINK, A. J., HOFFMAN, M. M., IYER, V. R., JUNG, Y. L., KARMAKAR, S., KELLIS, M., KHARCHENKO, P. V., LI, Q., LIU, T., LIU, X. S., MA, L., MILOSAVLJEVIC, A., MYERS, R. M., PARK, P. J., PAZIN, M. J., PERRY, M. D., RAHA, D., REDDY, T. E., ROZOWSKY, J., SHORESH, N., SIDOW, A., SLATTERY, M., STAMATOYANNOPOULOS, J. A.,

TOLSTORUKOV, M. Y., WHITE, K. P., XI, S., FARNHAM, P. J., LIEB, J. D., WOLD, B. J. & SNYDER, M. 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res,* 22**,** 1813-31.

LANGO ALLEN, H., FLANAGAN, S. E., SHAW-SMITH, C., DE FRANCO, E., AKERMAN, I., CASWELL, R., FERRER, J., HATTERSLEY, A. T. & ELLARD, S. 2012. GATA6 haploinsufficiency causes pancreatic agenesis in humans. *Nat Genet,* 44**,** 20-2.

LAVERRIERE, A. C., MACNEILL, C., MUELLER, C., POELMANN, R. E., BURCH, J. B. & EVANS, T. 1994. GATA-4/5/6, a subfamily of three transcription factors transcribed in developing heart and gut. *J Biol Chem,* 269**,** 23177-84.

LEDERMANN, H. M. 1995. Is maturity onset diabetes at young age (MODY) more common in Europe than previously assumed? *Lancet,* 345**,** 648.

LEE, H. J., KIM, E. & KIM, J. S. 2010. Targeted chromosomal deletions in human cells using zinc finger nucleases. *Genome Res,* 20**,** 81-9.

LEE, H. J., KWEON, J., KIM, E., KIM, S. & KIM, J. S. 2012. Targeted chromosomal duplications and inversions in the human genome using zinc finger nucleases. *Genome Res,* 22**,** 539-48.

LEE, M. E., TEMIZER, D. H., CLIFFORD, J. A. & QUERTERMOUS, T. 1991. Cloning of the GATA-binding protein that regulates endothelin-1 gene expression in endothelial cells. *J Biol Chem,* 266**,** 16188-92.

LEK, M., KARCZEWSKI, K. J., MINIKEL, E. V., SAMOCHA, K. E., BANKS, E., FENNELL, T., O'DONNELL-LURIA, A. H., WARE, J. S., HILL, A. J., CUMMINGS, B. B., TUKIAINEN, T., BIRNBAUM, D. P., KOSMICKI, J. A., DUNCAN, L. E., ESTRADA, K., ZHAO, F., ZOU, J., PIERCE-HOFFMAN, E., BERGHOUT, J., COOPER, D. N., DEFLAUX, N., DEPRISTO, M., DO, R., FLANNICK, J., FROMER, M., GAUTHIER, L., GOLDSTEIN, J., GUPTA, N., HOWRIGAN, D., KIEZUN, A., KURKI, M. I., MOONSHINE, A. L., NATARAJAN, P., OROZCO, L., PELOSO, G. M., POPLIN, R., RIVAS, M. A., RUANO-RUBIO, V., ROSE, S. A., RUDERFER, D. M., SHAKIR, K., STENSON, P. D., STEVENS, C., THOMAS, B. P., TIAO, G., TUSIE-LUNA, M. T., WEISBURD, B., WON, H. H., YU, D., ALTSHULER, D. M., ARDISSINO, D., BOEHNKE, M., DANESH, J., DONNELLY, S., ELOSUA, R., FLOREZ, J. C., GABRIEL, S. B., GETZ, G., GLATT, S. J., HULTMAN, C. M., KATHIRESAN, S., LAAKSO, M., MCCARROLL, S., MCCARTHY, M. I., MCGOVERN, D., MCPHERSON, R., NEALE, B. M., PALOTIE, A., PURCELL, S. M., SALEHEEN, D., SCHARF, J. M., SKLAR, P., SULLIVAN, P. F., TUOMILEHTO, J., TSUANG, M. T., WATKINS, H. C., WILSON, J. G., DALY, M. J., MACARTHUR, D. G. & EXOME AGGREGATION, C. 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature,* 536**,** 285-91.

LI, H. & DURBIN, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics,* 25**,** 1754-60.

LI, Q., BROWN, J. B., HUANG, H. & BICKEL, P. J. 2011. Measuring reproducibility of high-throughput experiments. 1752-1779.

LIAO, Y., SMYTH, G. K. & SHI, W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics,* 30**,** 923-30.

LIEBER, M. R. 2008. The mechanism of human nonhomologous DNA end joining. *J Biol Chem,* 283**,** 1-5.

LIN, X., HUO, Z., LIU, X., ZHANG, Y., LI, L., ZHAO, H., YAN, B., LIU, Y., YANG, Y. & CHEN, Y. H. 2010. A novel GATA6 mutation in patients with tetralogy of Fallot or atrial septal defect. *J Hum Genet,* 55**,** 662-7.

LINDNER, T. H., NJOLSTAD, P. R., HORIKAWA, Y., BOSTAD, L., BELL, G. I. & SOVIK, O. 1999. A novel syndrome of diabetes mellitus, renal dysfunction and genital malformation associated with a partial deletion of the pseudo-POU domain of hepatocyte nuclear factor-1beta. *Hum Mol Genet,* 8**,** 2001-8.

LOH, K. M., ANG, L. T., ZHANG, J., KUMAR, V., ANG, J., AUYEONG, J. Q., LEE, K. L., CHOO, S. H., LIM, C. Y., NICHANE, M., TAN, J., NOGHABI, M. S., AZZOLA, L., NG, E. S., DURRUTHY-DURRUTHY, J., SEBASTIANO, V., POELLINGER, L., ELEFANTY, A. G., STANLEY, E. G., CHEN, Q., PRABHAKAR, S., WEISSMAN, I. L. & LIM, B. 2014. Efficient endoderm induction from human pluripotent stem cells by logically directing signals controlling lineage bifurcations. *Cell Stem Cell,* 14**,** 237-52.

LOMBARDO, A., CESANA, D., GENOVESE, P., DI STEFANO, B., PROVASI, E., COLOMBO, D. F., NERI, M., MAGNANI, Z., CANTORE, A., LO RISO, P., DAMO, M., PELLO, O. M., HOLMES, M. C., GREGORY, P. D., GRITTI, A., BROCCOLI, V., BONINI, C. & NALDINI, L. 2011. Site-specific integration and tailoring of cassette design for sustainable gene transfer. *Nat Methods,* 8**,** 861-9.

LOVE, M. I., HUBER, W. & ANDERS, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol,* 15**,** 550.

LOWE, L. A., YAMADA, S. & KUEHN, M. R. 2001. Genetic dissection of nodal function in patterning the mouse embryo. *Development,* 128**,** 1831-43.

LYNN, F. C., SMITH, S. B., WILSON, M. E., YANG, K. Y., NEKREP, N. & GERMAN, M. S. 2007. Sox9 coordinates a transcriptional network in pancreatic progenitor cells. *Proc Natl Acad Sci U S A,* 104**,** 10500-5.

LYTTLE, B. M., LI, J., KRISHNAMURTHY, M., FELLOWS, F., WHEELER, M. B., GOODYER, C. G. & WANG, R. 2008. Transcription factor expression in the developing human fetal endocrine pancreas. *Diabetologia,* 51**,** 1169-80.

MAEHR, R., CHEN, S., SNITOW, M., LUDWIG, T., YAGASAKI, L., GOLAND, R., LEIBEL, R. L. & MELTON, D. A. 2009. Generation of pluripotent stem cells from patients with type 1 diabetes. *Proc Natl Acad Sci U S A,* 106**,** 15768-73.

MAESTRO, M. A., BOJ, S. F., LUCO, R. F., PIERREUX, C. E., CABEDO, J., SERVITJA, J. M., GERMAN, M. S., ROUSSEAU, G. G., LEMAIGRE, F. P. & FERRER, J. 2003. Hnf6 and Tcf2 (MODY5) are linked in a gene network operating in a precursor cell domain of the embryonic pancreas. *Hum Mol Genet,* 12**,** 3307-14.

MALI, P., YANG, L., ESVELT, K. M., AACH, J., GUELL, M., DICARLO, J. E., NORVILLE, J. E. & CHURCH, G. M. 2013. RNA-guided human genome engineering via Cas9. *Science,* 339**,** 823-6.

MANDEGAR, M. A., HUEBSCH, N., FROLOV, E. B., SHIN, E., TRUONG, A., OLVERA, M. P., CHAN, A. H., MIYAOKA, Y., HOLMES, K., SPENCER, C. I., JUDGE, L. M., GORDON, D. E., ESKILDSEN, T. V., VILLALTA, J. E., HORLBECK, M. A., GILBERT, L. A., KROGAN, N. J., SHEIKH, S. P., WEISSMAN, J. S., QI, L. S., SO, P. L. & CONKLIN, B. R. 2016. CRISPR Interference Efficiently Induces Specific and Reversible Gene Silencing in Human iPSCs. *Cell Stem Cell,* 18**,** 541-53.

MARTIN, G. R. 1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A,* 78**,** 7634-8.

MCDONALD, E., LI, J., KRISHNAMURTHY, M., FELLOWS, G. F., GOODYER, C. G. & WANG, R. 2012. SOX9 regulates endocrine cell differentiation during human fetal pancreas development. *Int J Biochem Cell Biol,* 44**,** 72-83.

MCGRATH, P. S., WATSON, C. L., INGRAM, C., HELMRATH, M. A. & WELLS, J. M. 2015. The Basic Helix-Loop-Helix Transcription Factor NEUROG3 Is Required for Development of the Human Endocrine Pancreas. *Diabetes,* 64**,** 2497-505.

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. & DALTON, S. 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.

MELLOUL, D., MARSHAK, S. & CERASI, E. 2002. Regulation of insulin gene transcription. *Diabetologia,* 45**,** 309-26.

MILLER, J. C., HOLMES, M. C., WANG, J., GUSCHIN, D. Y., LEE, Y.-L., RUPNIEWSKI, I., BEAUSEJOUR, C. M., WAITE, A. J., WANG, N. S., KIM, K. A., GREGORY, P. D., PABO, C. O. & REBAR, E. J. 2007. An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotech,* 25**,** 778-785.

MILLER, J. C., TAN, S., QIAO, G., BARLOW, K. A., WANG, J., XIA, D. F., MENG, X., PASCHON, D. E., LEUNG, E., HINKLEY, S. J., DULAY, G. P., HUA, K. L., ANKOUDINOVA, I., COST, G. J., URNOV, F. D., ZHANG, H. S., HOLMES, M. C., ZHANG, L., GREGORY, P. D. & REBAR, E. J. 2011. A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol,* 29**,** 143-8.

MOEHLE, E. A., ROCK, J. M., LEE, Y. L., JOUVENOT, Y., DEKELVER, R. C., GREGORY, P. D., URNOV, F. D. & HOLMES, M. C. 2007. Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proc Natl Acad Sci U S A,* 104**,** 3055-60.

MOLKENTIN, J. D. 2000. The zinc finger-containing transcription factors GATA-4, -5, and -6. Ubiquitously expressed regulators of tissue-specific gene expression. *J Biol Chem,* 275**,** 38949-52.

MOLKENTIN, J. D., LIN, Q., DUNCAN, S. A. & OLSON, E. N. 1997. Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis. *Genes & Development,* 11**,** 1061-1072.

MORRISEY, E. E., IP, H. S., LU, M. M. & PARMACEK, M. S. 1996a. GATA-6: A Zinc Finger Transcription Factor That Is Expressed in Multiple Cell Lineages Derived from Lateral Mesoderm. *Developmental Biology,* 177**,** 309-322.

MORRISEY, E. E., IP, H. S., LU, M. M. & PARMACEK, M. S. 1996b. GATA-6: a zinc finger transcription factor that is expressed in multiple cell lineages derived from lateral mesoderm. *Dev Biol,* 177**,** 309-22.

MORRISEY, E. E., IP, H. S., TANG, Z., LU, M. M. & PARMACEK, M. S. 1997. GATA-5: A Transcriptional Activator Expressed in a Novel Temporally and Spatially-Restricted Pattern during Embryonic Development. *Developmental Biology,* 183**,** 21-36.

MORRISEY, E. E., TANG, Z., SIGRIST, K., LU, M. M., JIANG, F., IP, H. S. & PARMACEK, M. S. 1998. GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. *Genes Dev,* 12**,** 3579-90.

MOSCOU, M. J. & BOGDANOVE, A. J. 2009. A simple cipher governs DNA recognition by TAL effectors. *Science,* 326**,** 1501.

MUSSOLINO, C. & CATHOMEN, T. 2012. TALE nucleases: tailored genome engineering made easy. *Curr Opin Biotechnol,* 23**,** 644-50.

NARITA, N., HEIKINHEIMO, M., BIELINSKA, M., WHITE, R. A. & WILSON, D. B. 1996. The gene for transcription factor GATA-6 resides on mouse chromosome 18 and is expressed in myocardium and vascular smooth muscle. *Genomics,* 36**,** 345-8.

NISHIGORI, H., YAMADA, S., KOHAMA, T., TOMURA, H., SHO, K., HORIKAWA, Y., BELL, G. I., TAKEUCHI, T. & TAKEDA, J. 1998. Frameshift mutation, A263fsinsGG, in the hepatocyte nuclear factor-1beta gene associated with diabetes and renal dysfunction. *Diabetes,* 47**,** 1354-5.

NOSTRO, M. C., SARANGI, F., OGAWA, S., HOLTZINGER, A., CORNEO, B., LI, X., MICALLEF, S. J., PARK, I. H., BASFORD, C., WHEELER, M. B., DALEY, G. Q., ELEFANTY, A. G., STANLEY, E. G. & KELLER, G. 2011. Stage-specific signaling through TGFbeta family members and WNT regulates patterning and pancreatic specification of human pluripotent stem cells. *Development,* 138**,** 861-71.

O'RAHILLY, R. & MÜLLER, F. 2010. Developmental Stages in Human Embryos: Revised and New Measurements. *Cells Tissues Organs,* 192**,** 73-84.

OBATA, J., YANO, M., MIMURA, H., GOTO, T., NAKAYAMA, R., MIBU, Y., OKA, C. & KAWAICHI, M. 2001. p48 subunit of mouse PTF1 binds to RBP-Jkappa/CBF-1, the intracellular mediator of Notch signalling, and is expressed in the neural tube of early stage embryos. *Genes Cells,* 6**,** 345-60.

OFFIELD, M. F., JETTON, T. L., LABOSKY, P. A., RAY, M., STEIN, R. W., MAGNUSON, M. A., HOGAN, B. L. & WRIGHT, C. V. 1996. PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. *Development,* 122**,** 983-95.

OLIVER-KRASINSKI, J. M., KASNER, M. T., YANG, J., CRUTCHLOW, M. F., RUSTGI, A. K., KAESTNER, K. H. & STOFFERS, D. A. 2009. The diabetes gene Pdx1 regulates the transcriptional network of pancreatic endocrine progenitor cells in mice. *J Clin Invest,* 119**,** 1888-98.

ORKIN, S. H. 1998. Embryonic stem cells and transgenic mice in the study of hematopoiesis. *Int J Dev Biol,* 42**,** 927-34.

OSAFUNE, K., CARON, L., BOROWIAK, M., MARTINEZ, R. J., FITZ-GERALD, C. S., SATO, Y., COWAN, C. A., CHIEN, K. R. & MELTON, D. A. 2008. Marked differences in differentiation propensity among human embryonic stem cell lines. *Nat Biotechnol,* 26**,** 313-5.

PAGLIUCA, F. W., MILLMAN, J. R., GURTLER, M., SEGEL, M., VAN DERVORT, A., RYU, J. H., PETERSON, Q. P., GREINER, D. & MELTON, D. A. 2014. Generation of functional human pancreatic beta cells in vitro. *Cell,* 159**,** 428-39.

PAN, F. C. & WRIGHT, C. 2011. Pancreas organogenesis: from bud to plexus to gland. *Dev Dyn,* 240**,** 530-65.

PARK, C. Y., KIM, J., KWEON, J., SON, J. S., LEE, J. S., YOO, J. E., CHO, S. R., KIM, J. H., KIM, J. S. & KIM, D. W. 2014. Targeted inversion and reversion of the blood coagulation factor 8 gene in human iPS cells using TALENs. *Proc Natl Acad Sci U S A,* 111**,** 9253-8.

PATIENT, R. K. & MCGHEE, J. D. 2002. The GATA family (vertebrates and invertebrates). *Curr Opin Genet Dev,* 12**,** 416-22.

PEARSON, E. R., FLECHTNER, I., NJOLSTAD, P. R., MALECKI, M. T., FLANAGAN, S. E., LARKIN, B., ASHCROFT, F. M., KLIMES, I., CODNER, E., IOTOVA, V., SLINGERLAND, A. S., SHIELD, J., ROBERT, J. J., HOLST, J. J., CLARK, P. M., ELLARD, S., SOVIK, O., POLAK, M. & HATTERSLEY, A. T. 2006. Switching from insulin to oral sulfonylureas in patients with diabetes due to Kir6.2 mutations. *N Engl J Med,* 355**,** 467-77.

PIPER, K., BALL*, S. G., KEELING, J. W., MANSOOR, S., WILSON, D. I. & HANLEY, N. A. 2002. Novel SOX9 expression during human pancreas development correlates to abnormalities in Campomelic dysplasia. *Mechanisms of Development,* 116**,** 223-226.

PIPER, K., BRICKWOOD, S., TURNPENNY, L. W., CAMERON, I. T., BALL, S. G., WILSON, D. I. & HANLEY, N. A. 2004. Beta cell differentiation during early human pancreas development. *J Endocrinol,* 181**,** 11-23.

PLESSIS, A., PERRIN, A., HABER, J. E. & DUJON, B. 1992. Site-specific recombination determined by I-SceI, a mitochondrial group I intron-encoded endonuclease expressed in the yeast nucleus. *Genetics,* 130**,** 451-60.

PROVASI, E., GENOVESE, P., LOMBARDO, A., MAGNANI, Z., LIU, P. Q., REIK, A., CHU, V., PASCHON, D. E., ZHANG, L., KUBALL, J., CAMISA, B., BONDANZA, A., CASORATI, G., PONZONI, M., CICERI, F., BORDIGNON, C., GREENBERG, P. D., HOLMES, M. C., GREGORY, P. D., NALDINI, L. & BONINI, C. 2012. Editing T cell specificity towards leukemia by zinc finger nucleases and lentiviral gene transfer. *Nat Med,* 18**,** 807-15.

RAFIQ, M., FLANAGAN, S. E., PATCH, A. M., SHIELDS, B. M., ELLARD, S. & HATTERSLEY, A. T. 2008. Effective treatment with oral sulfonylureas in patients with diabetes due to sulfonylurea receptor 1 (SUR1) mutations. *Diabetes Care,* 31**,** 204-9.

RAMIREZ, F., DUNDAR, F., DIEHL, S., GRUNING, B. A. & MANKE, T. 2014. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res,* 42**,** W187-91.

RASHID, S. T., CORBINEAU, S., HANNAN, N., MARCINIAK, S. J., MIRANDA, E., ALEXANDER, G., HUANG-DORAN, I., GRIFFIN, J., AHRLUND-RICHTER, L., SKEPPER, J., SEMPLE, R., WEBER, A., LOMAS, D. A. & VALLIER, L. 2010. Modeling inherited metabolic disorders of the liver using human induced pluripotent stem cells. *The Journal of Clinical Investigation,* 120**,** 3127-3136.

REYON, D., TSAI, S. Q., KHAYTER, C., FODEN, J. A., SANDER, J. D. & JOUNG, J. K. 2012. FLASH Assembly of TALENs Enables High-Throughput Genome Editing. *Nature biotechnology,* 30**,** 460-465.

REZANIA, A., BRUIN, J. E., ARORA, P., RUBIN, A., BATUSHANSKY, I., ASADI, A., O'DWYER, S., QUISKAMP, N., MOJIBIAN, M., ALBRECHT, T., YANG, Y. H. C., JOHNSON, J. D. & KIEFFER, T. J. 2014. Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. *Nat Biotech,* 32**,** 1121-1133.

ROBINSON, M. D., MCCARTHY, D. J. & SMYTH, G. K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics,* 26**,** 139-40.

ROBINTON, D. A. & DALEY, G. Q. 2012. The promise of induced pluripotent stem cells in research and therapy. *Nature,* 481**,** 295-305.

ROJAS, A., SCHACHTERLE, W., XU, S. M. & BLACK, B. L. 2009. An endoderm-specific transcriptional enhancer from the mouse Gata4 gene requires GATA and homeodomain protein-binding sites for function in vivo. *Dev Dyn,* 238**,** 2588-98.

ROSTOVSKAYA, M., BREDENKAMP, N. & SMITH, A. 2015. Towards consistent generation of pancreatic lineage progenitors from human pluripotent stem cells. *Philos Trans R Soc Lond B Biol Sci,* 370**,** 20140365.

ROUET, P., SMIH, F. & JASIN, M. 1994. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol Cell Biol,* 14**,** 8096-106.

RUBIO-CABEZAS, O., JENSEN, J. N., HODGSON, M. I., CODNER, E., ELLARD, S., SERUP, P. & HATTERSLEY, A. T. 2011. Permanent Neonatal Diabetes and Enteric Anendocrinosis Associated With Biallelic Mutations in NEUROG3. *Diabetes,* 60**,** 1349-53.

RUDIN, N., SUGARMAN, E. & HABER, J. E. 1989. Genetic and physical analysis of double-strand break repair and recombination in Saccharomyces cerevisiae. *Genetics,* 122**,** 519-34.

RUSS, H. A., PARENT, A. V., RINGLER, J. J., HENNINGS, T. G., NAIR, G. G., SHVEYGERT, M., GUO, T., PURI, S., HAATAJA, L., CIRULLI, V., BLELLOCH, R., SZOT, G. L., ARVAN, P. & HEBROK, M. 2015. Controlled induction of human pancreatic progenitors produces functional beta-like cells in vitro. *EMBO J,* 34**,** 1759-72.

SALISBURY, R. J., BLAYLOCK, J., BERRY, A. A., JENNINGS, R. E., DE KRIJGER, R., PIPER HANLEY, K. & HANLEY, N. A. 2014. The window period of NEUROGENIN3 during human gestation. *Islets,* 6**,** e954436.

SANDER, J. D., CADE, L., KHAYTER, C., REYON, D., PETERSON, R. T., JOUNG, J. K. & YEH, J. R. 2011. Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nat Biotechnol,* 29**,** 697-8.

SANDER, J. D., MAEDER, M. L., REYON, D., VOYTAS, D. F., JOUNG, J. K. & DOBBS, D. 2010. ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. *Nucleic Acids Research*.

SANDER, J. D., ZABACK, P., JOUNG, J. K., VOYTAS, D. F. & DOBBS, D. 2007. Zinc Finger Targeter (ZiFiT): an engineered zinc finger/target site design tool. *Nucleic Acids Research,* 35**,** W599-W605.

SANDER, M., SUSSEL, L., CONNERS, J., SCHEEL, D., KALAMARAS, J., DELA CRUZ, F., SCHWITZGEBEL, V., HAYES-JORDAN, A. & GERMAN, M. 2000. Homeobox gene Nkx6.1 lies downstream of Nkx2.2 in the major pathway of beta-cell formation in the pancreas. *Development,* 127**,** 5533-40.

SCHAFFER, A. E., FREUDE, K. K., NELSON, S. B. & SANDER, M. 2010. Ptf1a and Nkx6 transcription factors function as antagonistic lineage determinants in multipotent pancreatic progenitors. *Developmental cell,* 18**,** 1022-1029.

SCHAFFER, A. E., TAYLOR, B. L., BENTHUYSEN, J. R., LIU, J., THOREL, F., YUAN, W., JIAO, Y., KAESTNER, K. H., HERRERA, P. L., MAGNUSON, M. A., MAY, C. L. & SANDER, M. 2013. Nkx6.1 controls a gene regulatory network required for establishing and maintaining pancreatic Beta cell identity. *PLoS Genet,* 9**,** e1003274.

SCHWITZGEBEL, V. M., SCHEEL, D. W., CONNERS, J. R., KALAMARAS, J., LEE, J. E., ANDERSON, D. J., SUSSEL, L., JOHNSON, J. D. & GERMAN, M. S. 2000. Expression of neurogenin3 reveals an islet cell precursor population in the pancreas. *Development,* 127**,** 3533-42.

SEGUIN, C. A., DRAPER, J. S., NAGY, A. & ROSSANT, J. 2008. Establishment of endoderm progenitors by SOX transcription factor expression in human embryonic stem cells. *Cell Stem Cell,* 3**,** 182-95.

SEIDMAN, J. G. & SEIDMAN, C. 2002. Transcription factor haploinsufficiency: when half a loaf is not enough. *J Clin Invest,* 109**,** 451-5.

SELLICK, G. S., BARKER, K. T., STOLTE-DIJKSTRA, I., FLEISCHMANN, C., COLEMAN, R. J., GARRETT, C., GLOYN, A. L., EDGHILL, E. L., HATTERSLEY, A. T., WELLAUER, P. K., GOODWIN, G. & HOULSTON, R. S. 2004. Mutations in PTF1A cause pancreatic and cerebellar agenesis. *Nat Genet,* 36**,** 1301-5.

SEYMOUR, P. A., FREUDE, K. K., DUBOIS, C. L., SHIH, H. P., PATEL, N. A. & SANDER, M. 2008. A dosage-dependent requirement for Sox9 in pancreatic endocrine cell formation. *Dev Biol,* 323**,** 19-30.

SEYMOUR, P. A., FREUDE, K. K., TRAN, M. N., MAYES, E. E., JENSEN, J., KIST, R., SCHERER, G. & SANDER, M. 2007. SOX9 is required for maintenance of the pancreatic progenitor cell pool. *Proc Natl Acad Sci U S A,* 104**,** 1865-70.

SHAW-SMITH, C., DE FRANCO, E., LANGO ALLEN, H., BATLLE, M., FLANAGAN, S. E., BOROWIEC, M., TAPLIN, C. E., VAN ALFEN-VAN DER VELDEN, J., CRUZ-ROJO, J., PEREZ DE NANCLARES, G., MIEDZYBRODZKA, Z., DEJA, G., WLODARSKA, I., MLYNARSKI, W., FERRER, J., HATTERSLEY, A. T. & ELLARD, S. 2014. GATA4 mutations are a cause of neonatal and childhood-onset diabetes. *Diabetes,* 63**,** 2888-94.

SHI, Z. D., LEE, K., YANG, D., AMIN, S., VERMA, N., LI, Q. V., ZHU, Z., SOH, C. L., KUMAR, R., EVANS, T., CHEN, S. & HUANGFU, D. 2017. Genome Editing in hPSCs Reveals GATA6 Haploinsufficiency and a Genetic Interaction with GATA4 in Human Pancreatic Development. *Cell Stem Cell*.

SHIH, H. P., WANG, A. & SANDER, M. 2013. Pancreas organogenesis: from lineage determination to morphogenesis. *Annu Rev Cell Dev Biol,* 29**,** 81-105.

SINNER, D., RANKIN, S., LEE, M. & ZORN, A. M. 2004. Sox17 and beta-catenin cooperate to regulate the transcription of endodermal genes. *Development,* 131**,** 3069-80.

SMITH, A. G., HEATH, J. K., DONALDSON, D. D., WONG, G. G., MOREAU, J., STAHL, M. & ROGERS, D. 1988. Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature,* 336**,** 688-90.

SOLAR, M., CARDALDA, C., HOUBRACKEN, I., MARTIN, M., MAESTRO, M. A., DE MEDTS, N., XU, X., GRAU, V., HEIMBERG, H., BOUWENS, L. & FERRER, J. 2009. Pancreatic exocrine duct cells give rise to insulin-producing beta cells during embryogenesis but not after birth. *Dev Cell,* 17**,** 849-60.

SOLTER, D. 2006. From teratocarcinomas to embryonic stem cells and beyond: a history of embryonic stem cell research. *Nat Rev Genet,* 7**,** 319-327.

SORIA, B., ROCHE, E., BERNA, G., LEON-QUINTO, T., REIG, J. A. & MARTIN, F. 2000. Insulin-secreting cells derived from embryonic stem cells normalize glycemia in streptozotocin-induced diabetic mice. *Diabetes,* 49**,** 157-62.

SOSA-PINEDA, B., CHOWDHURY, K., TORRES, M., OLIVER, G. & GRUSS, P. 1997. The Pax4 gene is essential for differentiation of insulin-producing beta cells in the mammalian pancreas. *Nature,* 386**,** 399-402.

STANIK, J., GASPERIKOVA, D., PASKOVA, M., BARAK, L., JAVORKOVA, J., JANCOVA, E., CILJAKOVA, M., HLAVA, P., MICHALEK, J., FLANAGAN, S. E., PEARSON, E., HATTERSLEY, A. T., ELLARD, S. & KLIMES, I. 2007. Prevalence of permanent neonatal diabetes in Slovakia and successful replacement of insulin with sulfonylurea therapy in KCNJ11 and ABCC8 mutation carriers. *J Clin Endocrinol Metab,* 92**,** 1276-82.

STOFFERS, D. A., HELLER, R. S., MILLER, C. P. & HABENER, J. F. 1999. Developmental expression of the homeodomain protein IDX-1 in mice transgenic for an IDX-1 promoter/lacZ transcriptional reporter. *Endocrinology,* 140**,** 5374-81.

STOFFERS, D. A., ZINKIN, N. T., STANOJEVIC, V., CLARKE, W. L. & HABENER, J. F. 1997. Pancreatic agenesis attributable to a single nucleotide deletion in the human IPF1 gene coding sequence. *Nat Genet,* 15**,** 106-10.

STREUBEL, J., BLUCHER, C., LANDGRAF, A. & BOCH, J. 2012. TAL effector RVD specificities and efficiencies. *Nat Biotech,* 30**,** 593-595.

SUDA, Y., SUZUKI, M., IKAWA, Y. & AIZAWA, S. 1987. Mouse embryonic stem cells exhibit indefinite proliferative potential. *J Cell Physiol,* 133**,** 197-201.

SUMI, K., TANAKA, T., UCHIDA, A., MAGOORI, K., URASHIMA, Y., OHASHI, R., OHGUCHI, H., OKAMURA, M., KUDO, H., DAIGO, K., MAEJIMA, T., KOJIMA, N., SAKAKIBARA, I., JIANG, S., HASEGAWA, G., KIM, I., OSBORNE, T. F., NAITO, M., GONZALEZ, F. J., HAMAKUBO, T., KODAMA, T. & SAKAI, J. 2007. Cooperative Interaction between Hepatocyte Nuclear Factor 4α and GATA Transcription Factors Regulates ATP-Binding Cassette Sterol Transporters ABCG5 and ABCG8. *Molecular and Cellular Biology,* 27**,** 4248-4260.

SUZUKI, E., EVANS, T., LOWRY, J., TRUONG, L., BELL, D. W., TESTA, J. R. & WALSH, K. 1996. The HumanGATA-6Gene: Structure, Chromosomal Location, and Regulation of Expression by Tissue-Specific and Mitogen-Responsive Signals. *Genomics,* 38**,** 283-290.

SVENSSON, P., WILLIAMS, C., LUNDEBERG, J., RYDEN, P., BERGQVIST, I. & EDLUND, H. 2007. Gene array identification of Ipf1/Pdx1-/- regulated genes in pancreatic progenitor cells. *BMC Dev Biol,* 7**,** 129.

SZCZEPEK, M., BRONDANI, V., BUCHEL, J., SERRANO, L., SEGAL, D. J. & CATHOMEN, T. 2007. Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat Biotech,* 25**,** 786-793.

TAKAHASHI, K., TANABE, K., OHNUKI, M., NARITA, M., ICHISAKA, T., TOMODA, K. & YAMANAKA, S. 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell,* 131**,** 861-72.

TAKAHASHI, K. & YAMANAKA, S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell,* 126**,** 663-76.

TAKEDA, M., OBAYASHI, K., KOBAYASHI, A. & MAEDA, M. 2004. A unique role of an amino terminal 16-residue region of long-type GATA-6. *J Biochem,* 135**,** 639-50.

TATTERSALL, R. B. 1974. Mild familial diabetes with dominant inheritance. *Q J Med,* 43**,** 339-57.

TATTERSALL, R. B. & FAJANS, S. S. 1975. A difference between the inheritance of classical juvenile-onset and maturity-onset type diabetes of young people. *Diabetes,* 24**,** 44-53.

TEO, A. K., ARNOLD, S. J., TROTTER, M. W., BROWN, S., ANG, L. T., CHNG, Z., ROBERTSON, E. J., DUNN, N. R. & VALLIER, L. 2011. Pluripotency factors regulate definitive endoderm specification through eomesodermin. *Genes Dev,* 25**,** 238-50.

THOMSON, J. A., ITSKOVITZ-ELDOR, J., SHAPIRO, S. S., WAKNITZ, M. A., SWIERGIEL, J. J., MARSHALL, V. S. & JONES, J. M. 1998. Embryonic stem cell lines derived from human blastocysts. *Science,* 282**,** 1145-7.

THOMSON, J. A., KALISHMAN, J., GOLOS, T. G., DURNING, M., HARRIS, C. P., BECKER, R. A. & HEARN, J. P. 1995. Isolation of a primate embryonic stem cell line. *Proc Natl Acad Sci U S A,* 92**,** 7844-8.

THORVALDSDÓTTIR, H., ROBINSON, J. T. & MESIROV, J. P. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics,* 14**,** 178-192.

TIYABOONCHAI, A., CARDENAS-DIAZ, F. L., YING, L., MAGUIRE, J. A., SIM, X., JOBALIYA, C., GAGNE, A. L., KISHORE, S., STANESCU, D. E., HUGHES, N., DE LEON, D. D., FRENCH, D. L. & GADUE, P. 2017. GATA6 Plays an Important Role in the Induction of Human

Definitive Endoderm, Development of the Pancreas, and Functionality of Pancreatic beta Cells. *Stem Cell Reports,* 8**,** 589-604.

TORIKAI, H., REIK, A., LIU, P. Q., ZHOU, Y., ZHANG, L., MAITI, S., HULS, H., MILLER, J. C., KEBRIAEI, P., RABINOVICH, B., LEE, D. A., CHAMPLIN, R. E., BONINI, C., NALDINI, L., REBAR, E. J., GREGORY, P. D., HOLMES, M. C. & COOPER, L. J. 2012. A foundation for universal T-cell based immunotherapy: T cells engineered to express a CD19-specific chimeric-antigen-receptor and eliminate expression of endogenous TCR. *Blood,* 119**,** 5697-705.

TOWNSEND, J. A., WRIGHT, D. A., WINFREY, R. J., FU, F., MAEDER, M. L., JOUNG, J. K. & VOYTAS, D. F. 2009. High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature,* 459**,** 442-5.

TSAI, S. F., MARTIN, D. I., ZON, L. I., D'ANDREA, A. D., WONG, G. G. & ORKIN, S. H. 1989. Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. *Nature,* 339**,** 446-51.

VALLIER, L., MENDJAN, S., BROWN, S., CHNG, Z., TEO, A., SMITHERS, L. E., TROTTER, M. W., CHO, C. H., MARTINEZ, A., RUGG-GUNN, P., BRONS, G. & PEDERSEN, R. A. 2009a. Activin/Nodal signalling maintains pluripotency by controlling Nanog expression. *Development,* 136**,** 1339-49.

VALLIER, L., TOUBOUL, T., CHNG, Z., BRIMPARI, M., HANNAN, N., MILLAN, E., SMITHERS, L. E., TROTTER, M., RUGG-GUNN, P., WEBER, A. & PEDERSEN, R. A. 2009b. Early cell fate decisions of human embryonic stem cells and mouse epiblast stem cells are controlled by the same signalling pathways. *PLoS One,* 4**,** e6082.

VAN BELLE, T. L., COPPIETERS, K. T. & VON HERRATH, M. G. 2011. Type 1 Diabetes: Etiology, Immunology, and Therapeutic Strategies. *Physiological Reviews,* 91**,** 79.

VANAMEE, E. S., SANTAGATA, S. & AGGARWAL, A. K. 2001. FokI requires two specific DNA sites for cleavage. *J Mol Biol,* 309**,** 69-78.

VILLASENOR, A., CHONG, D. C. & CLEAVER, O. 2008. Biphasic Ngn3 expression in the developing pancreas. *Developmental dynamics : an official publication of the American Association of Anatomists,* 237**,** 3270-3279.

WANG, S., SUN, H., MA, J., ZANG, C., WANG, C., WANG, J., TANG, Q., MEYER, C. A., ZHANG, Y. & LIU, X. S. 2013. Target analysis by integration of transcriptome and ChIP-seq data with BETA. *Nat Protoc,* 8**,** 2502-15.

WANG, Z., LI, J., HUANG, H., WANG, G., JIANG, M., YIN, S., SUN, C., ZHANG, H., ZHUANG, F. & XI, J. J. 2012. An integrated chip for the high-throughput synthesis of transcription activator-like effectors. *Angew Chem Int Ed Engl,* 51**,** 8505-8.

WATT, A. J., BATTLE, M. A., LI, J. & DUNCAN, S. A. 2004. GATA4 is essential for formation of the proepicardium and regulates cardiogenesis. *Proceedings of the National Academy of Sciences of the United States of America,* 101**,** 12573-12578.

WATT, A. J., ZHAO, R., LI, J. & DUNCAN, S. A. 2007. Development of the mammalian liver and ventral pancreas is dependent on GATA4. *BMC Dev Biol,* 7**,** 37.

WEBER, E., GRUETZNER, R., WERNER, S., ENGLER, C. & MARILLONNET, S. 2011. Assembly of designer TAL effectors by Golden Gate cloning. *PLoS One,* 6**,** e19722.

WEEDON, M. N., CEBOLA, I., PATCH, A. M., FLANAGAN, S. E., DE FRANCO, E., CASWELL, R., RODRIGUEZ-SEGUI, S. A., SHAW-SMITH, C., CHO, C. H., LANGO ALLEN, H., HOUGHTON, J. A., ROTH, C. L., CHEN, R., HUSSAIN, K., MARSH, P., VALLIER, L., MURRAY, A., ELLARD, S., FERRER, J. & HATTERSLEY, A. T. 2014. Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. *Nat Genet,* 46**,** 61-4.

WERNIG, M., MEISSNER, A., FOREMAN, R., BRAMBRINK, T., KU, M., HOCHEDLINGER, K., BERNSTEIN, B. E. & JAENISCH, R. 2007. In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature,* 448**,** 318-324.

WILLIAMS, R. L., HILTON, D. J., PEASE, S., WILLSON, T. A., STEWART, C. L., GEARING, D. P., WAGNER, E. F., METCALF, D., NICOLA, N. A. & GOUGH, N. M. 1988. Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature,* 336**,** 684-7.

XUAN, S., BOROK, M. J., DECKER, K. J., BATTLE, M. A., DUNCAN, S. A., HALE, M. A., MACDONALD, R. J. & SUSSEL, L. 2012. Pancreas-specific deletion of mouse Gata4 and Gata6 causes pancreatic agenesis. *J Clin Invest,* 122**,** 3516-28.

YAMAMOTO, M., KO, L. J., LEONARD, M. W., BEUG, H., ORKIN, S. H. & ENGEL, J. D. 1990. Activity and tissue-specific expression of the transcription factor NF-E1 multigene family. *Genes Dev,* 4**,** 1650-62.

YAU, D., DE FRANCO, E., FLANAGAN, S. E., ELLARD, S., BLUMENKRANTZ, M. & MITCHELL, J. J. 2017. Case report: maternal mosaicism resulting in inheritance of a novel GATA6 mutation causing pancreatic agenesis and neonatal diabetes mellitus. *Diagnostic Pathology,* 12**,** 1.

YU, H. & COWAN, C. A. 2016. Minireview: Genome Editing of Human Pluripotent Stem Cells for Modeling Metabolic Disease. *Molecular Endocrinology,* 30**,** 575-586.

YUSA, K., RASHID, S. T., STRICK-MARCHAND, H., VARELA, I., LIU, P.-Q., PASCHON, D. E., MIRANDA, E., ORDONEZ, A., HANNAN, N. R. F., ROUHANI, F. J., DARCHE, S., ALEXANDER, G., MARCINIAK, S. J., FUSAKI, N., HASEGAWA, M., HOLMES, M. C., DI SANTO, J. P., LOMAS, D. A., BRADLEY, A. & VALLIER, L. 2011. Targeted gene correction of [agr]1-antitrypsin deficiency in induced pluripotent stem cells. *Nature,* 478**,** 391-394.

ZHANG, D., JIANG, W., LIU, M., SUI, X., YIN, X., CHEN, S., SHI, Y. & DENG, H. 2009. Highly efficient differentiation of human ES cells and iPS cells into mature pancreatic insulin-producing cells. *Cell Res,* 19**,** 429-38.

ZHANG, F., MAEDER, M. L., UNGER-WALLACE, E., HOSHAW, J. P., REYON, D., CHRISTIAN, M., LI, X., PIERICK, C. J., DOBBS, D., PETERSON, T., JOUNG, J. K. & VOYTAS, D. F. 2010. High frequency targeted mutagenesis in Arabidopsis thaliana using zinc finger nucleases. *Proc Natl Acad Sci U S A,* 107**,** 12028-33.

ZHANG, Y., LIU, T., MEYER, C. A., EECKHOUTE, J., JOHNSON, D. S., BERNSTEIN, B. E., NUSBAUM, C., MYERS, R. M., BROWN, M., LI, W. & LIU, X. S. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol,* 9**,** R137.

ZHAO, R., WATT, A. J., LI, J., LUEBKE-WHEELER, J., MORRISEY, E. E. & DUNCAN, S. A. 2005. GATA6 is essential for embryonic development of the liver but dispensable for early heart formation. *Mol Cell Biol,* 25**,** 2622-31.

ZWAKA, T. P. & THOMSON, J. A. 2005. A germ cell origin of embryonic stem cells? *Development,* 132**,** 227-33.