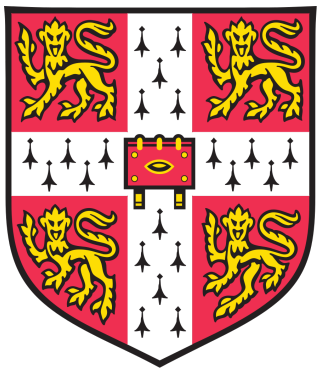
The use of human pluripotent stem cells to model HNF1B-associated diabetes



Ranna El Khairi

Clare College
University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy
September 2017

Abstract

Heterozygous mutations in the transcription factor, hepatocyte nuclear factor 1β (HNF1B), result in multisystem disease including diabetes due to beta-cell dysfunction and pancreatic hypoplasia. However, the mechanisms that underlie development of diabetes in HNF1B mutation carriers are still not fully understood due to lack of an appropriate model system. Human induced pluripotent stem cells (hiPSCs), which are capable of self-renewal and can differentiate into any cell type, provide an advantageous alternative to model human developmental diseases. The aim of this project was to develop a hiPSC based model system to determine the molecular mechanisms by which HNF1B mutations cause pancreatic hypoplasia and diabetes.

HNF1B mutant hiPSC lines were produced using CRISPR-Cas9 genome editing. Isogenic HNF1B wild-type, homozygous and heterozygous mutant hiPSC lines were directed to differentiate along the pancreatic lineage and cells were phenotyped at each stage of the differentiation process to check for appropriate expression of lineage markers. The normal expression pattern of HNF1B in human pancreas development was analysed and showed up-regulation of HNF1B at the foregut stage, and during pancreas specification. Homozygous knockout of HNF1B resulted in failure of foregut and pancreatic progenitor development, while heterozygous knockout of HNF1B resulted in impairment of pancreatic progenitor and endocrine cell differentiation as well as impaired insulin secretion upon glucose stimulation. Cell proliferation analyses showed a significant decrease in the proliferation rate in HNF1B heterozygous and homozygous mutant cells compared with wild-type cells at the foregut stage while no change in the apoptosis rate could be detected. RNA-sequencing and ATAC-sequencing, were used to further define the molecular mechanisms controlled by HNF1B and the effect HNF1B on modulation of chromatin accessibility during pancreas development.

These results provide further insights into the molecular mechanisms by which HNF1B regulates human pancreas development and function, revealing that HNF1B haploinsufficiency impairs the expansion and maintenance of pancreatic progenitor cells in vitro. In vivo, this would likely result in reduced beta cell numbers at birth and diabetes later in life in patients with HNF1B-associated disease. These mechanisms suggest that the capacity to produce pancreatic progenitor cells during embryonic life could determine individual susceptibility to diabetes.

Declaration

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

It does not exceed the prescribed word limit of 60,000 words (80,000 by special permission) excluding bibliography, figures, appendices etc. prescribed by the Degree Committee of the Department of Biology.

Ranna El Khairi

September 2017

Acknowledgements

I would like to thank my PhD supervisor, Dr Ludovic Vallier, for the opportunity to work on this project, and for his on-going support and guidance. I am grateful to my external supervisor, Professor Andrew Hattersley for his advice, access to his amazing resource of patients with monogenic diabetes and his enthusiasm towards the project. I would also like to thank Dr. Robert Semple for his encouragement and insightful comments as well as his assistance in recruiting patients for the study.

I would like to express my appreciation and gratitude to members of the Pancreatic Genetics team at the Sanger Institute and other members of the Vallier laboratory for teaching me experimental procedures, troubleshooting and for many stimulating discussions. Special thanks to Sapna, Kasia and Alexandra for their support and friendship and making the many hours I spent in the lab so enjoyable. I am particularly grateful to Dr Pedro Madrigal and my sister, Dr Muna El Khairi, for their assistance with the statistical and NGS data analysis.

I would also like to thank other collaborators who have contributed in various capacities to the project, particularly Dr Kouroush Saeb-Parsey, Dr Kathleen Elliot and my husband Dr Reza Motallebzadeh, for their assistance with the in vivo studies.

Finally, I would like to thank my parents, sister and husband for their endless support and encouragement.

This project was funded by the Wellcome Trust, and I am very grateful for their generous funding which made this project possible.

Table of Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

List of Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

List of Abbreviations

AFP	Alfa fetoprotein
ALB	Albumin
AP	Anterior-posterior
APOF	Apolipoprotein F
ARX	Aristaless Related Homeobox
ATAC	Assay for Transposase-Accessible Chromatin
ATP	Adenosine triphosphate
BDC	Bile duct cysts
BICC	Bicc Family RNA Binding Protein
BMP	Bone Morphogenetic Protein
BNZ	6-Benzoyladenosine-3',5'-cyclic monophosphate / 6-BNZ-cAMP
BSA	Bovine Serum Albumin
cAMP	Cyclic adenosine monophosphate
CAKUT	Congenital abnormalities of the kidneys and urinary tract
CBP	CREB-binding protein
CDM	Chemically defined medium
cDNA	complimentary DNA
CDX	Caudal type homeobox
ChIP	Chromatin Immunoprecipitation
CPEP	C-peptide
CRB3	Crumbs 3, Cell Polarity Complex Component
CRISPR	Clustered regularly interspaced short palindromic repeat
CS	Carnegie Stage
CXCR4	Chemokine (C-X-C motif) receptor 4
CYP	Cytochrome P450
CYS	Cystin
DAPI	4', 6-diamidino-2-phenylindole
DAPT	N-(N-((3,5-Difluorophenyl)acetyl)-L-alanyl)-L-phenylglycinetert-butyl ester
DE	Definitive Endoderm
DLL	Delta Like Canonical Notch Ligand
DM	Diabetes mellitus
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphates
DPC	Days post conception
DV	Dorsal-ventral
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
ELISA	Enzyme Linked Immunosorbent Assay
EMT	Epithelial-mesenchymal transition
EP	Endodermal progenitor
ERK	Extracellular signal-regulated kinase
ESC	Embryonic Stem Cell

FACS	Fluorescent Activated Cell Sorting
FAIRE	Formaldehyde-Assisted Isolation of Regulatory Elements
FASN	Fatty acid synthase
FBP1	Fructose-1,6-bisphosphatase 1
FBS	Fetal Bovine Serum
FFA	Free fatty acids
FGF	Fibroblast growth factor
FGFR	Fibroblast growth factor receptor
FOX	Forkhead box
FP	Foregut progenitor
FXYD	FXYD Domain Containing Ion Transport Regulator
G6PC	Glucose 6-phosphatase
GATA	GATA binding protein
GCG	Glucagon
GCK	Glucokinase
GCKD	Glomerulocystic kidney disease
GGT	Gamma-glutamyl transpeptidase
GLIS	GLIS Family Zinc Finger
GLUT	Glucose transporter
GMP	Good manufacturing practice
GSIS	Glucose-stimulated insulin secretion
HHEX	Haematopoietically expressed homeobox
HES	Hes Family BHLH Transcription Factor
HH	Hedgehog
HNF	Hepatic Nuclear Factor
HOX	Homeodomain
IHBD	Intrahepatic bile ducts
INS	Insulin
INSM	INSM Transcriptional Repressor
INSR	Insulin receptor
iPSC	Induced Pluripotent Stem Cell
ISL	ISL LIM Homeobox
KGF	Keratinocyte growth factor
KIF	Kinesin Family Member
KLF	Kruppel Like Factor
MAFA	v-maf avian musculoaponeurotic fibrosarcoma oncogene homolog A
MAPK	Mitogen-activated protein kinase
MEF	Mouse embryonic fibroblast
MLPA	Multiple ligation-dependent probe amplification
MNX	Motor neuron and pancreas homeobox
MODY	Maturity onset diabetes of the young
MPC	Multipotent progenitor cells
MRCP	Magnetic resonance cholangiopancreatography
MYC	MYC Proto-Oncogene, BHLH Transcription Factor
NANOG	Nanog homeobox
NDM	Neonatal diabetes mellitus
NEUROD	Neuronal differentiation
NEUROG	Neurogenin

NGS	Next Generation Sequencing
NKX	NK Homeobox
NMD	Nonsense mediated decay
OATP	Organic anion transporting polypeptide
OCT	Octamer-binding transcription factor
OSM	Oncostatin-M
PAGE	Polyacrylamide gel electrophoresis
PAX	Paired box
PBGD	Porphobilinogen deaminase
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDBU	Phorbol 12,13-dibutyrate
PDX	Pancreatic and Duodenal Homeobox
PE	Pancreatic Endoderm
PFA	Paraformaldehyde
PI3K	Phosphatidylinositol 3-kinase
PKA	Protein kinase A
PKD2	Polycystin 2, Transient Receptor Potential Cation Channel
PNDM	Permanent neonatal diabetes mellitus
PKHD1	PKHD1, Fibrocystin/Polyductin
POU1F1	POU Class 1 Homeobox 1
PP	Pancreatic polypeptide
PROX	Prospero Homeobox
PS	Primitive streak
PSC	Pluripotent Stem Cell
PTF	Pancreas Specific Transcription Factor
PTH	Parathyroid hormone
PVA	Poly vinyl alcohol
qRT-PCR	Real-time quantitative PCR
RA	All Trans-Retinoic Acid
RBPJ	Recombination Signal Binding Protein for Immunoglobulin Kappa J Region
RCAD	Renal cyst and diabetes syndrome
RCC	Renal cell carcinoma
RFX	Regulatory Factor
RNA	Ribonucleic acid
SDS	Sodium dodecyl sulfate
SEM	Standard error of the mean
SHH	Sonic hedgehog
SNP	Single nucleotide polymorphism
SOX	SRY-related HMG Box
SSH	Sonic hedgehog
SST	Somatostatin
T1D	Type 1 diabetes
T2D	Type 2 diabetes
TALEN	transcription activator-like effector nucleases
TBST	Tris-buffered saline with Tween-20
TGF	Transforming growth factor

TMEM	Transmembrane Protein
TNDM	Transient neonatal diabetes mellitus
TTR	Transthyretin
UMOD	Uromodulin
VE	Visceral endoderm
VLCAD	Very long-chain acyl-CoA dehydrogenase
WB	Western blot
WT	Wild Type
Wnt	Wingless-type MMTV integration site
ZFN	Zinc finger nuclease

Chapter 1 Introduction

1.1 Diabetes Mellitus

Diabetes mellitus is characterized by high blood glucose levels as a result of insufficient ability of the pancreas to secrete adequate amounts of insulin to maintain euglycaemia. The condition affects an estimated 336 million people and the incidence is increasing steadily worldwide (Danaei et al., 2011). It is estimated that by 2035 the number of people with diabetes will increase by 55% to 592 million (1 in 10 people) (Guariguata et al., 2014). In the United Kingdom (UK) there are approximately 3.9 million individuals with diabetes. Treatment of the complications of the disease consume £14 billion per year in the UK, which accounts for about 10% of the National Health Service (NHS) budget (Hex et al., 2012).

Diabetes is a heterogeneous disorder with multiple aetiologies. Type 1 diabetes (T1D) is an autoimmune disease, characterized by progressive destruction of pancreatic beta cells, resulting in absolute deficiency in insulin secretion. Type 2 diabetes (T2D) is the commonest form of diabetes (85-95%) and is the result of insulin resistance in peripheral organs (liver, muscle and adipose tissue) and beta cell dysfunction, dedifferentiation and death leading to relative insulin deficiency. The number of β-cells lost in T2D is unclear but can approach 60% (Butler et al., 2003, Rahier et al., 2008), and the remaining β-cells are likely to be dysfunctional.

Other, rarer forms of diabetes are associated with either monogenetic defects in β-cell development and function or abnormalities in insulin action i.e. insulin resistance. These monogenic forms of diabetes account for approximately 1-5% of diabetes cases, depending on the population (Malecki, 2010, Murphy et al., 2008). Maturity-onset diabetes of the young (MODY) is the commonest form of monogenic diabetes and is characterized by autosomal dominant inheritance and onset in children and young adults (<25 years of age). Monogenic diabetes that develops within the first 6 months of life is referred to as neonatal diabetes mellitus and may be transient (TNDM) or permanent (PNDM) or part of a syndrome, with extra-pancreatic features. Diabetes associated with monogenic severe insulin resistance is far less common than monogenic β-cell failure, and is caused by genetic defects affecting insulin signalling or adipocyte function (Semple et al., 2011).

Over 30 different genetic subtypes of MODY and NDM have been identified to date, each having a typical phenotype and a specific pattern of inheritance (Rubio-Cabezas et al., 2014, Murphy et al., 2008) (Table 1.1.). In approximately 50% of PNDM cases, the diabetes-causing

gene encodes a transcription factor, and the mutation leads to a structural defect of the pancreas and a decreased number of β-cells (Stekelenburg and Schwitzgebel, 2016). The identification of mutations in transcription factor genes in patients with diabetes, has provided great insights into human pancreas and beta cell development and function, and the clinical heterogeneity of diabetes. Next-generation sequencing (NGS), using targeted gene panels, enables simultaneous analysis of multiple genes, vastly reducing costs and leading to even higher rates of mutation detection (De Franco et al., 2015, Ellard et al., 2013). Whole exome or genome sequencing could subsequently be performed in cases when a pathogenic variant is not identified using the targeted NGS approach.

Gene	Function	Locus	Inheritance	Clinical features
HNF4A	TF	20q12-q13.1	AR	MODY, macrosomia and neonatal hypoglycemia, Renal Fanconi syndrome (mutation specific)
GCK	Hexokinase	7p13	AD or AR	MODY, mild asymptomatic hyperglycemia, isolated PNDM
HNF1A	TF	12q24.2	AD	MODY, renal glucosuria
HNF1B	TF	17q12	AD	MODY, TNDM, pancreatic hypoplasia, renal developmental abnormalities, genital tract malformations, gout, liver abnormalities
PLAGL1/ HYMAI	TF / Non-protein coding RNA	6q24	Variable (imprinting)	TNDM +/- macroglossia +/- umbilical hernia
ZFP57	TF	6p22.1	AR	TNDM, multiple hypomethylation syndrome +/- macroglossia +/- developmental delay +/- umbilical defects +/- congenital heart disease
PDX1	TF	13q12.1	AR	PNDM, pancreatic agenesis (steatorrhea)
PTF1A	TF	10p12.2	AR	PNDM, pancreatic agenesis, cerebellar hypoplasia/ aplasia, central respiratory dysfunction
PTF1A enhancer	TF	10p12.2	AR	PNDM, pancreatic agenesis without CNS features
RFX6	TF	6q22.1	AR	PNDM, intestinal atresia, gall bladder agenesis
GATA6	TF	18q11.1-q11.2	AD	PNDM, pancreatic agenesis, congenital heart defects, biliary abnormalities
GATA4	TF	8p23.1	AD	PNDM, pancreatic agenesis, congenital heart defects
GLIS3	TF	9p24.3-p23	AR	PNDM, congenital hypothyroidism, glaucoma, hepatic fibrosis + renal cysts
NEUROG3	TF	10q21.3	AR	PNDM + enteric anendocrinosis (malabsorptive diarrhoea)
NEUROD1	TF	2q32	AR	PNDM, cerebellar hypoplasia, visual impairment, deafness
PAX6	TF	11p13	AR	PNDM, microphthalmia, brain malformations
MNX1	TF	7q36.3	AR	PNDM, developmental delay, sacral agenesis, imperforate anus
NKX2-2	TF	20p11.22	AR	PNDM, developmental delay, hypotonia, short stature, deafness, constipation
KCNJ11	K^+ channel	11p15.1	Spontaneous or AD	PNDM/TNDM +/- DEND
ABCC8	K^+ channel	11p15.1	Spontaneous, AD or AR	PNDM/TNDM +/- DEND
INS	Hormone	11p15.5	Spontaneous, AD or AR	Isolated PNDM or TNDM
BLK	Protein tyrosine kinase	8p23.1	AD	MODY
KLF11	TF	2p25.1	AD	MODY
CEL	Lipolytic enzyme	9q34.13	AD	MODY, faecal elastase deficiency, steatorrhoea
SLC2A2 (GLUT2)	Glucose transporter	3q26.1-q26.3	AR	Fanconi–Bickel syndrome: PNDM, hypergalactosaemia,, liver dysfunction
SLC19A2	Thiamine transporter	1q23.3	AR	Roger's syndrome: PNDM, thiamine-responsive megaloblastic anemia, sensorineural deafness
SLC29A3	Nucleoside transporter	10q22.1	AR	Histiocytosis-lymphadenopathy plus syndrome, Pigmented hypertrichotic dermatosis
EIF2AK3	Protein kinase	2p11.2	AR	Wolcott–Rallison syndrome: PNDM, skeletal dysplasia, recurrent liver dysfunction
IER3IP1	ER membrane protein	18q21.2	AR	Microcephaly, epilepsy and diabetes syndrome: PNDM, microcephaly, lissencephaly, epileptic encephalopathy
FOXP3	TF	Xp11.23-p13.3	XR	IPEX syndrome: T1D, autoimmune enteropathy, eczema, autoimmune hypothyroidism, and elevated IgE
WFS1	Ion (Ca^{2+}) transporter	4p16.1	AR	Wolfram syndrome: DM, optic atrophy, diabetes insipidus, sensorineural deafness
CISD2	Ion (Ca^{2+}) transporter	4q24	AR	Wolfram syndrome 2: DM, sensorineural deafness, optic atrophy
STAT3	TF	17q21.31	AD	Early-onset polyautoimmunity; T1D, autoimmune hypothyroidism and celiac disease

Table 1.1. Common subtypes of monogenic diabetes (MODY and neonatal-onset diabetes). Modified from (Rubio-Cabezas et al., 2014).
DEND, developmental delay, epilepsy, and neonatal diabetes syndrome; ER, Endoplasmic reticulum; IPEX, immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome; MODY, maturity onset diabetes of the young; PNDM, permanent neonatal diabetes mellitus; T1D, type 1 diabetes; TF, Transcription factor; TNDM, transient neonatal diabetes mellitus.

1.2 Adult pancreas structure and function

The pancreas is an abdominal glandular organ, consisting of exocrine (90%) and endocrine (10%) cells. The exocrine tissue consists of acinar cells that secrete digestive enzymes (including trypsinogen, chymotrypsinogen, carboxypeptidase, pancreatic lipase, sterol esterase, phospholipase, pancreatic amylase) into a branched ductal network that drains into the gastrointestinal tract. The endocrine pancreas synthesizes and secretes hormones involved in the regulation of glucose, lipid and protein metabolism. The endocrine cells are located in clusters called the islets of Langerhans, scattered throughout pancreatic parenchyma. They include; insulin-secreting β-cells (50-60%), glucagon-secreting α-cells (30-45%), somatostatin-secreting δ-cells (5%), pancreatic polypeptide (PP)–secreting γ-cells (<5%) and gherlin-secreting ε-cells (<1%). Unlike in mice, where there is clear segregation of cell types in different regions of the islet, β-cells central, surrounded by α-, δ- and γ-cells; in humans β-cells are intermingled with α- and δ- cells, throughout the islet (Cabrera et al., 2006).

Single cell RNA-sequencing studies of the human pancreas and islet have been used to study heterogeneity and classify subpopulations within known cell types (Muraro et al., 2016). β-cells were generally regarded as a single, homogenous cell population. However, two recent studies have identified four antigenically distinct subtypes of human β-cells, which can be distinguished by diverse gene expression profiles and distinct basal and glucose-stimulated insulin secretion (GSIS) pattern (Bader et al., 2016, Dorrell et al., 2016).

1.3 Pancreas development

1.3.1 Endoderm development

Embryonic development is a complex process characterized by dynamic changes in gene expression, initiated by rapidly changing environmental cues, which affect cellular states by altering chromatin structure and gene transcription (Xie et al., 2013). Specification of the three germ layers, ectoderm, endoderm, and mesoderm, occurs during gastrulation in the developing embryo (14-19 d.p.c in humans and E6.5-7.5 in mice). It is from these three layers that all the organs of the body are derived through the process of organogenesis, which occurs during the 4th to 8th week of human embryonic development (E10.5-14.5 in mice). The pancreas, as well as other organs such as the liver, intestines, lungs and thyroid, are derived from the endoderm germ layer.

The definitive endoderm (DE) was first defined as the innermost germ layer found in all metazoan embryos (Zorn and Wells, 2009). The beginning of gastrulation is marked by the formation of a transient structure known as the primitive streak (PS) and molecular analyses and lineage mapping studies have defined posterior, mid, and anterior regions of the PS that differ in gene expression patterns and developmental potential. During gastrulation, uncommitted epiblast cells mobilise/undergo epithelial-mesenchymal transition (EMT), migrate through the PS, and exit either as mesoderm or DE. DE develops from epiblast cells that transit the most anterior region of the PS. After egression from the primitive streak, DE cells are incorporated into the pre-existing epithelium, the visceral endoderm (VE). As more DE is recruited into the endodermal layer during gastrulation, the incoming population expands anteriorly and laterally from the site of integration. The DE thereby progressively displaces the VE anteriorly and proximally to cover the extraembryonic region of the conceptus. At the end of gastrulation, the DE is a sheet of cells/squamous epithelium that surrounds the external surface of the mouse embryo (Lewis and Tam, 2006, Wells and Melton, 1999, Zorn and Wells, 2009).

In all vertebrates, the Nodal signalling pathway is necessary and sufficient to initiate endoderm and mesoderm development. It is required for initiation of gastrulation through regulation of wingless-type MMTV integration site (Wnt), fibroblast growth factor (FGF), and bone morphogenetic protein (BMP) pathways and for axial patterning. Many studies support a model in which high levels of Nodal signalling promote endoderm development, whereas lower doses specify mesoderm identity. Canonical Wnt signalling via β-catenin and Tcf transcription factors is also important for endoderm formation by stimulating high levels of Nodal gene transcription pre-gastrulation. Nodal signalling promotes the expression of a conserved network of transcription factors within the endodermal lineage, which include Mix-like proteins, Foxa2 (Forkhead box a2), Sox17 (SRY-related HMG Box 17), Eomesodermin, and Gata 4–6 (GATA binding protein 4-6) (Zorn and Wells, 2009).

1.3.2 Primitive gut tube development

Morphogenesis of the primitive gut, begins with folding of the epithelial sheet, resulting in formation of two intestinal pockets, in the anterior (foregut pocket) and the posterior (hindgut pocket) regions of the embryo. The foregut pocket initially develops as a crescent-shaped depression beneath the neural folds. It then deepens to form a pocket that extends rostrally.

Anterior/cephalo axial and lateral endoderm fold toward the ventral midline forming the foregut. The hindgut pocket forms in a similar manner in the posterior/caudal region. This results in the formation of a blind-ended endoderm-lined tubes. The middle part of the tube, the midgut, fuses more slowly due to the presence of the yolk sac. Gut tube formation, with foregut, midgut and hindgut domains, is complete by 22-26 d.p.c of development in humans and around E7.5-8 in mice (Jennings et al., 2013, Jennings et al., 2015). At this stage, the yolk sac is reduced to a thin stalk called the vitelline duct, which remains temporarily connected to the midgut.

As migration is occurring, the primitive gut simultaneously undergoes patterning along the anterior-posterior (AP) dorsal-ventral (DV) and axes and is delineated into three main areas: the foregut, midgut, and hindgut domains (Figure 1.1). The molecular mechanisms responsible for early patterning along the AP and DV axes in endoderm have been extensively studied. AP patterning is induced by the surrounding mesoderm, which secretes a variety of growth factors including transforming growth factor beta (TGFβ), FGF, Wnt, BMP, and retinoic acid (RA) signalling pathways. These pathways are known to regulate the expression of key transcription factors, including Fox/ Hepatic Nuclear Factor (HNF), ParaHox (i.e. Pdx1 and Cdx factors), and Homeodomain (Hox) factors, which are important mediators of cell fate (Grapin-Botton, 2005). Haematopoietically expressed homeobox (Hhex), FoxA2 and Sox2 (pluripotency marker which re-emerges as a foregut marker) are expressed in the anterior endoderm, and Pancreatic and Duodenal Homeobox 1 (Pdx1), Caudal type homeobox 1 and 2 (Cdx1 and Cdx2) in the posterior endoderm (Zorn and Wells, 2009). The pattern of expression of members of the HOX family of transcription factors also plays an important role in AP patterning.

The same signalling pathways are used reiteratively throughout development and that the same cells can respond differently depending on the stage. For example, Wnt and FGF signalling are first required to initiate gastrulation and generate the early anterior endoderm. However, in late gastrulation, Wnt and FGF4 repress anterior fate and must be excluded to maintain foregut progenitors (Zorn and Wells, 2009). RA, secreted by the mesoderm, is important in establishing the foregut-hindgut boundary and increasing concentrations promote a posterior fate (Chen et al., 2004, Stafford and Prince, 2002).

It is from the foregut region that many of the major organ systems develop. Anatomically, the foregut can further be divided into two portions, the anterior and posterior foregut, with the former giving rise to the thyroid, oesophagus, upper airways, and lungs and the latter to the stomach, pancreas, liver and biliary system. The foregut can also be divided into ventral and

dorsal foregut. DV patterning is more poorly understood compared to AP patterning, however it is important for separating the ventral trachea from the dorsal oesophagus and the ventral and dorsal buds of the pancreas.

There is evidence that in addition to the signals that lead to the regionalization of endoderm along the AP axis, other signals are necessary locally for the induction of organ primordia. For example, the high FGF/BMP environment adjacent to the cardiac mesoderm favours a liver fate with cells expressing Hhex, and foregut cells caudal to the Hhex domain co-express Pdx1 and Sox17 and contribute to the ventral pancreas and biliary system (Rossi et al., 2001, Spagnoli and Brivanlou, 2008, Spence et al., 2009). Signalling pathways and molecular mechanisms involved in pancreas development from the posterior foregut will be discussed in more detail in the next section.

Figure 1.1. Expression pattern of transcription factors along the AP axis of the gut tube (Zorn and Wells, 2009)

1.3.3 Pancreas development in mice

1.3.3.1 Pancreatic specification

In the mouse, pancreas specification within the foregut endoderm is first indicated by expression of the transcription factor Pdx1 (Jonsson et al., 1994, Ohlsson et al., 1993) and is first evident at E9-9.5. The pancreas arises from distinct dorsal and ventral buds, which are induced by different mesodermal structures. Important signalling pathways that are involved in establishing the pancreatic domain within the developing gut tube and further development of the pancreas include the FGF, BMP, Wnt, RA, Hedgehog (HH), TGFβ and Notch signalling pathways.

The ventral pancreas fate is induced in the portion of the ventral foregut that has low levels of cardiac FGF - Mitogen-activated protein kinases (MAPK) signalling. The low levels of Fgf2 promote expression of the homeobox gene Pdx1 in the ventral pancreas (Serls et al., 2005, Zaret, 2008). RA-mediated signalling functions to pattern the foregut before specification of the pancreatic domain. Dorsal pancreatic development is particularly disrupted in the absence of RA signalling due to absence of the enzyme retinaldehyde dehydrogenase 2 (Molotkov et al., 2005). Fgf10 is necessary for the development of several organs budding off the digestive tract (Bhushan et al., 2001). Genetic analysis suggests that Foxf1 and Fgf10 in the surrounding mesenchyme regulate proliferation and correct positioning of the liver, pancreas, and various hepatobiliary ducts (Dong et al., 2007, Kalinichenko et al., 2002).

In contrast to most of the digestive tract epithelium, the pancreas can only develop in an area free of Sonic hedgehog (Shh). Shh is broadly expressed by the endoderm of the developing gut tube along the AP axis. Coincident with the start of pancreas development, Shh expression is repressed in the dorsal and ventral pancreatic buds. The repression of Shh in the dorsal bud depends on signals from the notochord that include FGF2 and Activin (Hebrok et al., 1998, Kim et al., 1997). The TGFβ signalling molecule Activin, secreted from the notochord, is able to repress expression of Shh and induce Pdx1 expression in dorsal endoderm in a dose-dependent manner. Activin appears to play a role in AP patterning which has a secondary effect on pancreas specification. Additional studies are needed to clarify the role of TGF family members in endoderm patterning and pancreas specification (McCracken and Wells, 2012).

In addition to Pdx1, which is necessary for normal pancreas development, there are several

other transcription factors involved in foregut patterning and early pancreas development, including, Hnf4a, Hnf6 (Onecut1), Hnf1b, Motor Neuron And Pancreas Homeobox 1 (Mnx1; also known as Hlxb9), and Ptf1a (Grapin-Botton, 2005). Both the dorsal and ventral buds, pancreatic progenitors express the essential transcription factor Pdx1 and Pancreas Specific Transcription Factor 1a (Ptf1a), but only the dorsal bud expresses Hoxb9 (Sherwood et al., 2009). Genetic lineage-tracing in mice indicates that essentially all mature pancreatic cells, including acinar, ductal and endocrine cells, arise from *Pdx1*$^{+}$/*Ptf1a*$^{+}$ progenitor cells (Gu et al., 2002, Kawaguchi et al., 2002, Jonsson et al., 1994, Zhou et al., 2007).

1.3.3.2 Development of pancreatic endocrine, ductal and exocrine progenitor cells

After pancreatic specification and the onset of bud formation, there are continued signalling processes that regulate the proliferation and differentiation of progenitor cells for the endocrine, ductal and exocrine lineages. Multipotent progenitor cells initially go through a period of rapid proliferation under the influence of Fgf10, forming a branching epithelial network of the pancreas (Bhushan et al., 2001). At around E13 the epithelium becomes segregated into distinct ‘‘tip’’ and ‘‘trunk’’ domains. While the tip domains contain multipotent pancreatic progenitor cells, which are capable of differentiating into into pro-acinar, the adjacent trunk epithelial region consists of a duct/endocrine bipotential progenitor pool (Zhou et al., 2007).

Ptf1a appears to be the key player in acinar formation and its loss results in acinar agenesis. (Fukuda et al., 2008). The β-catenin/Wnt signalling pathway is required for maintenance of Ptf1a exocrine progenitor cells (Murtaugh et al., 2005, Wells et al., 2007). Loss-of-function experiments in mice have identified several transcription factors that play a role in specifying the multiple endocrine lineages in the developing pancreas. Ptf1a and NK Homeobox 6-1 (Nkx6-1) have an antagonistic relationship and the segregation of Nkx6-1 and Ptf1a expression into exclusive trunk and tip compartments, respectively, appears to correlate with the progressive restriction of trunk and tip cells towards the future duct/endocrine, or acinar lineages (Schaffer et al., 2010).

In the mouse, endocrine specification via the generation of Neurogenin 3 (Neurog3^{+}) endocrine precursors occurs in two major waves termed the primary and secondary transition. The primary transition takes place during the bud stages of early pancreas development, where Neurog3 transcripts can be detected in the presumptive pancreatic anlage as early as E8.5-E9.5. The majority of endocrine precursors are however generated from the trunk plexus during the

secondary transition, peaking in absolute numbers at around E15.5 (Larsen and Grapin-Botton, 2017). The level of Notch signalling has been demonstrated to exert antagonistic effects on Neurog3 transcription. Low levels of Notch signalling induce expression of Sox9, which positively regulates Neurog3 expression whereas, high levels of Notch signalling induce both Sox9 and Hes Family BHLH Transcription Factor 1 (Hes1) leading to repression of Neurog3 expression (Larsen and Grapin-Botton, 2017). The remaining non-delaminating Hnf1b^{+}/Sox9^{+}/Nkx6-1^{+} trunk cells are fated to become duct cells. Precocious endocrine differentiation e.g. from deficiency in Notch pathway components, such as Recombination Signal Binding Protein For Immunoglobulin Kappa J Region (Rbpj), Delta Like Canonical Notch Ligand 1 (Dll1), and Hes1 results in pancreatic hypoplasia due to Ptf1a^{+} progenitor pool depletion (Apelqvist et al., 1999, Fujikura et al., 2006, Jensen et al., 2000). Neuronal differentiation (NeuroD), INSM Transcriptional Repressor 1 (Insm1) and ISL LIM Homeobox 1 (Isl1) are all required for the development of multiple endocrine cell lineages and are thought to act downstream of Neurog3 (Ahlgren et al., 1997, Gierl et al., 2006, Naya et al., 1997).

Neurog3 expression defines the development of a progenitor population for all endocrine lineages, including α-cells, β-cells, δ-cells, γ-cells and ε-cells (Gu et al., 2002). Endocrine progenitor cells delaminate from the trunk epithelium and coalesce to form islets. With further maturation, complex networks of factors function to establish the different endocrine lineages. Expression of transcription factors in a combinatorial manner further restricts cells into distinct mono-hormonal endocrine lineages; α-cells express Paired box 6 (Pax6), Nkx2.2, Foxa2, Aristaless Related Homeobox (Arx) and MAF BZIP Transcription Factor B (MafB), β-cells express Pax4, Pax6, Nkx2.2, Nkx6-1, Pdx1, Mnx1 and MafA, and δ-cells express Pax4, Pax6 (Conrad et al., 2014, Oliver-Krasinski and Stoffers, 2008). Surprisingly little is known about the signalling pathways that promote these specific endocrine lineages during embryonic development in vivo (Spence and Wells, 2007).

1.3.4 Timeline and regulation of human pancreas development

Most of our understanding of the molecular mechanisms involved in human pancreas development is derived from studies initiated at 7–8 weeks of fetal development, by which time the two pancreatic buds have coalesced as a single organ, and endocrine commitment has already started (Figure 1.2). Recently however, a study by Jennings et al. (Jennings et al., 2013) examined specimens obtained from human embryos from as early as human embryonic stage Carnegie Stage 9 (CS9) or 22-26 days post conception (d.p.c.), providing important insights

into earlier development and pancreas specification.

In humans, as in mice, the pancreas develops from dorsal and ventral pancreatic buds, which emerge as distinct outgrowths from the posterior foregut endoderm. Foregut endoderm lying adjacent to the notochord can be observed at 25-27 d.p.c (CS9). SHH, which must be excluded for pancreatic buds to develop can be detected at CS10 and PDX1 expression is first seen at CS12 (29-31 d.p.c.) (Jennings et al., 2013, Jennings et al., 2015). It appears that human foregut and pancreatic patterning is similar to mice. In mice, a "primary transition" stage occurs at between E9.5–E12.5. The first differentiated endocrine cells, mainly glucagon-producing a-cells, begin to appear in the dorsal bud at this phase of development, in a cluster-budding process, rather than by individual delamination of epithelial cells. By contrast there has been no evidence of early pancreatic endocrine differentiation in human embryos possibly due to lack of proximity between the paired dorsal aortae and the early pancreatic endoderm (Jennings et al., 2013).

The ventral and dorsal pancreatic buds at CS13 (30-33 d.p.c.) are marked by SOX9, PDX1 and GATA4 all of which are required for human pancreas development based on the phenotype of patients with loss of function mutations in these genes (Piper et al., 2002, Shaw-Smith et al., 2014, Stoffers et al., 1997). Significant growth of the pancreas and proliferation of MPCs occurs between CS13 to CS19. Unlike in mice NKX2.2 is not detected in these cells. Further differences in pancreatic progenitor cells are seen from CS19 (45-47 d.p.c), equivalent to E14-E14.5 in mice. $SOX9^{+}/NKX6\text{-}1^{+}$ "trunk" cells show low expression of GATA4, whereas peripheral "tip" cells are $SOX9^{+}/NKX6\text{-}1^{+}/GATA4^{+}$. By 10 w.p.c. pro-acinar "tip" cells no longer express NKX6-1, similar to in mice. Loss of SOX9 expression in acinar cells is delayed in humans (at 10-14 weeks post conception (w.p.c)) compared with mice, in which Sox9 is lost from peripheral tip cells shortly following their formation, at around E14.5-E15 (Jennings et al., 2015).

In humans, detection of NEUROG3 (or NGN3) first occurs at CS21 (49-52 d.p.c) closely followed by detection of insulin positive cells, which is the first islet cell type to appear (Jennings et al., 2013, Piper et al., 2004). SOX9 expression persists in pancreatic duct cells, but is absent in NEUROG3 positive endocrine progenitors. Transient expression of NEUROG3 is also seen in humans as in mice (Jennings et al., 2015). The ventral pancreatic bud migrates toward the dorsal bud and both buds fuse together resulting a single pancreatic organ by the end of embryonic development (CS23, 8^{th} week). Between 10 to 12 weeks the proliferation of

branched tubules into primitive acini, islets and ducts occurs with limited differentiation (Conrad et al., 2014, Pan and Wright, 2011, Jennings et al., 2013). Islet clusters containing all four endocrine cell types are observed by the 12th to 13th week (Piper et al., 2004). (Conrad et al., 2014, Pan and Wright, 2011, Jennings et al., 2013).

Although most genes and signalling pathways play conserved roles in mice and humans, obvious species-specific differences exist in gestation period, morphology, spatial and temporal regulation of gene expression and gene-dose sensitivity during embryonic pancreas development. Consequently, mouse models do not always fully replicate the features of human pancreatic diseases (Elsea and Lucas, 2002). For example, discrepant phenotypes between mice and humans are seen in haploinsufficient forms of monogenic diabetes resulting from mutations in the transcription factors HNF1A, HNF4A, HNF1B or GATA6 (Flanagan et al., 2014a). There is still much to be learned regarding, foregut development, the potential for patterning by nearby structures, and the regulatory relationships between signalling pathways and transcription factors and how these processes guide pancreatic specification, endocrine differentiation, and endocrine and exocrine cell maturation (Jennings et al., 2013).

The next section of this chapter will focus on the transcription factor HNF1B and its role in development and human disease.

Figure 1.2. Transcription factor regulatory network in human pancreas development (Al-Khawaga et al., 2017).

1.4 Role of HNF1B in development and human disease

1.4.1 HNF1B-associated disease

Hepatocyte nuclear factor 1β (HNF1B, TCF2, HNF1beta, MODY5) is a member of the Pituitary-specific Pit-1, Octamer-binding transcription factor (Oct-1/2), UNC-86 (POU) homeodomain-containing superfamily of transcription factors. Early expression of HNF1B is seen in the kidney, liver, pancreas, bile ducts, urogenital tract, lung, thymus and gut, where it plays an important role in normal development and tissue-specific regulation of gene expression in these organs (Barbacci et al., 1999, Coffinier et al., 1999a).

In humans, heterozygous mutations in HNF1B result in a multisystem disorder. HNF1B was initially identified as a monogenic diabetes gene and the first HNF1B gene mutation (R177X) was described in a Japanese family with MODY in 1997 (Horikawa et al., 1997). MODY is the most common form of monogenic diabetes and is characterized by autosomal dominant inheritance, onset typically before 25 years of age and β-cell dysfunction. Renal disease, especially renal cysts, was subsequently found to be the main clinical feature in patients with HNF1B gene mutations and the association of renal cysts with diabetes led to the description of renal cyst and diabetes syndrome. Additional phenotypes observed in subsets of patients include pancreatic hypoplasia, reduced exocrine function, biliary disorders, abnormal liver function tests, urogenital tract abnormalities, hyperuricaemia, early-onset gout, hypomagnesaemia and primary hyperparathyroidism (Bingham and Hattersley, 2004, Clissold et al., 2014, Edghill et al., 2006a). The clinical features of HNF1B-associated disease are summarised in Figure 1.3.

1.4.2 Human molecular genetics

In humans, the HNF1B gene is located on chromosome 17q12 and the encoded protein has three functional domains: the N-terminal dimerization domain, the DNA-binding domain and the C-terminal transactivation domain. The DNA-binding domain is characterized by a POU-specific domain (POUs) and an atypical homeodomain, referred to as POU-homeodomain (POUh). HNF1B is structurally similar to hepatocyte nuclear factor 1A (HNF1A, TCF1, MODY3) and shares >80% sequence homology in the DNA-binding domain. The protein binds to DNA as either a homodimer or a heterodimer with HNF1A to regulate gene expression (Mendel and Crabtree, 1991, Mendel et al., 1991).

HNF1B mutations comprise base substitutions, small insertions or deletions, or whole-gene deletions. More than 50 splice-site, nonsense, missense and frame-shift mutations have been reported to date (Figure 1.4) (Clissold et al., 2014). The majority of identified mutations are clustered in the first four exons of the gene, particularly in exons 2 and 4 and the intron 2 splice site. Whole-gene deletions of HNF1B were identified 8 years after the first coding mutation (Bellanne-Chantelot et al., 2005). The region of chromosome 17 that encompasses the HNF1B gene is susceptible to genomic rearrangement and whole-gene deletions occur because of non-allelic homologous recombination of chromosome 17q12, mediated by segmental duplications flanking a 1.5-Mb region that encompasses HNF1B (Mefford et al., 2007). The chromosomal microdeletion at 17q12, also involves a further 14 genes. HNF1B gene mutations and small insertions and deletions can be detected by sequencing, whereas partial or whole gene deletions are detected by multiple ligation-dependent probe amplification (MLPA) HNF1B dosage assay. There is no evidence available to suggest that patients with whole-gene deletions exhibit a different phenotype to those with coding or splice site mutations, suggesting that dysfunction is because of a gene dosage effect, i.e, haploinsufficiency.

HNF1B-associated disease is generally considered to exhibit autosomal dominant inheritance; however, de novo mutations and whole-gene deletions account for up to 50% of cases, so a family history of renal disease or diabetes mellitus may be absent (Clissold et al., 2014, Edghill et al., 2006a). The phenotype of HNF1B mutation carriers is extremely variable and there is no clear evidence for a genotype-phenotype relationship (Clissold et al., 2014, Edghill et al., 2006a). The reasons for phenotypic variation in HNF1B-associated disease remain poorly understood and may reflect the functional effects of different gene anomalies, stochastic variation in temporal HNF1B gene expression during early development or additional genetic and/or environmental modifiers, which might influence the HNF1B-associated disease phenotype (Clissold et al., 2014, Edghill et al., 2006a).

Figure 1.3. Pancreatic and extra-pancreatic phenotypes frequently observed among patients with HNF1B-associated disease (Clissold et al., 2014)

Figure 1.4. Hepatocyte nuclear factor 1β (HNF1B) gene structure illustrating >50 different mutations identified to date, which span the first eight (out of nine) exons of the gene. The exons and protein domains are indicated. The HNF1B protein contains 557 amino acids with three distinct domains: the dimerization domain (1-32 amino acids), the DNA-binding domain (90-311 amino acids) and the transactivation domain (312-557 amino acids) (El-Khairi and Vallier, 2016).

To gain insight into the molecular mechanisms underlying the varied phenotypes in HNF1B-associated disease, eight disease-causing mutations predicted to produce protein truncations, amino acids substitutions or frameshift deletions in different domains of the protein were functionally characterized (Barbacci et al., 2004). It was found that diverse mechanisms may be involved. Some mutations, which affected DNA binding, possessed a weak dominant-negative activity, whereas others, with intact or partially decreased DNA-binding activity, primarily impaired the transactivation potential of the protein. The transactivation potential of HNF1B depends on the synergistic action of the histone-acetyltransferases CREB-binding protein (CBP) and p300/CBP-associated factor. The transcriptional impairment of HNF1B mutants with normal DNA-binding activity correlated with the loss of its association with one of these co-activators in co-immunoprecipitation studies. HNF1B-associated disease may therefore result not only from defective DNA binding but also from diminished transactivation function through impaired recruitment of co-activator proteins (Barbacci et al., 2004).

The mutational mechanisms of two splice site mutations in HNF1B, which result in the deletion of exon 2, and are predicted to cause premature termination of the HNF1B protein, have been examined (Harries et al., 2004). As tissues with high native HNF1B expression are not easily accessible, ectopic transcription was used to allow the examination of transcripts in transformed lymphocytes. It was shown that the mutated splice variants are present at a lower level than the wildtype transcripts and inclusion of a nonsense mediated decay (NMD) inhibitor failed to alter the relative transcript frequencies. This suggests that the reduction of the mutant transcript frequency is unlikely to be because of NMD and it is more likely that the stability of the mutant splice variants is compromised by other mechanisms (Harries et al., 2004).

Another study by Harries et al. (Harries et al., 2005) further examined the susceptibility of six truncating HNF1B mutations to NMD. Four of the six mutations (R181X, Q243fsdelC, P328L329fsdelCCTCT and A373fsdel29) showed evidence of NMD with varying reductions in levels of mutant transcripts (3%-71%). Two mutations (H69fsdelAC and P159fsdelT), located most proximally, produced transcripts that did not undergo NMD. Susceptibility to NMD was therefore found to depend on the position of the mutation and showed 50 to 30 polarity, suggesting that re-initiation of translation may occur and be an important mechanism in the evasion of NMD.

HNF1B has been identified as a bookmarking factor. During cell cycle progression in higher eukaryotes, after nuclear envelope breakdown, transcription factors are normally released from

chromatin leading to the transcriptional silencing that accompanies mitosis. Using time-lapse microscopy of IMCD3 mouse renal cells expressing fluorescence-tagged Hnf1b, Verdeguer et al. (Verdeguer et al., 2010) found that, a substantial fraction of Hnf1b remained associated with mitotic chromatin and travelled with condensed chromosomes throughout mitosis. The same effect was not observed for Hnf4a. In Hnf1b mutant mice, a severe polycystic kidney phenotype is observed only when this transcription factor is inactivated in highly proliferating tubular cells. These cells are cycling very intensely and probably need to promptly re-establish the expression of cystic disease–associated genes (Uromodulin [Umod], Fibrocystin/Polyductin [Pkhd1], Polycystin 2, Transient Receptor Potential Cation Channel [Pkd2], Kinesin Family Member 12 [Kif12], Crumbs 3, Cell Polarity Complex Component [Crb3], Tcfap2b, Transmembrane Protein 27 [Tmem27] and BicC Family RNA Binding Protein 1 [Bicc1]) as soon as the cells re-enter interphase. The fact that Hnf1b remains associated with mitotic chromatin during mitosis in these proliferating cells might provide a rapid mechanism to reactivate the expression of crucial target genes necessary for maintaining proper alignment of cell division during tubular elongation and preventing tubular dilation. In Hnf1b deficiency, the expression of these cystogenic target genes is lost and proliferation triggers a positive feedback loop leading to the persistence of a spatially distorted hyperproliferation. Hnf1b may therefore function as both a classic transcriptional activator and as a bookmarking factor that marks target genes for rapid transcriptional reactivation after mitosis.

A recent study by Lerner et al. (Lerner et al., 2016) looked at the bookmarking effect of HNF1B in transfected human IMCD3 cells with induced point mutations in GTP-HNF1B (P256S, V265L and G287S) and UREC cells in a patient with the HNF1B mutation P256S. The ability of HNF1B to localise to mitotic chromatin was impaired by naturally occurring mutations found in patients. Importin-β is involved in the maintenance of the mitotic retention of HNF1B, suggesting a functional link between the nuclear import system and the mitotic localization. This suggests that HNF1B is involved in mitotic transmission of epigenetic information via its ability to remain associated with mitotic chromatin and this may be a mechanism connecting genetically determined predispositions and environmentally induced proliferative stress conditions.

1.4.3 HNF1B-associated pancreatic disease

1.4.3.1 HNF1B-associated diabetes

Diabetes mellitus is the most frequent extra-renal phenotype detected in patients with HNF1B-associated disorder and usually presents following renal disease. HNF1B mutations are an uncommon cause of MODY (HNF1B-MODY or MODY5) and represent approximately 1%-6% of MODY cases in the UK depending on the cohort studied (Beards et al., 1998, Shields et al., 2010). Diabetes has been described in 58% of reported HNF1B mutation carriers and a further 4% have impaired glucose tolerance (Bingham and Hattersley, 2004). Diabetes typically develops during adolescence or early adulthood and the mean age at diagnosis of diabetes is 26 years, with a range of 10-61 years. Transient neonatal diabetes has also been reported in a small number of cases (Edghill et al., 2006b). Importantly, the age of onset of diabetes shows no correlation with the specific gene mutation.

The pathophysiology of diabetes mellitus in patients with HNF1B mutations is mainly attributed to β-cell dysfunction and reduced insulin secretion, which is likely to be a consequence of pancreatic hypoplasia. Of note, marked reduction in birth weight, with a median birth weight of 2.4 kg, is a consistent feature possibly reflecting reduced insulin secretion in utero (Edghill et al., 2006b). The majority of patients with HNF1B-associated diabetes mellitus require early treatment with insulin as they respond poorly to sulphonylureas. Thus, HNF1B haploinsufficiency could have an effect on early pancreatic development and the adult disease could be provoked by a decrease in the number of β-cells available at birth.

In contrast to patients with HNF1A-asociated diabetes, patients with HNF1B mutations can develop reduced insulin sensitivity to endogenous glucose production (i.e, hepatic insulin resistance) with normal peripheral insulin sensitivity, in addition to β-cell dysfunction (Brackenridge et al., 2006, Pearson et al., 2004). This can result in hyperinsulinaemia and associated dyslipidaemia, with raised levels of triglycerides and reduced levels of high-density lipoprotein.

In some patients, reduced insulin sensitivity owing to weight gain, puberty, infection, stress or other causes may unmask an underlying defect in β-cell function (Furuta et al., 2002). Diabetes mellitus may also present as new-onset diabetes mellitus after transplantation, and analysis of the HNF1B gene should be considered in all individuals with unexplained congenital anomalies

of the kidneys and urinary tract undergoing renal transplantation. Many immunosuppressive agents (calcineurin inhibitors, mTOR inhibitors, and steroids) inhibit insulin transcription and/or secretion, and steroids decrease insulin sensitivity (Bamgbola, 2016, Faguer et al., 2016). In fact, a recent in vitro study showed that exposure to a calcineurin inhibitor resulted in decreased HNF1B expression in epithelial cells, although the mechanism for this is not understood (Faguer et al., 2016). Therefore, careful monitoring of plasma glucose after transplantation is fundamental, irrespective of the pretransplantation diabetes status of patients with HNF1B.

Genome-wide association studies have found that approximately one third of monogenic diabetes genes, including HNF1B, are associated with common T2D but not with T1D. For example, Winckler et al. found that the single nucleotide polymorphism (SNP) rs757210A in intron 2 of HNF1B showed a strong association with T2D in multiple populations (Morris et al., 2012, Winckler et al., 2007). Thus, modulation of HNF1B expression by common genetic variants can also have in role in predisposition to T2D.

1.4.3.2 HNF1B-associated pancreatic exocrine disease

Pancreatic hypoplasia or agenesis and exocrine dysfunction, which is usually asymptomatic and mainly manifested by reduced faecal elastase concentration, have been reported in some patients with HNF1B mutations. Extensive radiological examinations of the pancreas in a small cohort of patients showed absence of the pancreatic body and tail, and a slightly atrophic pancreatic head, indicating agenesis of the dorsal pancreas (Haldorsen et al., 2008). These findings are consistent with data from studies on mice, showing the essential role of Hnf1b in the development of ventral and dorsal pancreatic buds. Pancreatic exocrine hypersecretion has also been observed in a number of affected individuals, and could be a compensatory mechanism for diminished pancreatic volume (Clissold et al., 2014).

The pancreatic and renal phenotype of two foetuses, a 27-week male and a 31.5-week female, carrying a frameshift mutation in exons 2 and 7 of HNF1B, respectively, and whose gestation was terminated following the diagnosis of a foetal form of polycystic renal disease, has been described (Haumaitre et al., 2006). Both foetuses present the more severe phenotypes described in HNF1B mutation carriers, specifically bilateral enlarged polycystic kidneys, severe pancreas hypoplasia and abnormal genital tract. Three further foetuses with severe HNF1B-associated kidney and renal disease have also been described (Body-Bechou et al., 2014). Global

histological analysis of the mutant pancreatic tissue from one foetus revealed a severe hypoplasia of both the tail and body, which resulted mainly from underdeveloped, small acini. The exocrine pancreas, although severely hypoplastic, appeared functionally normal as shown by the expression of the digestive hormone carboxypeptidase-A. In addition, although there was no significant decrease in insulin, glucagon and somatostatin expression, the islets of Langerhans appeared slightly disorganized and the density of β-cells was moderately reduced (Haumaitre et al., 2006). The authors postulated that the partial disorganization of both exocrine and endocrine pancreas could be explained by the low expression of β-catenin and E-cadherin that was observed in the mutant pancreatic tissue. Expression of glucose transporter 2 (GLUT2), the low-affinity GLUT essential for glucose sensing, was undetectable, suggesting defective β-cell differentiation. Expression of GLUT2 is likely to be mediated directly by HNF1B, as functional HNF1A and HNF1B binding sites are present in the GLUT2 promoter (Cha et al., 2000).

In summary, haploinsufficiency of HNF1B expression during development results in impaired pancreas organogenesis and differentiation to endocrine and exocrine cells. Much of our knowledge regarding pancreas development is derived from studies in mice; however, the precise molecular mechanisms by which HNF1B mutations impair human pancreas development and function remain poorly understood.

1.4.4 HNF1B-associated liver and biliary disease

Liver dysfunction is a common clinical finding in patients with HNF1B gene mutations and usually manifests as an asymptomatic rise in the levels of liver enzymes, particularly alanine aminotransferase and γ-glutamyl transferase. The severity of liver disease in patients ranges from neonatal cholestasis with or without an atrophic gallbladder to adult onset cholestasis to non-cholestatic liver impairment, which may be asymptomatic (Kotalova et al., 2015). Liver biopsies in neonates with cholestasis, showed a reduction in the number of intrahepatic bile ducts (IHBDs), biliary stasis and periportal fibrosis (Beckers et al., 2007, Kitanaka et al., 2004, Kotalova et al., 2015, Raile et al., 2009). Liver biopsies performed in three adult patients with cholestasis and abnormal liver function tests found no structural abnormalities of the bile ducts; however, immunohistochemistry showed fibrosis around the bile ducts and electron microscopy showed a reduction of normal primary cilia on the bile duct epithelial cells, which could contribute to cholestasis (Roelandt et al., 2012).

In a recent study, contrast-enhanced MRI and magnetic resonance cholangiopancreatography (MRCP), were used to detect hepatic and pancreatico-biliary pathologies in patients with HNF1B mutations (Kettunen et al., 2017). Structural anomalies of the bile ducts were found in seven of 14 patients (50%). These included varying types of choledochal cysts or bile duct cysts (BDCs) in the extrahepatic and/or intrahepatic bile ducts (n=6), adenomyomatosis of the gallbladder (n=2), adenomatous tissue near the ampulla of Vater (n=1) and intraductal papillary mucinous neoplasia-like cystic dilatations of the pancreatic branch ducts (n=1). The malignant potential of HNF1B-associated choledochal cysts warrants further studies. The liver parenchyma appeared normal in all cases however, seven patients had abnormal serum ALT levels, which were generally less than twice the upper reference limit.

HNF1B has a key function in controlling biliary system development in humans. However, little information has been obtained so far concerning the role of HNF1B in hepatocyte-specific enzymatic and metabolic function. The finding of hepatic insulin resistance in patients with HNF1B mutations described previously is particularly interesting, and further studies are required to define the role of HNF1B in mediating insulin action and in the regulation of hepatic lipid and glucose metabolism.

1.4.5 HNF1B-associated renal disease

Renal disease is the most consistent phenotype and is present in almost all reported probands with HNF1B gene mutations. The wide spectrum of renal diseases associated with HNF1B mutations can all be classified as disorders of renal development also known as congenital abnormalities of the kidneys and urinary tract (CAKUT). CAKUT are the most predominant developmental disorders in humans comprising ~20-30% of all anomalies identified in the prenatal period (Clissold et al., 2014). Mutations in HNF1B are the most common single gene cause of CAKUT (Vivante et al., 2014). Analysis of nine study cohorts of over 50 individuals with renal disease estimated that the prevalence of HNF1B gene anomalies was 19%, ranging from 5% to 31% depending on the phenotypic selection of the cohort.

Renal abnormalities are frequently detected on antenatal ultrasound scans from as early as 17-weeks gestation in patients with HNF1B mutations. Renal cystic disease is the most common manifestation in both paediatric and adult population however, other renal histologies and morphological abnormalities that have been described include; glomerulocystic kidney disease (GCKD), cystic renal dysplasia, enlarged glomeruli and/or nephrons (including

oligomeganephronia), interstitial fibrosis, renal hypoplasia, single kidney, horseshoe kidney, duplex kidney, collecting system abnormalities and bilateral hydronephrosis. *HNF1B* mutations have also been found in 2 of 34 individuals with prune-belly syndrome, which is characterized by a triad of dilatation of the urinary tract, deficiency or absence of the abdominal wall musculature and bilateral undescended testes. Patients with HNF1B mutations may have renal function that ranges from normal to end-stage renal failure requiring dialysis or transplantation (Clissold et al., 2014).

Patients with HNF1B gene anomalies may develop electrolyte abnormalities and hypomagnesaemia is a common feature of HNF1B-associated disease and can be accompanied by hypermagnesuria and hypocalciuria. HNF1B regulates the transcription of FXYD Domain Containing Ion Transport Regulator 2 (FXYD2), a gene that encodes the γ subunit of the Na^+/K^+–ATPase and is involved in the reabsorption of magnesium in the distal convoluted tubule. Hyperuricaemia is also frequently found in patients with HNF1B mutations. Patients may present with early-onset gout, and some affected individuals with hyperuricaemia, early-onset gout and renal disease meet established criteria for familial juvenile hyperuricaemic nephropathy, a condition usually caused by mutations in UMOD. The cause of hyperuricaemia in HNF1B-associated disease is considered to reflect both altered urate transport in the kidney and an early manifestation of renal impairment.

1.4.6 HNF1B-associated genital tract malformations

Congenital abnormalities of the genital tract are frequently found in patients with HNF1B mutations. These genital tract malformations can occur in both derivatives of the Mullerian and Wolffian ducts. The majority of malformations occur in females and result from aplasia and failure of the fusion of the Müllerian ducts, resulting in congenital uterine and upper vaginal abnormalities or aplasia. In a cohort of 108 females with congenital uterine abnormalities, heterozygous mutation or deletion of HNF1B was found in 9 of 50 patients (18%) who had both uterine and renal abnormalities, but interestingly in none of the 58 cases with isolated uterine abnormalities (Oram et al., 2010). HNF1B is also a candidate gene for Mayer–Rokitansky–Küster syndrome, which involves congenital aplasia of the uterus, cervix and upper vagina resulting in primary amenorrhoea and infertility. Genetic studies so far have only associated this syndrome with chromosomal microdeletions, including the 17q12 deletion (Ledig et al., 2011). Various genital tract malformations, including cryptorchidism, agenesis of the vas deferens, hypospadias, epididymal cysts and asthenospermia, have been sporadically

reported among males, however due to the small number of cases it is unclear whether these are coincidental findings, or part of the CAKUT spectrum, in which HNF1B mutations are the most commonly identified genetic cause (Clissold et al., 2014).

1.4.7 Role of HNF1B in malignancy

It has been suggested that imaging to screen for chromophobe renal cell carcinoma (RCC), a rare subtype of kidney cancer, should be considered for individuals with HNF1B gene anomalies. In a series of 34 randomly selected renal neoplasms screened for HNF1B gene inactivation, biallelic inactivation was identified in one of eleven tissue samples owing to the development of a somatic HNF1B gene deletion in addition to a germline mutation (Rebouissou et al., 2005). Overexpression of HNF1B has been found to be common in clear cell ovarian cancer (Tsuchiya et al., 2003). Furthermore, several genome-wide association studies have also linked genetic variation in the HNF1B region with a risk of endometrial and prostate cancer (Gudmundsson et al., 2007, Spurdle et al., 2011, Sun et al., 2008, Thomas et al., 2008).

1.4.8 Other clinical features

The possible association between HNF1B gene anomalies with neurodevelopmental disorders has been receiving increasing interest. Chromosome 17q12 deletions have been found in 18 of 15,749 patients referred for clinical genetic testing for autism spectrum disorders, developmental delay or cognitive impairment (Moreno-De-Luca et al., 2010). Seizures, structural brain abnormalities, mild facial dysmorphic features and macrocephaly have also been reported. The deleted stretch of DNA contains 15 genes, including HNF1B; therefore, it is not clear what genetic mechanism gives rise to these observed neurodevelopmental phenotypes (Clissold et al., 2014). In a cohort of 53 children with whole gene deletion of HNF1B and cystic kidney disease, three had a diagnosis of autism, which is more common than the prevalence of autism recorded in the general paediatric population (1:150–1:300) (Loirat et al., 2010). This suggests that further work is needed to ascertain the incidence of autism in patients with HNF1B gene mutations.

Another previously unrecognized feature of HNF1B-associated disease that has been observed is early development of hyperparathyroidism. Inappropriately high levels of parathyroid hormone (PTH), ranging from 6.6 pmol/l to 16.4 pmol/l, compared with the degree of decline in renal function, have been reported in six of eleven unselected patients with known HNF1B

gene anomalies undergoing follow-up at a single centre. Five of these six patients had hypomagnesaemia, which usually inhibits the release of PTH, whereas their plasma levels of calcium and phosphate were within the normal range. *In vitro* studies demonstrated that HNF1B is expressed in PTH-positive cells in the human parathyroid gland and wild-type HNF1B can inhibit transcription of PTH, whereas HNF1B mutants lacked this inhibitory function (Ferre et al., 2013).

Other clinical features that have been reported in a small number of individuals with HNF1B gene mutations include hearing loss and pyloric stenosis (Chen et al., 2010). However, a causal link with HNF1B gene anomalies remains to be established.

1.5 The role of HNF1B in development of endoderm lineage derivatives

1.5.1 Hnf1b function in murine pancreas development

Embryonic development is a complex process characterized by dynamic changes in gene expression. The pancreas, as well other organs such as the liver, part of the duodenum, lungs and thyroid, is derived from the foregut of the primitive gut tube, which originates from the endoderm primary germ layer. Studies on mice have shown that Hnf1b plays an important role in the regulation of endoderm patterning and is expressed broadly throughout the foregut-midgut region at E8, and in the liver and both the ventral and dorsal pancreas buds at E9.5. (Barbacci et al., 1999, Coffinier et al., 1999a, Coffinier et al., 1999b, Haumaitre et al., 2005). The sequential activation of Hnf1b, Hnf6 and Pdx1 directs the differentiation of endodermal cells into pancreatic progenitors (Poll et al., 2006). Hnf1b is a key member of the network of transcription factors (Pdx1, Sox9, Nkx6-1 and Ptf1a) controlling the differentiation of pancreatic multipotent progenitor cells (MPCs) to acinar, ductal and endocrine cells (Solar et al., 2009). Hnf1b expression later becomes confined to the bipotent ductal/endocrine progenitor domain of the central trunk epithelium in the developing organ. In the postnatal or adult pancreas, Hnf1b is restricted to ductal epithelial cells (Solar et al., 2009).

In mice, heterozygous deletion of Hnf1b shows no obvious phenotype, while homozygous mutations are embryonic lethal owing to defective visceral endoderm formation (Barbacci et al., 1999, Coffinier et al., 1999b). Consequently, studies in mice rely on sophisticated transgenic models that either rescue visceral expression during development or that allow conditional

knockout in specific organs and/or developmental stages (Bockenhauer and Jaureguiberry, 2016). To investigate the role of Hnf1b in pancreas development, the early lethality of homozygous Hnf1b-null embryos was rescued through generation of tetraploid chimaeras. It was shown that Hnf1b is required for normal pancreas morphogenesis and regional specification of the gut. Hnf1b-null embryos showed complete absence of the ventral pancreatic bud and an extremely reduced and transient dorsal pancreatic bud, leading to pancreatic agenesis by E13.5. The transient dorsal bud rudiment expressed the transcription factors Pdx1 and Mnx1, but lacked Ptf1a (Haumaitre et al., 2005). The dorsal pancreas was subsequently not maintained because of altered regionalization of the gut through deregulation of Hedgehog signalling. In addition, Hnf1b-null mutants displayed significantly reduced Hnf6 expression and no Neurog3 expression in the remnant dorsal pancreatic bud, resulting in the absence of endocrine progenitor induction (Haumaitre et al., 2005, Maestro et al., 2003, Poll et al., 2006).

In a recent study, constitutive and inducible conditional inactivation of Hnf1b at key developmental stages was performed and the authors proposed a model for Hnf1b function in the regulatory networks governing pancreas development (De Vas et al., 2015). Conditional deletion of Hnf1b in pancreatic MPCs at E12.5 resulted in severe pancreatic hypoplasia, associated with cystic ducts, reduction in number of acinar cells and absence of endocrine cells. Hnf1b was shown to be required for the proliferation and survival of pancreatic MPCs through modulation of FGF and Notch pathways. Deletion of Hnf1b resulted in down-regulation of FGF receptor 4 (Fgfr4), up-regulation of Hey factors, which are inhibitors of the Ptf1a transcriptional complex, and down-regulation of the Notch ligand Dll1, which is activated by Ptf1a.
Hnf1b was found to be a key regulator of both early (E12.5) and late (~E15) duct morphogenesis. Hnf1b deletion resulted in cystic ducts with altered polarity and lack of primary cilia. Hnf1b chromatin immunoprecipitation (ChIP) analyses showed that Hnf1b may control duct differentiation through direct control of the key cystic disease genes GLIS Family Zinc Finger 3 (Glis3), Fibrocystin/Polyductin (Pkhd1), Kif12, Cystin 1 (Cys1), Bicc1 and Hnf6. Early and late Hnf1b deletion also resulted in defects in the acquisition of acinar cell identity and acinar cell maintenance. Hnf1b deletion led to a severe reduction in the number of acinar cells with repression of acinar genes, ectopic expression of ductal genes and replacement of acini through expansion of the ductal compartment. Hnf1b inactivation during the first and secondary transitions resulted in a marked reduction of Neurog3^{+} endocrine progenitors and a novel Hnf1b-binding site was identified in putative regulatory regions of Neurog3. This study therefore showed that in mice Hnf1b is indispensable for the generation of endocrine precursors, likely by directly regulating expression of the pancreatic islet lineage-defining

transcription factor Neurog3. Hnf1b also appears to act directly on Glis3, but in parallel with Sox9 and Hes1 in the control of Neurog3 (De Vas et al., 2015).

The specific function of Hnf1b in pancreatic β-cells was investigated by deletion of Hnf1b in murine β-cells using Cre/Lox conditional gene targeting (Wang et al., 2004). Selective homozygous deletion of Hnf1b in β-cells led to dysregulated islet gene expression, including up-regulation of Pdx1 and Hnf1a and down-regulation of Hnf4a, impaired glucose tolerance and reduced glucose-stimulated insulin secretion. The mice showed normal islet morphology, insulin content, Glut2 and glucokinase (Gck) expression, and preserved normal insulin secretory response to arginine. This phenotype contrasts with Hnf1a-null β-cells, which exhibit impaired glucose and arginine-stimulated insulin secretion because of reduced adenosine triphosphate (ATP) synthesis and defective glucose sensing owing to lower Glut2 and Gck expression. Together, these results suggest that loss of Hnf1b disrupts the regulatory pathways necessary for glucose signalling in β-cells; however, the metabolic pathways affected differ from those disrupted by loss of Hnf1a and are still not fully understood.

1.5.2 Hnf1b function in murine liver development

The mammalian liver is derived from the endoderm of the ventral foregut in close contact with the adjacent mesenchyme of the septum transversum. In the developing liver in mice, Hnf1b is detectable in the hepatic bud from E9.5 and is abundantly expressed in the gallbladder primordium. It is detected later in the IHBDs as they develop within the liver. Expression increases transiently in parenchymal hepatocytes by E17.5-E18.5 and then decreases sharply after birth. In the adult liver, Hnf1b is strongly expressed in the biliary epithelial cells and more weakly expressed in periportal hepatocytes localized at the periphery of the hepatic lobules (Coffinier et al., 1999a).

The role of Hnf1b in early hepatic development in mice has been examined by rescuing Hnf1b-deficient embryos using tetraploid embryo complementation (Lokmane et al., 2008). Hnf1b-null mice exhibited severe liver hypoplasia because of disruption in hepatic specification and failure of formation of the liver bud. The liver architecture was severely disorganized and cells were rounded and lacked hepatoblast characteristics. Mutant livers were devoid of hepatoblasts and were mainly composed of mesenchymal and haematopoietic cells. None of the markers of hepatoblast cells, such as Albumin (Alb), Alfa fetoprotein (Afp), Transthyretin (Ttr), Prospero Homeobox 1 (Prox1), Forkhead Box A2 (Foxa2), Hnf1a, Hnf4a and Hnf6, were expressed at

any stage in the Hnf1b-null mutant liver.

In the same study, the molecular mechanisms underlying the failure of hepatic specification in Hnf1b-null embryos were elucidated by culturing ventral foregut cells in vitro, under conditions reported to be sufficient to induce hepatic specification, in the absence of cardiac mesoderm (Lokmane et al., 2008). These experiments showed that mutant ventral endoderm loses responsiveness to inductive Fgf signals and fails to induce the hepatic-specific genes; however, Hnf1b was found not to be a direct target of Fgf signals. The authors postulated that Hnf1b could mediate competence of the endoderm to respond to Fgf signalling by controlling the expression of a downstream target of this pathway. Analyses in zebrafish confirmed an evolutionary conserved role of Hnf1b in hepatic specification and development and in intrahepatic biliary morphogenesis (Lokmane et al., 2008).

To determine the role of Hnf1b in liver and biliary system organogenesis, tissue-specific inactivation in the liver was performed in a study by Coffinier et al using the Cre/loxP system (Coffinier et al., 2002). The loss of Hnf1b in both hepatocytes and bile ducts resulted in a severe phenotype, including growth retardation, owing to reduced adipose tissue and muscular atrophy, liver hypertrophy and jaundice. Epithelial dysplasia of the gallbladder and the larger IHBDs, a strong decrease in the number of the smaller IHBD and a persistence of the ductal plate from which IHBDs are formed, was also observed in the study. The ductal epithelium presented abnormal characteristics with a delocalization of the nuclei from the base of the cell, the organization of multiple layers of cells or the presence of ectopic epithelium. Hnf1b could therefore be involved in regulation of the fundamental characteristics of epithelium identity such as the basal position of the nuclei and the growth in monolayer of cells. A lack of formation of interlobular arteries was also seen in mutant mice. As Hnf1b is not expressed in endothelial cells, this indicates that artery-bile duct interactions may play an important role in establishing three-dimensional liver structures during organogenesis. Together, these data suggest a requirement for Hnf1b in biliary epithelium formation from the onset of development of the biliary system (Coffinier et al., 2002).

Biochemical analyses showed raised levels of total bilirubin, bile acids and serum cholesterol, which could result from reduced bile flow owing to IHBD paucity and/or defects in hepatocyte metabolism (Coffinier et al., 2002). High levels of triglycerides were found suggesting that hepatocyte lipid metabolism may be affected by Hnf1b liver-specific deletion. Biochemical and molecular studies of hepatic markers led to the identification of two target genes, which were

specifically down-regulated in the absence of Hnf1b, the organic anion transporting polypeptide 1 (Oatp1) transporter, which participates in the reabsorption of bile acids, and very long-chain acyl-CoA dehydrogenase (Vlcad), one of the four key enzymes of fatty acid oxidation (Coffinier et al., 2002).

A recent study in mice also revealed an important role for Hnf1b in the control of hepatic insulin sensitivity and glucose metabolism in vivo (Kornfeld et al., 2013), consistent with the finding of reduced hepatic insulin sensitivity described in humans with HNF1B mutations. Hnf1b was identified as a target of microRNA-802-dependent silencing. Short hairpin RNA-mediated knockdown of Hnf1b in the liver resulted in impaired glucose tolerance, impaired insulin signalling and promotion of hepatic gluconeogenesis, while hepatic overexpression of Hnf1b improved insulin sensitivity in $Lepr^{db/db}$ mice.

Thus, similar to pancreatic development, Hnf1b is necessary both for the induction of liver organogenesis and also for the activity of mature hepatocytes, thereby indicating distinct functions in differentiation and metabolic activity.

1.6 Use of human pluripotent stem cells as an in vitro model to study pancreas development and diabetes

1.6.1 Directed differentiation of human pluripotent stem cells along the pancreatic lineage

Insights gained from developmental biology have made it possible to recapitulate in vitro the key events that regulate early lineage commitment in the embryo, resulting in the efficient and reproducible generation of differentiated cell populations, which can be used for disease modeling, drug screening, and in the future, for cell replacement therapy.

Using an informed approach from developmental biology, landmark studies by D'Amour et al. (D'Amour et al., 2006) and Kroon et al. (Kroon et al., 2008) were the first to succeed in generating pancreatic progenitors and hormone expressing endocrine cells in vitro from human ESCs, using extracellular signalling molecules to direct differentiation choices. The protocols involve step-wise differentiation of human ESCs through the stages of definitive endoderm, primitive gut tube, posterior foregut, and pancreatic endoderm. The cells are exposed to various combinations of growth factors and small molecules in a specific dose and sequence to

differentiate the population of cells as uniformly as possible and the process takes approximately 14-28 days. Several groups have since developed improved multi-step protocols for guided differentiation of human ESCs and induced pluripotent stem cells (iPSCs) towards the pancreatic lineage (Cho et al., 2012, Russ et al., 2015, Nostro et al., 2015, Pagliuca et al., 2014, Rezania et al., 2014, Chen et al., 2009). Most protocols now use serum-free chemically defined conditions.

The expression patterns of various transcription factors are used as markers at each stage of the differentiation process. Most pancreatic differentiation protocols first establish a DE cell population as occurs in vivo. Wnt/beta-catenin and Nodal signals were identified as the main inducers of DE in the vertebrate embryo and during in vitro differentiation from both mouse and human embryonic stem cells (hESCs) and iPSCs. Nodal signalling can be simulated by high doses of other TGFβ family members such as Activin A or GDF8, while transcriptional activation of endodermal target genes (e.g. SOX17) by beta-catenin, can be stimulated by WNT3a or by chemical inhibition of GSK3b using Chir99021. Additional signals such as BMPs, FGFs and inhibition of PI3K/Akt (Ly294002) may also contribute to DE specification. Differentiation of DE to primitive gut tube or foregut progenitors (FP) tries to mimic the appropriate signals during AP patterning and inhibits hepatic lineage commitment. This step involves using FGFs (e.g. FGF7 or FGF10), BMP inhibitors such as Noggin or LDN193189, RA and Vitamin C. Further differentiation to posterior foregut or $PDX1^{+}/NKX6.1^{-}$ and then $PDX1^{+}/NKX6.1^{+}$ pancreatic progenitor cells can be stimulated by use of a SHH inhibitor such as SANT1 or cyclopamine and a Protein Kinase C activator such as Phorbol 12,13-dibutyrate (PdBU) or TPB in addition to FGFs, BMP inhibitor, RA (lower concentration) and vitamin C as in the earlier stage. The combination of epidermal growth factor (EGF) and nicotinamide signalling together with inhibition of BMP pathways also promotes the efficient development of $PDX1^{+}/NKX6\text{-}1^{+}$ progenitors. Introduction of vitamin C at early stages of differentiation, improves cell confluency and survival and prevents premature induction of NEUROG3, resulting in production of $PDX1^{+/}NKX6.1^{+}$ pancreatic progenitors with low expression of NEUROG3. Cell density may also play a role in determining $NKX6.1^{+}$ population size, as hESCs and hiPSCs cultured at high densities during their differentiation into pancreatic progenitors gave larger numbers of $PDX1^{+}/NKX6.1^{+}$ cells (Toyoda et al., 2015).

The next step involves differentiation of pancreatic progenitor cells to endocrine progenitor cells, aiming to up-regulate of NEUROG3 and to produce cell populations of which a substantial fraction co-expresses PDX1, NKX6.1, NEUROD1 and NKX2.2. Gamma-secretase

inhibitor (DAPT) treatment can be used to enhance the expression of NEUROG3 by direct inhibition of NOTCH signalling, directing the cells towards an endocrine fate. The combination of inhibition of TGFβ using TGFβ type I receptor kinase (Alk5 inhibitor) along with a BMP receptor inhibitor and thyroid hormone (T3) efficiently induces NEUROG3 expression (Rezania et al., 2014). EGF (or EGF family member Betacellulin) in combination with cyclopamine/SANT1, DAPT, Alk5 inhibitor, T3 and low concentration of RA is another strategy for induction of NEUROG3^{+} EP cells (Pagliuca et al., 2014).

Continued exposure to ALK5 inhibitor, BMP receptor inhibitor and T3, with addition of a notch inhibitor results in the generation of cell populations in which a substantial percentage of PDX1^{+}/NKX6.1^{+}/NEUROD1^{+} cells express insulin but not glucagon or somatostatin. An inhibitor of AXL (R428), in combination with ALK5 inhibitor and T3, has been found to potently induce MAFA expression in PDX1^{+}/NKX6.1^{+}/NEUROD1^{+} cells resulting in cells that are mono-hormonal, insulin^{+}/glucagon^{-}/somatostatin^{-} and secrete insulin in response to glucose stimulation (Rezania et al., 2014). Alternatively, simpler culture systems have been used where EP cells are allowed to mature in media containing RA and T3 (Pagliuca et al., 2014, Russ et al., 2015). Use of a 3D or cell suspension culture system following pancreatic progenitor specification has been shown to improve the efficiency of β-cell differentiation.

1.6.2 Use of PSCs to study human development and disease

Human induced pluripotent stem cells (hiPSCs) or hESCs offer a unique platform for generating patient-specific disease models and elucidating novel genes and molecular pathways that underlie monogenic and complex traits and diseases, such as diabetes (Teo et al., 2013) (Figure 1.5). hiPSCs can be derived by direct reprogramming of somatic cells using overexpression of transcription factors (Yamanaka Factors – OCT3/4, SOX2, Kruppel Like Factor 4 [KLF4], MYC Proto-Oncogene, BHLH Transcription Factor [cMYC]). The resulting cells can self-renew almost indefinitely in vitro while maintaining their capacity to differentiate into a diversity of cell types including pancreatic cells. Therefore, human iPSCs can allow the production of large quantities of cell types, which are not available through biopsy or primary tissue. Furthermore, human iPSCs allow performance of large-scale studies, including genome-wide analyses, which are challenging to achieve in vivo especially in humans.

Genome editing of hESCs and hiPSCs, using zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) or most recently clustered regularly interspaced

short palindromic repeat-Cas9 (CRISPR-Cas9), permits the introduction or correction of specific disease mutations. The resulting isogenic hiPSC or hESC lines differ only at specific loci, while all other locations remain identical. Using this approach, the effects of a chosen mutation can be dissected from the modifying effects of the genetic background and the precise mechanisms inducing disease can be studied (Sterneckert et al., 2014, Wu and Hochedlinger, 2011, Zhu and Huangfu, 2013).

hiPSC-derived pancreatic cells and hepatocytes have been used as in vitro models to study diabetes and other metabolic diseases (Rashid et al., 2010, Zhu et al., 2016, Shi et al., 2017, Tiyaboonchai et al., 2017, Teo et al., 2016, Yusa et al., 2011, Ding et al., 2013, Tulpule et al., 2013). Several studies in the past 5-10 years have used PSC culture systems and differentiation to pancreatic cells, to study the roles of monogenic diabetes genes (e.g. HNF1B, GATA6, GATA4, PDX1, NEUROG3) in pancreatic progenitor and endocrine cell specification during early human development.

Figure 1.5. Advancing developmental biology and regenerative medicine through studies of hPSCs and model organisms. Genetic studies from the mouse and other model organisms have identified many genes and signalling pathways that govern various aspects of development. Such information has guided the search for defined conditions to turn hPSCs into specific cell types of all three germ layers. With the development of new genetic tools, it is now possible to use hPSCs as a new model system for studies of human development. The generation of desired cell types from hPSCs will also advance regenerative medicine, including cell replacement therapy, disease modeling and drug discovery (Zhu and Huangfu, 2013).

1.6.3 Use of hPSCs to study monogenic diabetes

1.6.3.1 HNF1B

The mechanisms that underlie development of diabetes in patients with HNF1B mutations are still not fully understood because of the lack of an appropriate model system. Hnf1b heterozygous knockout mice show no obvious phenotype in contrast with the phenotype seen in humans (Servitja and Ferrer, 2004). Biallelic mutations in Hnf1b are embryonic lethal in mice because of defective visceral endoderm formation (Barbacci et al., 1999, Coffinier et al., 1999b). This implies that mouse models do not reproduce the phenotypic variability seen in the human disease and therefore cannot be used to determine genetic factors that modify the phenotypic outcome.

A recent report described the derivation and characterization of human iPSC lines from patients with HNF1B-associated diabetes owing to an S148L/+ mutation (Teo et al., 2016). This study showed that a decrease in HNF1B expression did not affect endoderm differentiation into foregut or the subsequent induction of pancreatic progenitors. HNF1B haploinsufficiency appeared to increase the expression of endoderm makers such as SOX17 and pancreatic bud markers such as PDX1. These observations appear to contradict clinical data suggesting that HNF1B haploinsufficiency during human development can result in pancreatic hypoplasia, and also genetic studies in the mouse showing the importance of Hnf1b in the patterning of the foregut. These apparent differences could be explained by divergence of in vitro culture conditions with normal development in vivo. However, another explanation could be that pancreatic hypoplasia due to HNF1B haploinsufficiency is cell non-autonomous and thus the phenotype is different in cells generated in isolation compared with cells generated in a more complex in vivo environment.

Importantly however, this study revealed a decrease in PAX6 expression during the differentiation of HNF1B mutant iPSCs into endocrine cells. This observation could be of particular relevance as PAX6 is necessary for differentiation of endocrine progenitors into hormonal cells. Recently, decrease in PAX6 expression has also been associated with the dedifferentiation process affecting β-cells in T2D (Ahmad et al., 2015). The absence of PAX6 during differentiation of HNF1B mutant iPSCs is therefore likely to result in major defects in hormonal cell production or function. Intriguingly, HNF1B is not known to bind to the PAX6 promoter or to directly control its expression, indicating that the decrease in PAX6 gene

expression is an indirect effect of the mutant HNF1B$^{S148L/+}$.

It must be noted that the function of HNF1B in β-like cells could not be comprehensively addressed in this study as the protocol utilized does not allow the production of glucose responsive β-like cells. An additional limitation of the study is that although the study used iPSC lines from a single family, including members with HNF1B mutations and diabetes, HNF1B mutations with no diabetes and normal family controls, isogenic cell lines were not used, therefore genetic background could account for the differences in gene expression seen. Thus, additional studies are required to further uncover the molecular mechanisms by which HNF1B could control β-cell production and function in humans through the regulation of PAX6.

1.6.3.2 HNF4A

A recent study by Vethe et al. (Vethe et al., 2017) looked at the effect of HNF4A mutations on the production of insulin-expressing cells in vitro. iPSC lines were derived from four family members, including three HNF4A mutation carriers (two with diabetes and one non-diabetic mutation carrier) and their family control (non-diabetic non-mutation carrier). The synchronous differentiation of the four iPSC lines towards insulin-producing cells was performed using the seven-step differentiation protocol established by Rezania et al. (Rezania et al., 2014) and resulted in the successful generation of insulin-expressing cells in all four samples regardless of the mutation or diabetes status. The final stage cell population also expressed other key markers of pancreatic β-cells, including NKX6-1 and MAFA, as detected by immunofluorescence. Furthermore, the global proteomic comparison of the final stage cells derived from HNF4A mutation carriers and their family control detected similar levels of insulin and most key β-cell specific markers, regardless of the mutation or diabetes status. Overall, these data suggest that the HNF4A mutation does not affect the development of insulin-producing β-cells in vitro.

1.6.3.3 HNF1A

A study by Stepniewski et al. (Stepniewski et al., 2015) assessed the ability of iPSC lines from patients with HNF1A mutations to differentiate toward insulin producing cells using two different protocols. No differences were observed between the HNF1A mutant and control cell lines however the authors commented that the efficiency of the differentiations was not optimal

and produced immature polyhormonal cells which secreted glucagon but not insulin, precluding reliable quantitative comparisons and mechanistic study of molecular pathways.

1.6.3.4 GATA6

Two recently published studies by Shi et al. (Shi et al., 2017) and Tiyaboonchai et al. (Tiyaboonchai et al., 2017) used hPSCs to investigate the effect of GATA6 mutations on endoderm and pancreas specification and β-like cell functionality. In the study by Shi et al., (Shi et al., 2017) CRISPR/Cas9 engineered hESCs and GATA6 patient-derived iPSCs, along with isogenic controls were differentiated along the pancreatic lineage. Directed differentiation assays revealed a requirement for GATA6 in the formation of DE and a dosage-sensitive requirement for GATA6 in the formation of $PDX1^{+}/NKX6.1^{+}$ pancreatic progenitors and, subsequently, glucose-responsive β-cells. No difference in glucose-stimulated insulin secretion was observed between GATA6 heterozyous mutant and wild-type cells in vitro or in vivo. Dosage-dependent effects of GATA4 alleles on phenotypes associated with GATA6 heterozygosity were observed during in vitro differentiation at the pancreatic progenitor cell stage.

In addition, Tiyaboonchai et al. (Tiyaboonchai et al., 2017) used CRISPR/Cas9 to insert a second mutation in a patient-derived hiPSC line with a heterozygous mutation in GATA6 and to correct the mutation to create an isogenic wild-type control hiPSC line. GATA6 wild-type and isogenic genetically engineered hESCs with homozygous knockout of GATA6 were also used for these experiments. hPSC lines with homozygous mutations failed to differentiate to definitive endoderm cells however re-expression of GATA6 or other GATA family members restored this defect. Endodermal progenitor (EP) cells were therefore used as a tool to bypass the endoderm defect and pancreatic β-cell differentiation was examined. Heterozygous and homozygous mutants unexpectedly produced $PDX1^{+}$ pancreatic progenitor cells and $PDX1^{+}/CPEP^{+}$ β-like cells although at reduced efficiency compared with isogenic wild-type control. However, simply reducing the concentration of RA 80-fold led to statistically significantly fewer PDX1+ cells when compared to a wild-type iPSC line that showed negligible sensitivity to the same culture regime. Reduced glucose-stimulated insulin secretion was seen in mutant cells as well as an increase in proinsulin compared with insulin, suggesting that GATA6 may be involved in insulin processing in EP-derived β-like cells.

1.6.3.5 Other monogenic diabetes genes

Zhu et al. (Zhu et al., 2016) combined TALEN and CRISPR/Cas9-mediated genome editing and human PSC directed differentiation to systematically examine eight transcription factors (PDX1, Regulatory Factor X6 [RFX6], PTF1A, GLIS3, MNX1, NEUROG3, HES1 and Aristaless Related Homeobox [ARX]) for their roles in pancreatic development. Homozygous loss of function mutations in PDX1, RFX6, PTF1A, GLIS3, MNX1, NEUROG3 are known to cause PNDM. ARX and HES1 were also chosen because in contrast to the pro-endocrine roles of PNDM-associated genes above, Hes1 inhibits endocrine development whereas Arx is necessary for the formation of α-cells, but not β-cells, in mice. Novel or divergent roles between mice and humans were observed for RFX6, PDX1 and NEUROG3. PDX1 haploinsufficincy resulted in a 65% decrease in INS^+ cells. In contrast, no developmental abnormalities are known to occur in $Pdx1^{+/-}$ mice (Jonsson et al., 1994, Offield et al., 1996) and in humans PDX1 haploinsufficiency was previously thought to cause defects in β-cell function and/or the maintenance of β-cell mass (Brissova et al., 2005, Brissova et al., 2002, Johnson et al., 2003). The study also showed a reduction in $PDX1^+$ cells in RFX6 homozygous knockout mutants. This was unexpected as RFX6 is known to be important in the later stages of differentiation of pancreatic progenitors to endocrine cells (Smith et al., 2010, Soyer et al., 2010) and for β-cell function (Chandra et al., 2014, Piccand et al., 2014).

Most interestingly, NEUROG3 homozygous mutant hESCs exhibited a distinct phenotype from the mouse knockouts. Loss of NEUROG3 resulted in impairment of endocrine cell formation however small numbers of β-cell were still able to form, unlike in mouse knockouts (Gradwohl et al., 2000). These cells produced low amounts of insulin but were not responsive to glucose stimulation. This is consistent with studies in humans which showed that patients with NEUROG3 homozygous mutations can have low but detectable blood C-peptide levels indicative of the presence of residual β-cells (Jensen et al., 2007). To explore the competence window responsive to NEUROG3 expression, gain- and loss-of-function approaches were combined. NEUROG3 homozygous mutant cells were differentiated to the early pancreatic progenitor stage and then subjected to doxycycline treatment at four differentiation stages, each of a 3-day interval. Expression of the NEUROG3 transgene in the NEUROG3 knockout background induces endocrine differentiation in all stages, with the strongest effects corresponding to the time of formation of $PDX1^+/NKX6.1^+$ pancreatic progenitor cells.

Considered together, these results illustrate the interest of using human iPSCs to study developmental mechanisms controlled by disease-associated genes in humans. However, important challenges remain including the advancement of differentiation protocols, which fully recapitulate natural development while remaining compatible with large-scale genome-wide analyses. The recent publications of protocols, using 3D culture systems and allowing large-scale production of β-like cells displaying a functional response to glucose (Pagliuca et al., 2014, Rezania et al., 2014, Russ et al., 2015) has provided further opportunities to investigate the mechanisms controlling functional β-cell development and insulin secretion.

1.6.4 Beta cell replacement therapy for diabetes

Treatment of T1D requires continuous administration of exogenous insulin, whereas T2D can often be controlled by other oral or injected therapeutics that act on the β-cells or peripheral tissues until later stages when patients also require exogenous insulin. However, none of these therapies matches the precision of endogenous production of insulin from pancreatic β-cells. Patients on insulin require multiple daily blood glucose tests and insulin doses for their entire life. Improving glycemic control can prevent complications of diabetes including retinopathy, nephropathy, neuropathy, and enhanced risk of cardiovascular disease and results in improved patient health.

Current cell replacement therapies consist of whole pancreas or islet transplantation. However, these approaches require life-time immunosuppression, and are limited by the paucity of available donors. Alternative therapeutic options are crucial in order to address these healthcare challenges. hPSCs (hESCs or hiPSCs) hold great potential as the basis for cell-based therapies of diabetes. hPSC-derived β-like cells have been used as a cell therapy in diabetic animal models, such as the Akita diabetic mouse model or mice with streptozotocin-induced diabetes. Transplantation of hPSC-derived pancreatic progenitor cells under the kidney capsule or fat pad in immunocompromised mice, results in differentiation of these cells into acinar, ductal and islet cells. Mature β-cells capable of GSIS and reversing streptozotocin-induced diabetes arise after 3-4 months (Kroon et al., 2008). hPSC-derived β-like cells have also been transplanted in diabetes mouse models and reverse diabetes after 2-4 weeks ((Pagliuca et al., 2014, Rezania et al., 2014).

Although there have been significant improvements, robust protocols allowing for the production of homogenous populations of β-like cells from hPSCs under Good Manufacturing

Practice (GMP) culture conditions compatible with clinical applications have not yet been established. It has proven to be extremely challenging to make authentic, glucose-responsive, insulin-secreting beta cells *in vitro* and further advances in our understanding of pancreatic development, particularly signalling pathways involved in endocrine cell specification and maturation are required.

1.7 Aims and objectives

Heterozygous mutations in the transcription factor, HNF1B, result in multisystem disease including diabetes due to beta-cell dysfunction and pancreatic hypoplasia. However, the mechanisms that underlie development of diabetes in patients with HNF1B mutations are still not fully understood due to the lack of an appropriate model system. The ability to differentiate human iPSCs along the pancreatic lineage represents a unique tool to explore the key roles of specific monogenic diabetes genes (e.g HNF1B) in human pancreas development and function (Hua et al., 2013, Teo et al., 2013).

The overall aim of the project is to determine the molecular mechanisms by which HNF1B mutations cause pancreatic hypoplasia and diabetes. This will be achieved by the following objectives:

1. To establish an in vitro model of HNF1B-associated diabetes by;
a) Derivation and characterisation of iPSCs from patients with HNF1B-associated disease.
b) Generation of heterozygous and homozygous HNF1B knockout hPSC lines.
c) Pancreatic differentiation and functional characterisation of HNF1B mutant cell lines and isogenic control PSC lines.

2. To determine the molecular mechanisms by which HNF1B mutations affect pancreatic development by;
a) Examining the transcriptional network controlled by HNF1B using RNA-seq and ChIP-seq.
b) Exploring the epigenetic function of HNF1B using ATAC-seq.

3. To elucidate the molecular mechanisms by which HNF1B mutations affect the function of adult-like pancreatic β-cells by generation of β-like cells from HNF1B wild-type, heterozygous and homozygous knockout hiPSCs in vivo.

Chapter 2 Methods

2.1 Research subjects and derivation of hiPSCs

Two hiPSC lines were used for most of the experiments in this project.; FSPS13.B (HPSI0813i-fpdm_2) and Eipl_1 (HPSI0114i-eipl_1). These hiPSC lines were derived in-house at the Wellcome Trust Sanger Institute, from human skin fibroblasts (Eipl_1) and peripheral whole blood (FSPS13.B), using Sendai virus reprogramming. Skin biopsy or whole blood samples were collected from consented normal volunteers recruited through the National Institute for Health Research (NIHR) Cambridge BioResource (www.cambridgebioresource.org.uk). Ethics approval for hiPSC derivation was obtained from the National Research Ethics Service (NRES) Committee East of England, Cambridge East (ref: 09/H0304/77).

The generation of patient-derived hiPSCs was approved by the Great Ormond Street Hospital and Institute of Child Health Research Ethics Committee (ref: 08/H0713/82) and the Cambridge Research Ethics Committee (ref: 06/Q0108/373). Skin biopsy samples were collected from a cohort of patients with monogenic diabetes, after obtaining informed consent (Table 2.1.). Fibroblasts were derived from the donated skin biopsy samples by the Cambridge University Hospitals NHS Genetics Laboratory. These fibroblasts were reprogrammed to hiPSCs by the Cellular Genetics and Phenotyping (CGaP) facility, Wellcome Trust Sanger Institute and Cambridge Biomedical Research Campus (BRC) hiPSC core facility, Cambridge, using Sendai virus reprogramming. hiPSC lines were successfully derived from 8 patients with HNF1B point mutations or whole gene deletions. Three independent hiPSC clones were isolated for each patient line.

All normal (wild-type) and patient-derived hiPSC lines used were validated by the Cambridge Biomedical Research Centre iPSC Core Facility or by the CGaP facility, Wellcome Trust Sanger Institute, for expression of endogenous pluripotency markers versus transgenes by qPCR and by in vitro differentiation into the three germ layers.

All hiPSC lines were regularly tested for mycoplasma contamination and karyotype analysis was performed using M-FISH and DAPI banding, following derivation and after every 20 passages, by the Molecular Cytogenetics Core Facility at the Wellcome Trust Sanger Institute.

Gene	Inheritance	Condition	Number of patient samples
HNF1A	AD	MODY	8
HNF4A	AD	MODY	3
HNF1B	AD	MODY	11
ABCC8	AR	Permanent and transient ND	3
KCNJ11	AR	Permanent and transient ND	12
GATA6	AD	Permanent ND and pancreatic agenesis	4
GCK	AR	Permanent ND	2
INS	AR	Permanent ND	7
GLIS3	AR	Permanent ND	1
SLC19A2	AR	Thiamine responsive megalobalstic anaemia (TRMA) syndrome	1
FOXP3	AR	Immunodysregulation polyendocrinopathy enteropathy X-linked (IPEX) syndrome	1
STAT3	AD	Autoimmune ND	1
MT-TL1		Maternally inherited diabetes and deafness (MIDD)	3
Unknown		Permanent ND +/- pancreatic agenesis	3

Table 2.1. Skin biopsy samples obtained from patients with monogenic forms of diabetes AD = Autosomal Dominant, AR = Autosomal Recessive, MODY = Maturity Onset Diabetes of the Young, ND = Neonatal Diabetes

2.2 Gene targeting to generate HNF1B knockout lines

2.2.1 Construction of gRNA vectors for Cas9-mediated genome engineering

Cas9 nuclease target regions in exon 1 of the HNF1B gene were identified using the CRISPR design tool provided by the Zhang laboratory (http://crispr.mit.edu) (Ran et al., 2013). The tool identifies suitable target sites in a genomic sequence of interest. gRNAs are scored and ranked based on their specificity and off-target genomic sites for each gRNA are provided. Two gRNA sequences with the highest scores were selected to test their efficiencies in iPSCs. To construct the gRNA vector, gRNA sequences were cloned into a U6 BsaI gRNA backbone vector. Two oligonucleotides that anneal together to create a double stranded DNA sequence with compatible ends to those generated by the BsaI digest were designed. The U6 promoter requires a "g" at the start of the sequence to be transcribed, therefore this was added before the 5' end of the gRNA (if the guide RNA does not start with a "g"). The PAM site was not included in the gRNA sequence. The gRNA oligonucleotides were ordered from Sigma-Aldrich (Haverhill, UK). The Cas9 nuclease vector was purchased from Addgene (hCas9 Plasmid #41815, Addgene, Cambridge, MA, USA) and the U6 BsaI gRNA backbone vector was obtained from Professor Bill Skarnes' group at the Wellcome Trust Sanger Institute, Hinxton, Cambridge. Plasmid maps are shown in Figure 2.1.

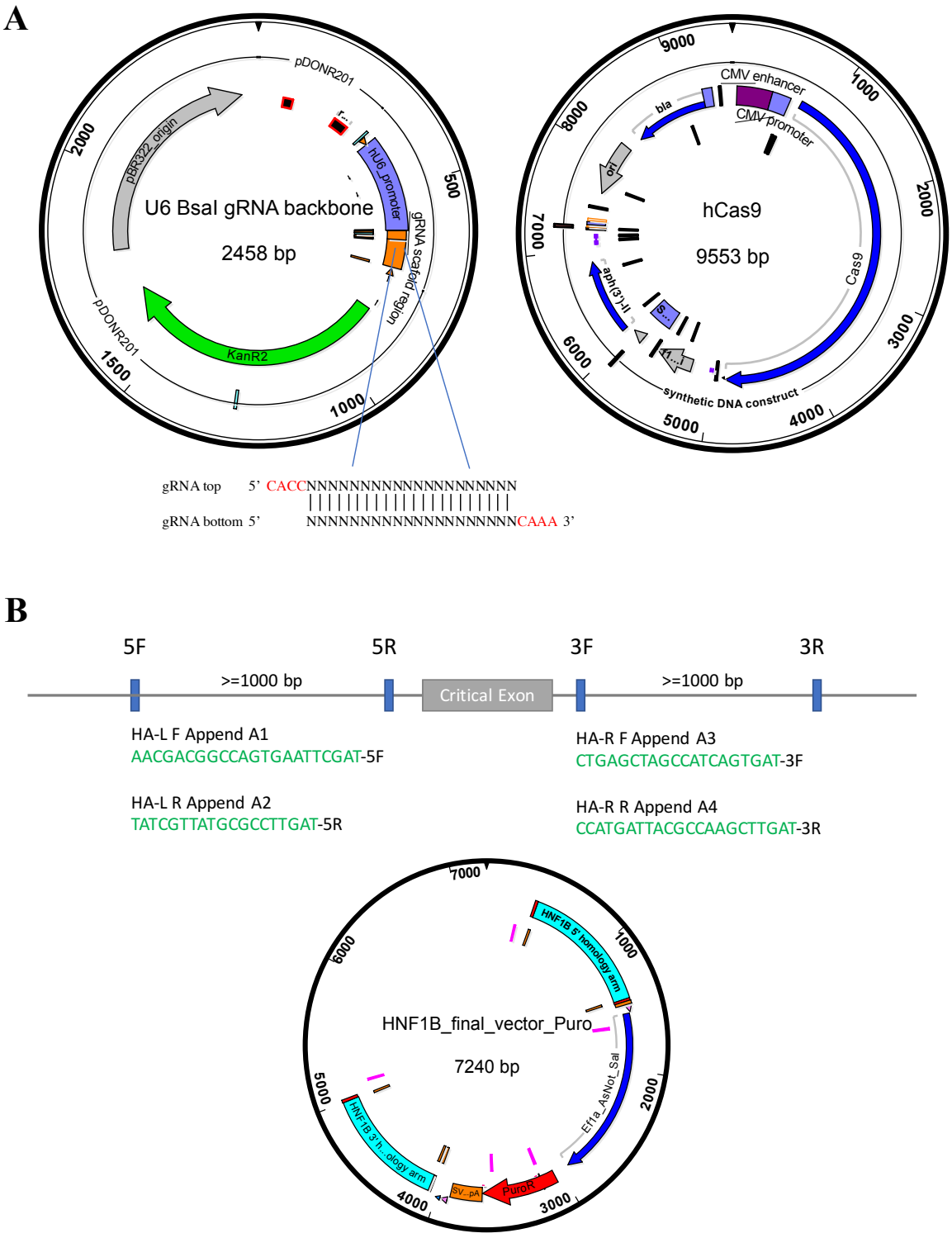


Figure 2.1. Targeting strategy for genetic engineering of HNF1B heterozygous knockout and HNF1B homozygous knockout hiPSCs. **(A)** Cas9 and gRNA vectors showing cloning site for guide RNA, **(B)** Donor vector containing repair template.

The U6 BsaI gRNA backbone vector was digested using the BsaI restriction enzyme (NEB, MA, USA), which cuts outside of its recognition sequence leaving non-compatible sticky ends to allow the ligation of the gRNA sequence. 0.1µg of the vector (1µl), was incubated with 40 Units of BsaI High Fidelity enzyme (2µl), 1x CutSmart Buffer (5µl) (NEB) and 42µl nuclease free water (to make up to a total volume of 50µl), overnight (16 hours) at 37°C, followed by heat inactivation at 65°C for 20 minutes. Following agarose gel electrophoresis, gel extraction was used to isolate a desired fragment of DNA (correct size of band = 2458 - 25 bp) (QIAquick Gel Extraction Kit, Qiagen, Hilden, Germany).

Annealing and phosphorylation of oligonucleotides was performed using the following reaction mixture: gRNA top (1µl, 10µM final concentration), gRNA bottom (1µl, 10µM final concentration), T4 ligase buffer (10x, 1µl) (NEB), T4 Polynucleotide Kinase (1µl) (NEB) and nuclease free water (6µl) to make up final volume to 10µl. The reaction mixture was incubated in a thermocycler at 37°C for 30 minutes, then 95°C for 5 minutes then temperature was ramped down to 25°C at 5°C per minute.

For ligation of the annealed oligonucleotides into the U6 BsaI gRNA backbone, the following reaction mixture was set up: linearised vector (20ng/µl, 1µl), diluted ds-oligos (7.6fmol/µl, 2µl) or H2O negative control (2µl), T4 DNA ligation buffer (10x, 1µl) (NEB), T4 ligase (1µl) (NEB), nuclease free water (5µl) to make up final volume to 10µl. The reaction mixture was incubated for at least 6 hours or overnight at 16°C. 5µl of the ligation reaction was transformed into 50µl of DH5-alpha competent bacteria (NEB) and 100µl of the culture was plated out onto LB agar Kanomycin plates. Colonies were picked and minipreps were grown overnight in LB containing Kanomycin. Bacterial glycerol stocks, for long-term storage of plasmids, were prepared. Plasmid DNA purified from miniprep cultures was sent for sequencing to confirm the presence/sequence of the gRNA cassette. Transfection quality DNA was purified using an endotoxin free maxiprep kit (EndoFree Plasmid Maxi Kit, Qiagen, Hilden, Germany).

2.2.2 Construction of intermediate homology arm donor vector – Gibson Assembly

Homology arms (HA) for the gene of interest (HNF1B) were designed to be around 1Kb in length and flank the critical exon (exon 1) and DNA cut site previously determined by the design of gRNA sequences. The homology arms were flanked by the appropriate Gibson Append sequence. Homology arms were obtained by PCR from genomic DNA using Q5 Hi

Fidelity enzyme (NEB, Ipswich, MA, USA) followed by gel extraction of the required band using the QIAquick Gel Extraction Kit (Qiagen).

The PCR/DNA synthesised homology arm fragments were combined with a Zeo/PheS selection cassette (positive and negative selection) and EcoRV linearised pUC19_RV vector to create an intermediate construct using the Gibson Assembly method (NEB). pUC19_RV vectors and Zeo/PheS selection cassette (excised from pR6K-R1R2-ZP vector) used for these experiments were obtained from Professor Bill Skarnes' group at the Wellcome Trust Sanger Institute, Hinxton, Cambridge. Plasmid maps are shown in Figure 2.1. For 4-6 fragment assembly, 100ng of vector was incubated with 2 -3 times excess of insert. The total amount of DNA should be between 0.2-1pmol with a final reaction volume of 20ul. The following reaction mixture was set up and incubated at 50°C for 1 hour: pUC19_RV (15ng/µl), Zeo/PheS (70ng/µl), 5' Homology arm (HA-L) (40ng/µl), 3'Homology arm (HA-R) (40ng/µl), Gibson Mix (2x, 10µl) and nuclease free water to make up final reaction volume to 20µl. 2µl of the reaction was transformed into NEB dam-/dcm- competent E.coli (C2925H) and 100µl was plated out onto LB agar plate containing Zeocin (50µg/ml) and Amplicillin (100µg/ml). Colonies were picked and minipreps were grown overnight in LB containing Zeocin (50µg/ml) and Amplicillin (100µg/ml). Plasmid DNA was isolated using QIAprep Spin Miniprep kit. Purified plasmid DNA was checked for the presence of all fragments by PCR and positive colonies were then sent for sequencing.

2.2.3 Construction of final homology arm donor vector

Ef1a promotor-driven puromycin resistance gene (Ef1a-Puro) cassette was obtained from Professor Bill Skarnes' group at the Wellcome Trust Sanger Institute, Hinxton, Cambridge. Replacement of the Zeo/Phes selection cassette between the homology arms with the Ef1a-Puro cassette was performed using the Gateway Recombination LR reaction (Invitrogen, Carlsbad, CA, USA). The following reaction mixture was set up and incubated overnight at 25°C: Ef1a-Puro cassette (60ng/µl), Intermediate vector (100ng/µl), LR clonase II (2µl) and Nuclease free water to make up final reaction volume to 10µl. Proteinase K (1µl) was then added and the reaction mixture was incubated at 37°C for 30 minutes. 1µl of the reaction was transformed into alpha select competent cells (OneShot TOP10 cells, Invitrogen) and 100µl of the bacteria was plated onto YEG-Cl Amp plates and grown overnight. The only bacteria that can grow on these plates are those containing the pUC19RV backbone (Amp^r) without the presence of the

ZeoPhes cassette (as the D,L-p-Cl-Phe acts as negative selection against the PheS gene). Any bacteria transformed with the non-recombined intermediate vector cannot grow due to the ZeoPhes cassette and any transformed with the non-recombined pL1L2 vector cannot grow due to the lack of Amp^r. Plasmid DNA was isolated using QIAprep Spin Miniprep kit (Qiagen).

Purified plasmid DNA was initially checked for the presence of all fragments by PCR and positive colonies were sent for sequencing (Beckman Coulter Genomics, Sanger sequencing facility). A larger amount of DNA, suitable for transfection, was purified using an endotoxin free maxiprep kit (EndoFree Plasmid Maxi Kit, Qiagen).

2.2.4 Nucleofection of hESCs

iPSCs were transfected using the Amaxa Nucleofector® Technology and Human Stem Cell Nucleofector® Kit 1 (Lonza, Basel, Switzerland). A total of 8µg of DNA; final targeting vector(s), gRNA vector, Cas9 nuclease (Addgene, Cambridge, MA, USA) in equivalent ratios; was transfected into 1-2x10^6 cells prepared into a single cells suspension. 8µg DNA was added to supplemented Nucleofector solution. Cells were pelleted and resuspended in 120µl of DNA and supplemented Nucleofector® solution (82µl Nucleofector Solution 1 + 18µl Supplement 1). The cell/DNA suspension was transferred into a cuvette and inserted into the Nucleofector® machine. Nucleofector® Program B-016 was applied. Following nucleofection, 500µl E8 media supplemented with Rock inhibitor was added to the cuvette and the sample was transferred to a vitronectin-coated 10mm plate, filled with a desired volume of culture media. Media was replaced with fresh E8 media the next day.

2.2.5 Selection and screening of targeted hESC colonies

48 hours after nucleofection, selection was commenced using 1 µg/ml puromycin added to E8 media for 5 days. Surviving colonies were allowed to grow in E8 media for a further 5-7 days. Surviving colonies were picked, once they had reached an appropriate size. Half of each colony was placed in PCR tubes for DNA extraction using the GenElute Mammalian Genomic DNA Miniprep Kit (Sigma-Aldrich, St. Louis, MO, USA) and the other half was transferred to one well of a 24-well plate for further culture.

Transfected PSC colonies were screened by PCR to detect heterozygous and homozygous HNF1B knockout clones. The targeted segment of DNA was amplified using primer pairs, with

one primer contained within the selection cassette and the other outside the homology arm (5' end and 3' end) of the donor vector. Sanger sequencing was also performed to check if integration of the puromycin cassette occurred in one or both alleles of the HNF1B gene. The HNF1B targeted locus was amplified using primer pairs around the targeted locus.

2.2.6 Correction of mutations in patient iPSC lines

Two patient iPSC lines (R2 and R4) were selected, all with mutations in exon 4 of the HNF1B gene. Patient iPSC line R2 had a 1bp duplication mutation, g.36091625dup resulting in a frameshift (p.His336fs) and patient R4 had nonsense mutation, g.36091805G>A (p.Arg276*).

Cas9 nuclease target regions in exon 4 of the HNF1B gene, as close as possible to the patient mutations, were identified using the CRISPR design tool. Oligonucleotide pairs encoding the 20-nt gRNA sequences were annealed and ligated into; (1) the pSpCas9(BB)-2A-Puro plasmid containing Cas9 nuclease, a cloning backbone for sgRNA and 2A-Puro (Addgene PX459, Plasmid #62988) and (2) into a U6 BsaI gRNA backbone vector, to be used in conjunction with a Cas9 nuclease protein.

A single stranded oligodeoxynucleotides (ssODN) repair template was used in place of a targeting plasmid to introduce short modifications within the patient mutation loci. The ssODN (ordered from Integrated DNA Technologies (IDT), Coralville, IA, USA) was orientated in the sense/antisense direction, contained flanking sequences of 90 bp on each side of the target locus and a synonymous mutation was introduced in the PAM site to prevent the Cas9 from cutting the ssODN donor template.

Method 1: The pSpCas9(BB)-2A-Puro vector (4ug) and a ssODN (100pmol) were co-nucleofected into the patient iPSC lines, in single cells suspension ($2x10^6$ cells). After 24 hours, the cells were incubated with puromycin for 72 hours. The cells were then cultured in E8 medium without antibiotics. Surviving colonies were picked, once they had reached an appropriate size. Half of each colony was placed in PCR tubes for DNA extraction using the GenElute Mammalian Genomic DNA Miniprep Kit (Sigma-Aldrich) and the other half was transferred to one well of a 24-well plate. Targeted genome modifications were detected by Sanger sequencing. Amplicons should be subcloned into a plasmid such as pUC19 for transformation, and individual colonies should be sequenced to reveal the clonal genotype.

Method 2: The Cas9-recombinant protein (Toolgen) (20g), a chemically synthesised tracrRNA (Dharmacon, CO, USA), a chemically synthesised crRNA, composed of the 20 nucleotide site-specific RNA and 22 nucleotide S.pyogenes guide sequence that interacts with the tracrRNA (Dharmacon), and a ssODN (500pmol) were co-nucleofected into the patient iPSC lines, in single cells suspension ($8x10^5$ cells). The cells were then plated in a 10cm dish at different densities (8,000, 16,000 and 40,000 cells) and cultured in E8 medium. 96 colonies were picked, once they had reached an appropriate size. Half of each colony was placed in PCR tubes for DNA extraction using the GenElute Mammalian Genomic DNA Miniprep Kit (Sigma-Aldrich) and the other half was transferred to one well of a 24-well plate. HDR was detected via PCR amplification, followed by Sanger sequencing of the modified region.

2.2.7 Off target effects

The clones selected for use in further experiments were analyzed for off-target effects of RNA-guided Cas9 nuclease. Off-target sites for the HNF1B gRNA used in this project were identified using the CRISPR design tool (http://crispr.mit.edu). Nine potential off-target regions, including three in exonic regions in protein coding genes were identified: 1) sema domain, transmembrane domain (TM), and cytoplasmic domain, (semaphorin) 6D (SEMA6D); 2) Rho GTPase activating protein 23 (ARHGAP23) and 3) Fragile Histidine Triad (FHIT). Sanger sequencing was performed for the regions in the 3 protein coding genes, and no mutations (single nucleotide variants, insertions and deletions) within 500 bp of a predicted off-target site in these genes were found.

2.3 Differentiation of hiPSCs to pancreatic cells

Once derived, hiPSCs were cultured in standard hiPSC culture conditions using irradiated mouse feeder plates and Knockout Serum Replacement (KSR) medium supplemented with fibroblast growth factor 2 (FGF2). Cells were then transferred to a feeder-free culture system; using Essential 8 (E8) medium (Invitrogen) and Vitronectin XF™ (Stem Cell Technologies) coated plates. Human iPSCs were passaged routinely every 4-7 days. Details of all media recipes and preparation of cytokine stock solutions and their manufacturers can be found in Appendices 1 and 2.

For pancreatic differentiations, hiPSCs were passaged and seeded onto 12-well plates using E8 supplemented with Y27632 (Rho-associated, coiled-coil containing protein kinase, ROCK

Inhibitor) (10μM). Cells were plated as single cells and the plating density was optimised for each hiPSC line (approximately 80,000 to 100,000 cells per well of a 12-well plate). After 24 hours, the media was replaced with fresh E8 media without ROCK Inhibitor. To induce definitive-endoderm differentiation (first stage; days 1-3), cells were cultured in CDM-PVA supplemented with Activin (100ng/ml), FGF2 (80ng/ml), BMP4 (10ng/ml), Ly294002 (10μM) and CHIR99021 (3μM) on day 1, CDM-PVA supplemented with Activin (100ng/ml), FGF2 (80ng/ml), BMP4 (10ng/ml) and Ly294002 (10μM) on day 2 then RPMI/B27 media containing Activin (100ng/ml), FGF2 (80ng/ml) on day 3.

For the second stage of the protocol (primitive gut tube differentiation), cells were cultured in Adv-BSA media supplemented with SB-431542 (10μM), FGF10 (50ng/ml), RA (3μM), Noggin (150μg/ml) and L-Ascorbic acid (250μM) for 3 days (days 4-6). For the third stage (posterior foregut; days 7-8), cells were cultured in Adv-BSA with FGF10 (50ng/ml), RA (3μM), Noggin (150μg/ml), KAAD-cyclopamine (0.238μM), PdBU (50nM) and L-Ascorbic acid (250μM). For the fourth stage of the protocol (pancreatic progenitor specification; days 9-13), the cells were cultured in RA (1μM), Noggin (150μg/ml), KAAD-cyclopamine (200ng/ml), EGF (100ng/ml), Nicotinamide (10mM), and L-Ascorbic acid (250μM) for 5 days.

Cells were then grown in Adv-BSA containing glucose (final concentration 25mM), B27 (1%), RA (100nM), DAPT (1μM), Alk5i (10μM) and the small molecule BNZ (0.1mM) for 3 days to induce maturation of pancreatic progenitors to endocrine progenitor cells (fifth stage; days 14-16). For maturation of endocrine cells and further differentiation into C-peptide-producing beta cells, cells were cultured for 3 additional days in Adv-BSA containing B27 (1%), RA (100nM) and Alk5i (10μM) followed by 11 days in Adv-BSA containing B27 (1%), RA (100nM) (sixth stage; days 17-27). See Appendices 1 and 2 for media recipes and cytokine information.

2.4 RNA extraction and quantitative polymerase chain reaction

For quantitative analysis of gene expression at the transcript level, total RNAs were extracted from hESCs or differentiated cells using the GenElute Mammalian Total RNA Miniprep Kit (Sigma-Aldrich, St. Louis, MO, USA). For each sample, 500ng of total RNA was reverse transcribed into first strand cDNA using Superscript II Reverse Transcriptase (Invitrogen,

Carlsbad, CA, USA) and then diluted 1:30 in nuclease-free water. Quantitative real-time PCR was performed using Sensi Mix Sybr Low Rox Kit (Bioline QT625-20, London, UK). The enzyme reaction mixture was prepared as follows: F primer (final concentration, 0.3μM), R primer (final concentration, 0.3μM), Sensi Mix (2x), nuclease free water (Invitrogen), to make up final reaction volume of 10μl. The enzyme reaction mixture was added to 5μl of sample cDNA. The reaction was run on a Stratagene Mx3005P (Agilent Technologies, Santa Clara, CA, USA) using the qPCR program: 10 minutes at 95°C (step 1), 30 seconds at 95°C (step 2), 30 seconds at 60°C (step 3), 30 seconds at 72°C (step 4), with steps 2 – 4 repeated 40 times. Reactions were performed in triplicate, with technical duplicates, and expression of the gene of interest was normalized to that of an internal housekeeping gene; porphobilinogen deaminase (PBGD) or ribosomal phosphoprotein P0 (RPLP0 / 36B4) in the same run. A negative control with no cDNA (no template control, NTC) was run for each primer pair used.

Primers were designed in house using the NCBI Primer-BLAST tool and synthesized by Sigma-Aldrich. Sequences for all primers used for RT-PCR analyses are listed in Appendix 3. All new pairs of QPCR primers were tested to ensure that they are specific and that they amplify at approximately 100% efficiency. To test new primers a standard curve of cDNA which expresses the gene of interest was set up, using 1 in 5 dilutions (100%, 20%, 4%, 0.8%) of the initial starting cDNA in duplicate as well as a NTC. Primers were used if they were between 90-110% efficient as defined by the software on the qPCR machine and had only one clean peak in the dissociation curve when tested.

2.5 Immunofluorescence and confocal microscopy

Immunofluorescence staining was performed to analyse protein expression of relevant transcription factors at different stages of the pancreatic differentiation protocol. hiPSCs or their differentiated progenitors were fixed for 20 minutes at 4°C in 4% paraformaldehyde, washed twice in phosphate buffered saline (PBS) and then stored in 1ml PBS at 4°C. For intracellular immunostaining, cells were initially incubated for 5 minutes at room temperature in PBST (0.1% Triton-X 100 (Sigma-Aldrich) in PBS). Cells were incubated in 10% donkey serum (Bio-Rad, Kidlington, UK) in PBST for 1 hour (to block non-specific binding of antibodies) and subsequently incubated overnight at 4°C with primary antibody diluted to the appropriate concentration in 1% donkey serum in PBST (Appendix 4). Cells were then washed three times in 1% donkey serum in PBST and incubated with the appropriate secondary antibodies diluted

1:1000 in 1% donkey serum in PBST for 1 hour at room temperature in the dark. Cells were then washed three times in 1% donkey serum in PBST in the dark to remove unbound secondary antibodies. Cells were then incubated in DAPI (1:1000) for 15 minutes, then washed twice in PBS and stored in 1ml PBS at 4°C in the dark. Analysis was performed using a Zeiss LSM 700 confocal microscope (Carl Zeiss, Jena, Germany).

2.6 Flow cytometry

Adherent cells at the specific stage of the pancreatic differentiation protocol were washed once in PBS and then incubated for 5-10 minutes at 37°C in 500µl Accutase (Stem Cell Technologies, Vancouver, Canada). Cells were dissociated by gentle pipetting. Accutase was inactivated by adding two volumes of 5% fetal calf serum (FCS) (Invitrogen) in PBS to each well. The cell suspension was collected and then washed twice with PBS. Before the last wash, the cell suspension was filtered through a 40µm filter, to get a single cell suspension and remove any residual cell clumps. Live/dead staining (LIVE⁄DEAD Fixable Violet Dead Cell Stain Kit from Molecular Probes, Eugene, OR, USA, at a dilution of1:1,000) was then performed for 30 minutes on ice in FACS buffer. Cells were then resuspended in 200µl of Fixation solution (4% PFA) and incubated for 20 minutes on ice. Cells were washed twice with PBS and then stored in 500µl 2% FCS in PBS at 4°C.

For intracellular staining, cells were resuspended in 200µl of 1% Saponin (Sigma-Aldrich) in PBS and incubated for 1 hour at room temperature. Cells were incubated in primary antibody, diluted to an appropriate concentration (Appendix 4), in 100µl staining solution (5% FCS and 1% Saponin in PBS). Cells were subsequently washed twice with staining solution and then incubated in secondary antibody, diluted 1:1000 in 100µl staining solution, for 30 minutes at room temperature in the dark. Unbound secondary antibody was removed by three washes in staining solution and cells were resuspended in 300µl of 2% FCS in PBS prior to flow cytometric analysis. An unstained control and secondary antibody only control was prepared for each of the primary antibodies analyzed. Flow cytometric analysis was performed using a LSR Fortessa machine (BD Biosciences, CA, USA).

2.7 Annexin V apoptosis assay

Apoptotic cells were detected using the Annexin V Apoptosis Detection Kit FITC (Thermo Fisher Scientific, Waltham, MA, USA). Cells were detached and harvested using Accutase (see flow cytometry section) and washed once in PBS, then once in 1x Binding Buffer. Cells were filtered and then resuspended in 1x Binding Buffer at $2.5x10^6$ cells/ml. Cells were subsequently aliquoted in four conditions – Annexin V only, Annexin V and Propidium Iodide (PI), PI only, and no staining. 5µL of fluorochrome-conjugated Annexin V was added to 100µl of the cell suspension and incubated for 15 minutes at room temperature. After two washes with 1x Binding Buffer, cells were incubated for 10 minutes with 5µL PI (10µg/ml) to visualize dead cells. Cells were analysed by flow cytometry within four hours of staining. Flow cytometric data were analysed with FlowJo software.

2.8 Cell proliferation assay

Cell proliferation was measured using the Click-iT® EdU Flow Cytometry Assay Kit (Invitrogen). 5-ethynyl-2'-deoxyuridine (EdU) was added to the culture medium at 10µM for 2 hours. Cells were then detached and harvested using Accutase as previously discussed. $1x10^6$ cells were aliquoted and washed once with 3mL of 1% BSA in PBS. Cells were then incubated in 100µL of Click-iT fixative for 15 minutes at room temperature, protected from light. Cells were washed with 3mL of 1% BSA in PBS and incubated in 100µl of 1x Click-iT saponin-based permeabilization and wash reagent for 15 mins. 500µl of Click-iT reaction cocktail (CuSO4 10µl, Pacific Blue azide 2.5µl, 1x Reaction Buffer Additive 50µl, and PBS 438µl) was added to each sample and the reaction mixture was incubated for 30minutes at room temperature, protected from light. Cells were washed once with 3mL of 1x Click-iT saponin-based permeabilization and wash reagent and then resuspended in 1mL of 1x Click-iT saponin-based for staining for DNA content using the FxCycle™ Far Red stain (Invitrogen). 1µl of FxCycle™ Far Red stain (final concentration 200nM) and 20mg/mL of RNase A was added to each flow cytometry sample and incubated for 30 minutes at room temperature. FxCycle™ Far Red stains RNA as well as dsDNA, so addition of RNase A (Thermo Fisher Scientific) is required for DNA content analysis. Samples were analysed on a flow cytometer without washing, using 405/450nm (Pacific Blue) and 640/670nm (Far Red) excitation and emission spectra.

2.9 SDS-PAGE and Western blot analysis

Western blot analysis was employed to assess cellular protein expression. For harvesting, hiPSC-derived foregut and pancreatic progenitor cells were washed once with ice-cold PBS then incubated for 15 minutes in the required amount of CelLytic M reagent (Sigma-Aldrich) with PhosStop phosphatase inhibitor cocktail (Roche, Basel, Switzerland) and Complete protease inhibitor cocktail (Roche). The cells were removed from the plate using a cell scraper and transferred into a tube. To remove insoluble material, lysates were centrifuged full speed (14,000 rpm) at 4°C for 5 minutes. Pellets were discarded and the supernatant was stored at -80°C.

The supernatant was subjected to colorimetric protein quantification against a bovine serum albumin (BSA) standard curve using a protein quantification kit (Sigma-Aldrich), following manufacturer's guidelines. Samples were measured on a plate reader at 600nm and the protein concentration as compared to the BSA standard curve was calculated.
Protein samples were denatured, followed by protein separation, performed by SDS-PAGE. 10μg protein was loaded onto each well of a 4-12% gradient Bis-Tris gradient gels (Invitrogen) assembled in an XCell SureLock Mini-Cell (Invitrogen) with 1x NuPage MOPS running buffer (Invitrogen). Transfer of proteins onto a PVDF membrane was performed using the Bio-Rad mini-Trans Blot system (Bio-Rad).

Membranes were blocked for 1 hour in 4% powdered skimmed milk or 5% BSA (for phospho-antibodies), diluted in PBS – 0.05% Tween-20 for 1 hour, shaking, at room temperature or overnight at 4°C. Membranes were then incubated in primary antibody (Appendix 4), on a rotator, overnight at 4°C or for 1 hour at room temperature. Membranes were rinsed three times in PBS-0.05% Tween-20 on a shaker for 5 minutes then incubated in horseradish peroxidise-linked anti-rabbit or anti-mouse IgG secondary antibody, diluted 1:10,000 in blocking solution on a rotator for 1 hour at room temperature. This was followed by three 5-minute washes in PBS-0.05% Tween-20 on a shaker.

Proteins were visualized using enhanced chemoluminescence (ECL detection kit; Thermo Fisher Scientific). The blot was developed using an imaging system and imaging film was exposed for 1 second to 15 minutes, depending on the strength of the signal.

2.10 ELISA for C-peptide

Human iPSC-derived pancreatic endocrine cells at day 27 were preincubated in DMEM supplemented with 2.5μM glucose (low glucose media) for 60 minutes at 37°C. To measure basal C-peptide release, cells were incubated in low glucose media for 60 minutes at 37°C. To estimate glucose-induced C-peptide secretion, the media was replaced by DMEM supplemented with 22.5 mM glucose (high glucose media) and alternatively with DMEM supplemented with 2.5 mM glucose for 60 minutes at 37°C. The low glucose – high glucose stimulation was repeated for a second cycle. Finally, cells were incubated in low-glucose DMEM with 30mM potassium chloride (KCl) for 30 min. The supernatant at each stage was collected and stored at -80°C for determination of C-peptide release.

C-peptide ELISA was measured using the Mercodia C-peptide ELISA kit (Mercodia, Uppsala, Sweden), following the manufacturer's recommendations. Absorbance was read at 450nm on an Infinite 200 Pro plate reader (Tecan, Männedorf, Switzerland).

2.11 In vivo transplantation studies

2.11.1 Animal models and ethics

Animal experiments were performed in collaboration with Mr Kourosh Saeb-Parsy at the Department of Surgery, University of Cambridge. All animals were maintained in specific-pathogen-free facilities and experiments have been regulated under the Animals (Scientific Procedures) Act 1986 Amendment Regulations 2012 following ethical review by the University of Cambridge Animal Welfare and Ethical Review Body (AWERB) under the project license PPL 70/8702. Animals were sourced from the Jackson Laboratory (Bar Harbor, ME, USA). Immunodeficient NOD/SCID/Gamma (NSG) mice, which lack B, T and NK lymphocytes, were bred in house. A mix of male and female animals aged approximately 6–8 weeks was used.

2.11.2 Pancreatic progenitor transplantation studies

Human iPSC-derived pancreatic progenitor cell (day 13) aggregates in Matrigel (Corning, Corning, NY, USA) were loaded into a syringe for cell delivery under the kidney capsule of NSG mice (5 million cells per animal). In vivo glucose stimulated insulin secretion (GSIS) was

performed at 18 weeks after transplantation. Mice were fasted for 6-8 hours and blood was collected from tail veins under fasting condition and at 30 and 60 minutes after intraperitoneal (IP) injection with 2g D-(+)-glucose/1kg body weight. Human insulin levels in the mouse sera were quantified immediately after blood collection using the Human Ultrasensitive C-Peptide ELISA kit (Mercodia). The grafted and normal control kidneys were removed at 18 weeks after transplantation for hematoxylin and eosin (H&E) staining and immunostaining analysis. Tissue samples were embedded in optimum cutting temperature (OCT, VWR International, Radnor, PA, USA) and stored at -80°C. Frozen blocks were cryostat sectioned at 10μm thickness and slides were stored in -80°C. Frozen graft samples embedded in OCT were sectioned to 10μl thickness using a cryostat. Immunofluorescence staining was performed as described in section 2.5. Some extracted grafts were fixed in 4% paraformaldehyde, wax embedded and processed for immunohistochemistry by Dr Rebecca Brais, Department of Histopathology, University Hospital Cambridge.

2.12 Statistical analysis

Quantification data are presented as mean ± SEM. Data from clonal lines of the same genotype were combined for calculating the significance of the differences between different genotypes. To directly compare two groups, Student's t test with two-tailed distribution was used to test for statistical significance. P values less than 0.05 was considered statistically significant. All statistical analyses were performed using GraphPad Prism 6.0 (GraphPad Software Inc., San Diego, CA, USA) or the R statistical environment.

2.13 Next Generation Sequencing (NGS)

2.13.1 RNA-sequencing (RNA-seq)

RNA was extracted using the GenElute Mammalian Total RNA Miniprep Kit (Sigma-Aldrich) according to manufacturer's protocol. RNA-seq library construction and sequencing was carried out by the DNA pipelines core facility at the Wellcome Trust Sanger Institute. Standard Illumina unstranded poly-A enriched libraries were prepared and sequenced using Illumina HiSeq 2500 v4 (Illumina, San Diego, CA, USA), with 75bp paired-end reads per sample and a library fragment size of 100-1000 bp. RNA-seq samples were run in biological triplicates.

Sequencing reads were aligned to the reference human genome assembly GRCh38 with Ensembl version 76 annotation using BWA version 1.14.9 (Li and Durbin, 2009). Reads overlapping gene annotations were counted using summarizeOverlaps (Lawrence et al., 2013) and genes with no counts were filtered out. Differentially expressed genes were identified using DESeq2 R/Bioconductor package (Love et al., 2014). Genes with fold change ≥2 and adjusted p-value (using Benjamin-Hochberg correction) <0.05 were identified as differentially expressed. Counts were normalised using the fragments per kilobase of transcript per million mapped reads (fpkm) function in the DESeq2 package. All analysis was performed on genes classified as expressed (fpkm >1) in all conditions being compared, except where noted otherwise. Gene Ontology and KEGG Pathway analyses were performed using the limma package (Ritchie et al., 2015). Samtools (Ritchie et al., 2015) was used to construct and filter BAM files. All downstream analysis was performed using R and ggplot2 was used to construct figures.

2.13.2 Chromatin immunoprecipitation and sequencing

Co-binding of DNA to DNA-binding proteins was determined by ChIP against HNF1B (sc22840-X Santa Cruz, Dallas, TX, USA) on approximately 1 x 10^7 cells per antibody or control sample. Cells were crosslinked with 1% formaldehyde (Thermo Fisher Scientific) for 10 minutes at room temperature. The reaction was quenched with 0.125M glycine (Millipore, Billerica, MA, USA) for 5 minutes. Cells were washed twice with ice-cold PBS then collected in ice-cold PBS containing freshly-added protease inhibitors (10µl/ml of 5mg/ml phenyl-methyl-sulfonyl-fluoride (PMSF; Sigma-Aldrich), 10µl/ml of 1M Sodium Butyrate (Sigma-Aldrich) and 1µl/ml of 1mg/ml Leupeptin (Roche). Harvested cells were centrifuged for 5

minutes 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 (10mM Tris-Cl pH 8.0, 10mM NaCl and 0.2% NP-40), incubated on ice for 10 min, and then centrifuged for 5 min at 1,800 rpm at 4°C. The supernatant was discarded and the pellet was gently re-suspended in 1.25ml of ice-cold Nuclear Lysis Buffer (50mM Tris-Cl pH 8.0, 10mM EDTA and 1% SDS) and incubated on ice for 10 minutes. 0.75ml of ice-cold immuno-precipitation (IP) dilution buffer (20mM Tris-Cl pH 8.0, 2mM EDTA, 150mM NaCl, 0.01% SDS, 1% Triton X-100) was then added.

The chromatin was sonicated using Diagenode Biorupter Pico in 15 ml Diagenode sonication tubes containing sonication beads (Diagenode, Denville, NJ, USA) pre-washed with 10ml D-PBS and 10ml IP dilution buffer for 10 cycles of 30 seconds on/45 seconds off. Chromatin fragments were determined by a Bioanalyser (Agilent 2100 Bioanalyzer) and analysed using High Sensitivity DNA Kit (Agilent Technologies) according to the manufacturer's protocol. The sonicated chromatin was then centrifuged at 14,000 rpm for 10 minutes at 4°C to pellet debris. 3.5ml of IP dilution buffer was added to the supernatant and mixed gently.

The cross-linked DNA was pre-cleared by incubation with 10µg of rabbit IgG (Sigma-Aldrich) rotating overnight at 4°C, followed by incubation with 100µl of Protein G agarose beads (50% v/v; Roche) pre-washed twice with D-PBS for 1 hour at 4°C. The samples were then centrifuged for 3 minutes at 3,000 rpm at 4°C and the supernatant was transferred to a fresh 15ml tube. An aliquot of 300µl for input sample was taken and stored at 4°C. 10µg of HNF1B antibody or rabbit HRP IgG control was added per sample and incubated rotating overnight at 4°C. Antibody-bound chromatin was then collected using 60µl of Protein G agarose beads (50% v/v) pre-washed twice with D-PBS by incubating with rotation for 1 hour at 4°C. Thereafter, the tubes were centrifuged for 3 minutes 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. Samples were washed twice with 500µl of IP wash buffer 1 (20mM Tris-Cl pH 8.0, 2mM EDTA, 50mM NaCl, 0.1% SDS and 1% Triton X-100), twice with 500µl of IP wash buffer 2 (10mM Tris- Cl pH 8.0, 1mM EDTA, 0.25M LiCl, 1% NP-40 and 1% Sodium deoxycholic acid), twice with 500µl of TE Buffer (10mM Tris-Cl pH 8.0, 1mM EDTA) and then eluted by

washing twice with 150µl of Elution Buffer (100mM $NaHCO_3$ and 1% SDS).

ChIP and input DNA cross-linking was reversed and RNA degraded by adding 1µl of 1mg/ml RNase A and 18µl of 5M NaCl and incubating at 67°C in a heat block with shaking at 1,300 rpm overnight. Protein was degraded by adding 3µl of 20mg/ml Proteinase K and incubating for 3 hours at 45°C in a heat block with shaking at 1,300 rpm. Pulled-down genomic DNA was extracted using 300µl of phenol/chloroform wash. The samples were next incubated with 30µl of 3M sodium acetate pH 5.2 (Ambion, Carlsbad, CA, USA), 30µg glycoblue (Ambion) and 750µl of 100% ethanol for at least 30 minutes at -80°C to precipitate the DNA. Precipitated DNA was pelleted by centrifuging at 14,000 rpm for 30 minutes at 4°C. The DNA pellet was then washed with ice-cold 70% ethanol then air dried. 70µl of deionised water was added to input samples whereas 30µl of deionised water was added to ChIP samples.

2.13.3 Assay for transposase-accessible chromatin using sequencing (ATAC-seq)

The ATAC-seq protocol used in this project was a modified version of the original protocol developed in the Greenleaf laboratory (Buenrostro et al., 2013). To prepare nuclei, adherent cells (one well of 12-well plate) were washed twice with ice-cold PBS then incubated on ice for 12 minutes in 500µl ice-cold sucrose buffer (10mM Tris-Cl pH 7.5, 3mM $CaCl_2$, 2mM $MgCl_2$, 0.32M sucrose). 25µL of 10 % Triton-X-100 was added (final concentration 0.5%) and incubated on ice for 6 minutes. Cells were then scraped and transferred to an Eppendorf tube. 100,000 cells were aliquoted and centrifuged at 300g for 5 minutes at 4°C. The supernatant lysis buffer was discarded prior to tagmentation, which involves transposon cleaving and tagging of the double-stranded DNA with a universal overhang.

Tagmentation was performed using the Illumina Nextera DNA Preparation Kit. Immediately following the nuclei prep, the pellet was resuspended in the transposase reaction mix (25µl 2× TD buffer, 5µl transposase (Illumina) and 20µl nuclease-free water). The transposition reaction was carried out for 30 minutes at 37°C. Following transposition, the reaction was stopped with 500µl of binding buffer (PB buffer, Qiagen GmbH, Hilden, Germany) and the sample was purified using a Qiagen MinElute PCR purification kit (Qiagen). The sample was eluted in 10-12µl elution buffer (EB). Following purification, library fragments were amplified using the following reaction: 7.5µl Nextera PCR mastermix, 2.5µl Nextera i7 and i5 primers, 2.5µl primer cocktail, 10µl template from tagmentation, using the following PCR conditions: 72°C

for 3 minutes; 98°C for 30 seconds; and thermocycling x12 at 98°C for 10 seconds, 63°C for 30 seconds and 72°C for 3 minutes.

Size selection and library purification was performed by running the sample on 50mL 1% agarose TAE at 90 V for 25 minutes; cutting PCR bands from ~120 to 1000 bp followed by gel extraction (Qiagen MiniElute gel extraction kit). Purified samples were eluted in 20µl EB. Samples were cleaned up further using SPRI beads (Agencourt SPRIselect, Beckman Coulter, Indianapolis, IN, USA) and finally eluted in 20µl EB. Library profiles were checked using an Agilent High Sensitivity DNA assay chip on an Agilent 2100 Bioanalyzer (Agilent Technologies).

Sequencing was performed by the DNA pipelines core facility at the Wellcome Trust Sanger Institute, using Illumina HiSeq 2500 v4 (Illumina Inc, CA, USA) with 75bp paired-end reads per sample and a library fragment size of 100-1000 bp. Twelve samples were pooled and sequenced over 4 lanes RNA-seq to obtain at least 100 million PE reads per sample. Reads were aligned to GRCh38 human reference genome Ensembl version 76 using BWA version 1.14.9 (Li and Durbin, 2009). For all data files, duplicate fragments were removed using samtools rmdup. Reads mapped to mitochondria, and mapped reads with quality score less than 10 were filtered out.

Peak-calling was performed using MACS2 (Zhang et al., 2008) version 2.1.1 with "--nomodel –shift -100 --extsize 200 –q 0.01" to identify open chromatin regions (peaks) that were enriched for transposase integration sites compared to the background at 1% FDR. Default settings were used for other parameters.

Differentially accessible chromatin regions were identified using diffReps (Shen et al., 2013). diffReps is designed as a PERL program that uses a sliding window to scan the genome and identifies the ones that show read count differences. A sliding window is defined as a fix-sized genomic region where the reads falling into this region can be counted.

Wellington Bootstrap (Piper et al., 2015), was used to identify unique footprints between regions of gained-open and gained-closed chromatin accessibility. Sample replicates (aligned reads) were merged to achieve more depth. The genome was scanned in regions aggregated from peak calls, and merged if closer than 100bp (extsize/MACS -total 396,572 ATAC-seq genomic regions). ATAC-seq signal / Tn5 frequencies were obtained using NucleoATAC.

Chapter 3 Creation of HNF1B knockout human iPSC lines

3.1 Introduction

The combination of hPSCs and genome-editing technology (and genome wide association studies) provides a powerful platform not only to systematically model human metabolic disease in relevant cell types but also to study the molecular mechanisms underlying these diseases. Generation of isogenic hPSC lines, which only contain differences in disease-causing mutations, has proved to be an effective strategy for modelling human diseases. These isogenic PSC lines have the same genetic background, epigenetic state and similar differentiation capacities. These similarities can substantially reduce variability among different lines and simplify analyses of interactions between genotype and phenotype.

Genome editing techniques enable efficient modification of targeted loci (site mutations, deletions and insertions) through the introduction of DNA double-strand breaks (DSBs) by programmable sequence-specific endonucleases to facilitate DNA damage repair by homologous recombination (HR) or non-homologous end joining (NHEJ). The three most widely used genome editing technologies are 1) zinc-finger nucleases (ZFNs), 2) transcription activator-like effector nucleases (TALENs) and 3) Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas.

ZFNs and TALENs rely on the generation of chimeric proteins consisting of an endonuclease catalytic domain and a DNA binding motif for inducing DSBs at the desired genomic loci. CRISPR was first identified in a prokaryotic immune system that uses RNA-guided nucleases to cleave foreign genetic elements. The Type II CRISPR system is the best characterized, consisting of the nuclease Cas9, the crRNA array that encodes the guide RNAs and a required auxiliary trans-activating crRNA (tracrRNA) that facilitates the processing of the crRNA array into discrete units. Each crRNA unit then contains a 20-nt guide sequence and a partial direct repeat, where the former directs the Cas9 to a 20-bp DNA target via base pairing. gRNAs must match a 20-nucleotide target sequence (protospacer sequence) in the genomic DNA and must be followed by a protospacer adjacent motif (PAM) sequence of NGG. This NGG motif is essential for DNA cleavage, which occurs at ~3 bp upstream of the PAM (Figure 3.1.) (Ran et al., 2013).

The CRISPR/Cas9 system has been widely used for genome editing of cultured mammalian cells, due to its precision and higher efficiency and ease of use compared with ZFNs and TALENs. In this project, we chose to use the CRISPR/Cas system to generate HNF1B heterozygous and homozygous knockout hiPSC lines and to correct mutations in 2 patient-derived iPSC lines

Figure 3.1. Schematic of RNA-guided Cas9 nuclease targeting and DSB repair pathways
(A) The Cas9 nuclease from *S. pyogenes* (in yellow) is targeted to genomic DNA by an sgRNA consisting of a 20-nt guide sequence (blue) and a scaffold (red). The guide sequence pairs with the DNA target (blue bar on top strand), directly upstream of a requisite 5′-NGG adjacent motif (PAM; pink). Cas9 mediates a DSB ~3 bp upstream of the PAM (red triangle).
(B) DSBs induced by Cas9 can be repaired by NHEJ or HDR
(Ran et al., 2013)

3.2 Results

3.2.1 Gene targeting strategy to generate HNF1B knockout lines

To create HNF1B homozygous and heterozygous knockout hiPSCs, a customized CRISPR system was used consisting of three components: (1) a plasmid encoding the full length Cas9 protein codon optimized for optimal expression in human cells and driven by the CMV promoter; and (2) a separate plasmid containing a U6 promoter-driven gRNA scaffold region into which custom gRNA sequences are cloned (3) a plasmid-based donor repair template that contains homology arms (1000bp in length) flanking the site of alteration and a stem cell-compatible eukaryotic transcription elongation factor 1 alpha 1 (EF1a) promoter-driven puromycin resistance (Puro-R) cassette for integration into the endogenous genomic locus. This method can be used to generate large modifications, including insertion of reporter genes such as fluorescent proteins or antibiotic resistance markers. Replacement of an essential coding region in a genetic locus with a fluorescent proteins or antibiotic resistance marker results in disruption of the gene's open reading frame and blocks its expression. This replacement occurs via recombination at the two flanking homology arms (Figure 3.2.).

FSPS13.B and Eipl_1 hiPSCs were received from the CGaP facility, Wellcome Trust Sanger Institute at passages (P) 56 (26 passages in E8 media) and 44 (21 passages in E8 media) respectively. FSPS13.B and Eipl_1 hiPSCs were karyotyped prior to any experiments (gene targeting or differentiations) at P61 (E8 31) and P51 (E8 28) respectively. The Cas9 vector, HNF1B gRNA vector and donor vector were co-nucleofected into two iPSC lines (FSPS13.B and Eipl_1). The FSPS13.B hiPSC line was nucleofected at passage P69 (E8 39) and the Eipl_1 hiPSC line was nucleofected at passage P46 (E8 23).

Following selection with puromycin, the surviving colonies were individually picked and expanded for genotyping to identify HNF1B targeted wild-type (WT), heterozygous and homozygous knockout clones (Figure 3.2. and Figure 3.3.). The targeted segment of DNA was amplified using primer pairs, with one primer contained within the selection cassette and the other outside the homology arm (5' end and 3' end) of the donor vector (Figure 3.2.). PCR primers for these purposes should anneal outside the region spanned by the homology arms to avoid false detection of residual repair template. PCR products arising from clones screened with this combination of primers, spanning the junction between the target and insert, indicate that these puromycin-resistant clones contain the integrated selection cassette in the correct

location and orientation in at least one allele of the gene. Sanger sequencing was performed to check if integration of the puromycin cassette occurred in one or both alleles of the HNF1B gene (Figure 3.3). The HNF1B targeted locus was amplified using primer pairs around the targeted locus. PCR products arising from clones screened with this combination of primers were sequenced to check for NHEJ or if the sequence was wild-type (Figure 3.3).

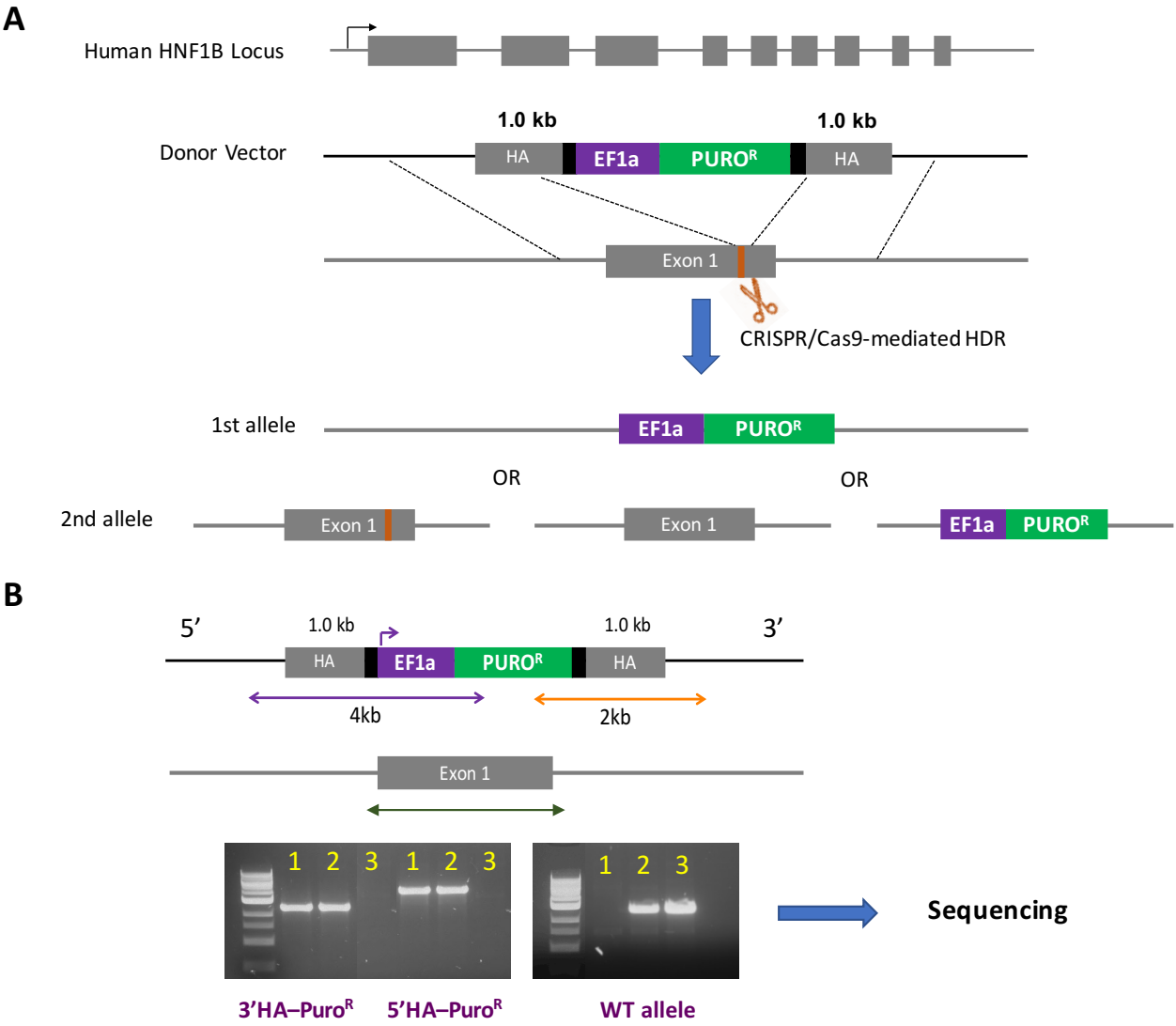


Figure 3.2. Schematic for gene targeting of the HNF1B gene using CRISPR/Cas9 mediated HDR

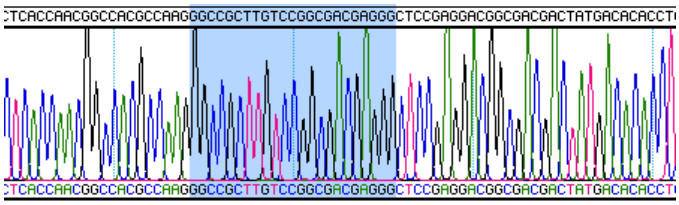
(A) Exon 1 of the HNF1B gene is targeted using a Cas9 expression plasmid and a U6-driven gRNA plasmid and a donor repair template containing 1kb homology arms flanking an EF1a promotor driven Puro-Resistance expression cassette (Puro^R).

(B) Genotyping of HNF1B targeted clones using PCR. Bands 1, 2 and 3 are representative of:

1. HNF1B homozygous knockout (HNF1B$^{-/-}$) – Puro^R insertion in both alleles, no WT band
2. HNF1B heterozygous knockout (HNF1B$^{+/-}$) - Puro^R insertion in 1st allele, WT band in 2nd allele
3. HNF1B wild-type (HNF1B$^{+/+}$) – no Puro^R insertion, WT band in both alleles

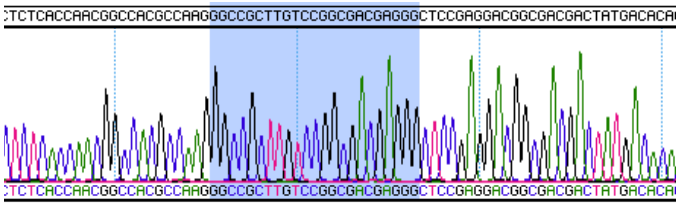
Alleles 1 and 2

WT



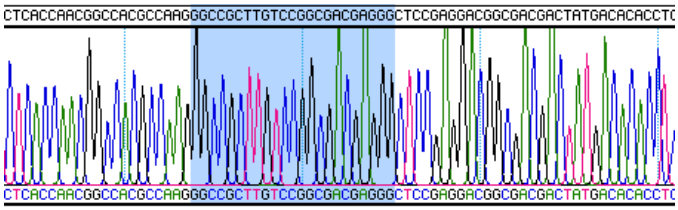
2 WT alleles

cl.15



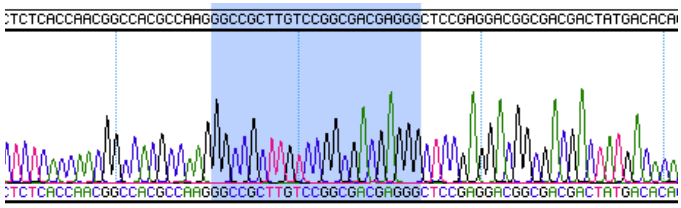
2 WT alleles

cl.45



2 WT alleles

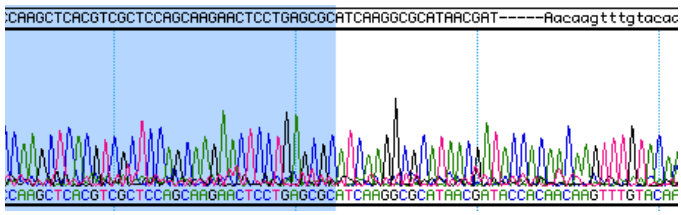
Allele 1



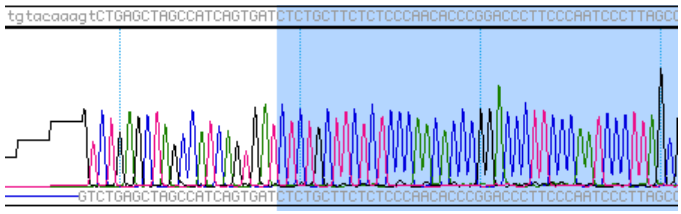
WT allele

Allele 2

cl.6

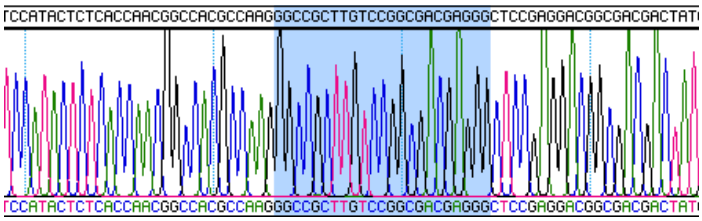


5'HA - Ef1a-PURO^{R}



Ef1a-PURO^{R} - 3'HA

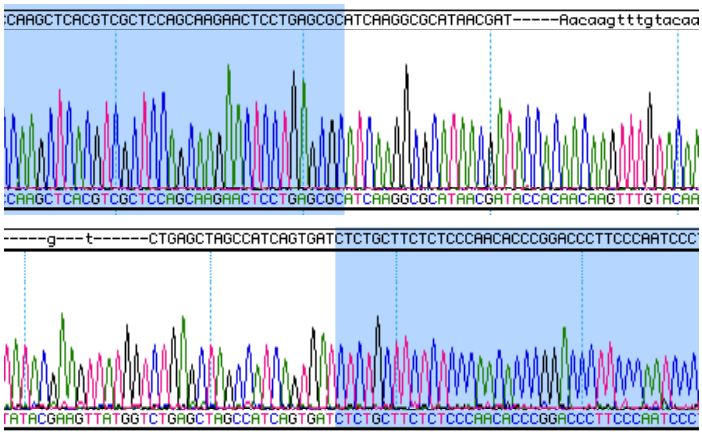
Allele 1



WT allele

Allele 2

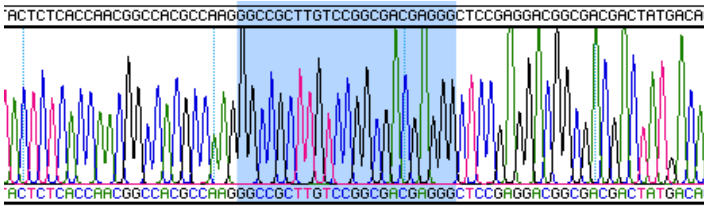
cl.28



5’HA - Ef1a-PURO^{R}

Ef1a-PURO^{R} - 3’HA

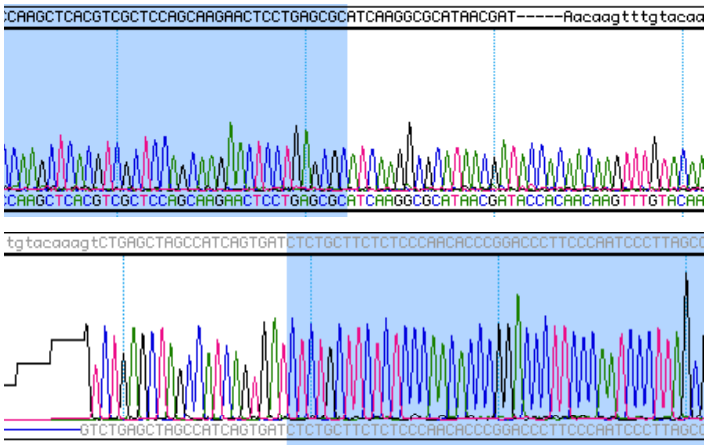
Allele 1



WT allele

Allele 2

cl.44



5’HA - Ef1a-PURO^{R}

Ef1a-PURO^{R} - 3’HA

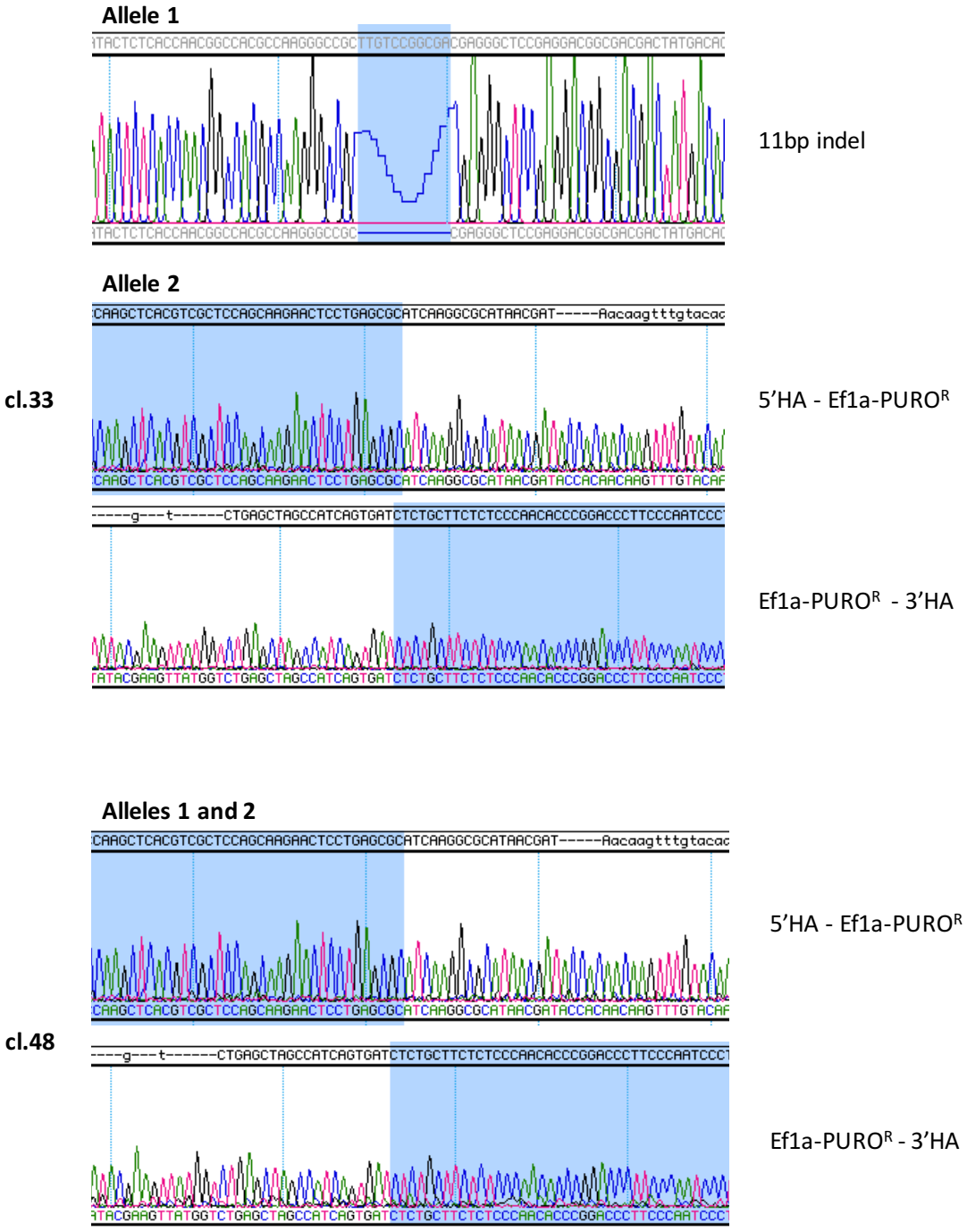


Figure 3.3. Sanger sequencing of FSPS13.B targeted clones to determine genotype.
3 HNF1B wild-type clones (HNF1B$^{+/+}$; one non-targeted wild-type (WT) and two targeted WT clones (cl.15 and cl.45)), 3 HNF1B heterozygous knockout clones (HNF1B$^{+/-}$; cl.6, cl.28, cl.44) and 2 HNF1B homozygous knockout clones (HNF1B$^{-/-}$; one with puromycin resistance cassette (Puro^R) in 1st allele and indel in 2nd allele (cl.33) and one with puromycin resistance cassette (Puro^R) in both alleles (cl.48)).

For the FSPS13.B hiPSC line, out of the 48 clones picked and genotyped (Table 3.1A), 19 were positive for correct integration of the puromycin cassette in at least 1 allele, at an overall efficiency of 40%. Of the 19 clones, the puromycin cassette was integrated into both alleles in one clone (no wild-type allele detected on PCR). A wild-type band was detected by PCR in 18 clones and the PCR product was sent for sequencing to check for NHEJ. Small insertions/deletions (indels) were found in 14 out of the 18 clones. An out-of–frame deletion or insertion was found in 4 out of the 18 clones. Absence of puromycin integration in the targeted locus was observed in 4 clones which were puromycin-resistant suggesting random integration of the puromycin cassette in the genome. These clones were used as targeted wild-type controls in experiments.

For the Eipl_1 hiPSC line, out of the 96 clones picked and genotyped (Table 3.1B), 36 were positive for correct integration of the puromycin cassette in at least 1 allele, at an overall efficiency of 37%. Of the 36 clones, the puromycin cassette was integrated into both alleles in one clone (no wild-type allele on PCR). A wild-type band was detected by PCR in 35 clones and the PCR product was sent for sequencing to check for NHEJ. Small (indels) were found in 29 out of the 35 clones, 10 of which would result in a frameshift. No puromycin integration was found in 8 clones which were puromycin-resistant suggesting random integration of the puromycin cassette in the genome. These clones were used as targeted wild-type controls in experiments.

A

FSPS13.B	1st allele	2nd allele	Number of targeted clones
WT	WT	WT	4
Heterozygous	Puro^R cassette	WT	3
Homozygous	Puro^R cassette	Puro^R cassette	1
	Puro^R cassette	NHEJ (frameshift mutation)	14 (4)

B

Eipl_1	1st allele	2nd allele	Number of targeted clones
WT	WT	WT	8
Heterozygous	Puro^R cassette	WT	5
Homozygous	Puro^R cassette	Puro^R cassette	1
	Puro^R cassette	NHEJ (frameshift mutation)	29 (10)

Table 3.1. Summary of genotypes for the targeted clones for **(A)** FSPS13.B and **(B)** Eipl_1 hiPSC lines. The number of clones with no integration of the puromycin resistance cassette (HNF1B WT clones) or integration of the puromycin resistance cassette in one or two alleles (HNF1B homozygous knockout) of the HNF1B gene is shown. For clones where there is integration of the puromycin resistance cassette in one allele, the 2nd allele was either WT (HNF1B heterozygous knockout) or contained an in-frame or frameshift mutation (HNF1B homozygous knockout).

3.2.2 Off target effects

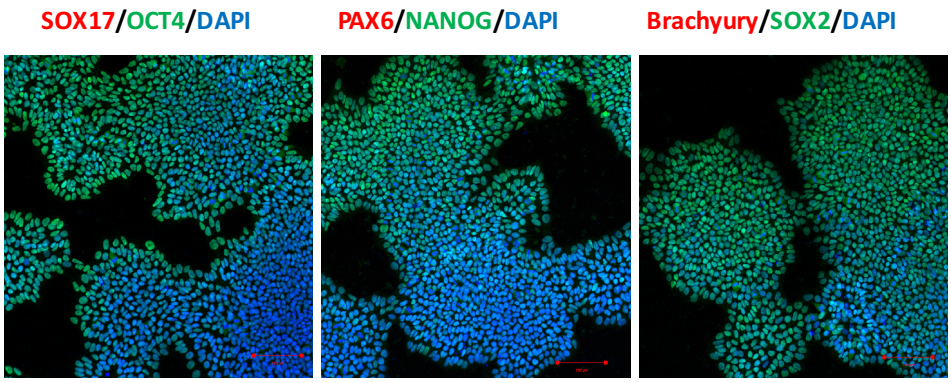
The clones selected for use in further experiments were analyzed for off-target effects of RNA-guided Cas9 nuclease. Nine potential off-target regions, including three in exonic regions in protein coding genes were identified. Sanger sequencing was performed for the regions in the 3 protein coding genes; 1) sema domain, transmembrane domain (TM), and cytoplasmic domain, (semaphorin) 6D (SEMA6D); 2) Rho GTPase activating protein 23 (ARHGAP23) and 3) Fragile Histidine Triad (FHIT). No mutations (single nucleotide variants, insertions and deletions) within 500 bp of a predicted off-target site in these genes were found (Appendix 5).

3.2.3 Characterisation of HNF1B targeted hiPSC lines

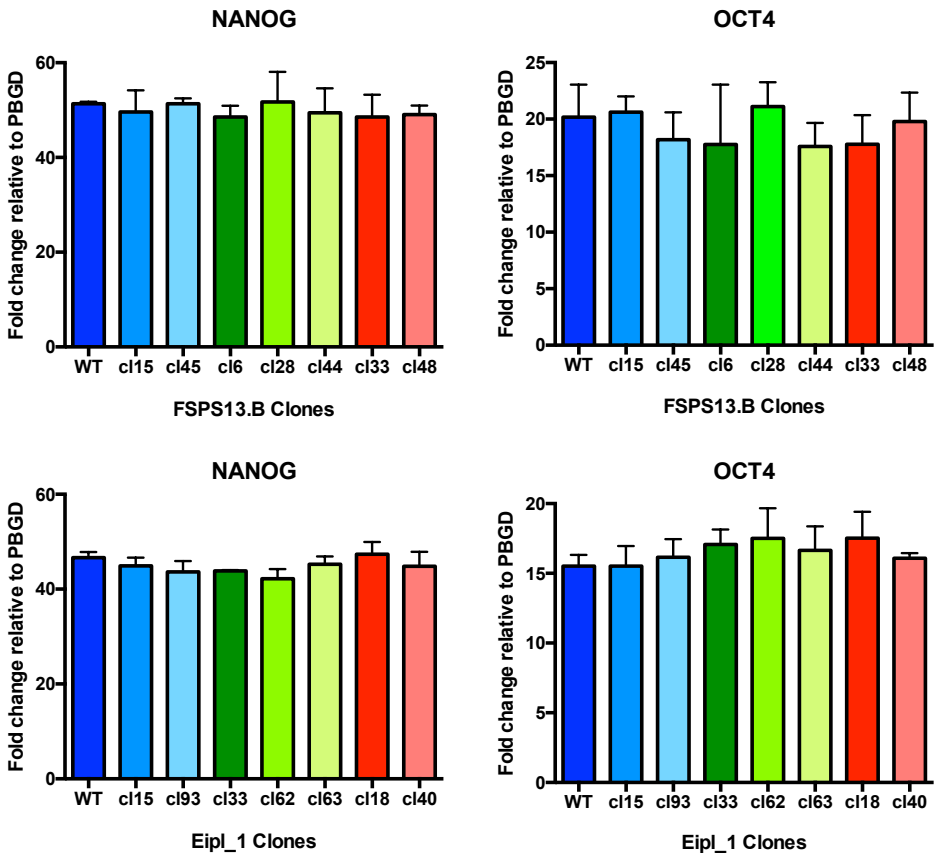
The FSPS13.B and Eipl_1 hiPSC clones selected for use in subsequent experiments included: 3 wild-type clones ($HNF1B^{+/+}$; one non-targeted wild-type and two targeted wild-type clones), 3 heterozygous clones ($HNF1B^{+/-}$) and 2 homozygous clones ($HNF1B^{-/-}$; one with puromycin cassette in both alleles and one with puromycin cassette in 1st allele and indel in 2nd allele) (Figure 3.3). The $HNF1B^{+/+}$, $HNF1B^{+/-}$ and $HNF1B^{-/-}$ hiPSCs selected for use in further experiments were characterised for pluripotency markers using immunofluorescence and qPCR. Undifferentiated $HNF1B^{+/+}$, $HNF1B^{+/-}$ and $HNF1B^{-/-}$ hiPSCs showed similar expression of pluripotency markers (OCT4, NANOG and SOX2) by qPCR and immunofluorescence (Figure 3.4A and B). All hiPSC lines were karyotyped prior to the first differentiation; at passage 71 (E8 41) for the FSPS13.B hiPSC lines and at passage 50 (E8 27) for the Eipl_1 hiPSC lines, and then every 20 passages to check for chromosomal abnormalities (Figure 3.4C and Appendix 6).

Human iPSCs were differentiated along the pancreatic lineage up to the foregut progenitor stage when HNF1B is initially expressed. The differentiation protocol is described in more detail in chapter 4. Western blot, immunofluorescence and qPCR confirmed absence of HNF1B at the protein (Figure 3.4D) and RNA level (Figure 3.4E) respectively in the $HNF1B^{-/-}$ cell lines and qPCR showed ~50% reduction in HNF1B expression in the $HNF1B^{+/-}$ cell lines.

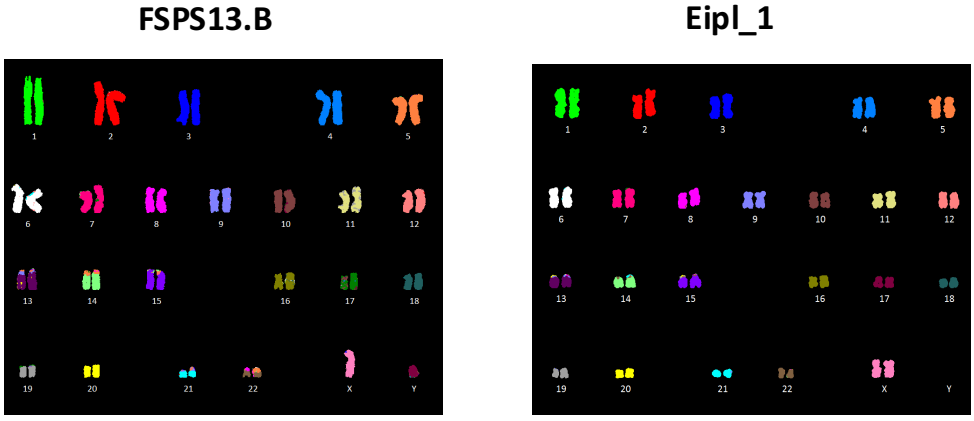
A


SOX17/OCT4/DAPI
PAX6/NANOG/DAPI
Brachyury/SOX2/DAPI

B


NANOG
OCT4
Fold change relative to PBGD
WT cl15 cl45 cl6 cl28 cl44 cl33 cl48
FSPS13.B Clones
NANOG
OCT4
Fold change relative to PBGD
WT cl15 cl93 cl33 cl62 cl63 cl18 cl40
Eipl_1 Clones

C


FSPS13.B
Eipl_1

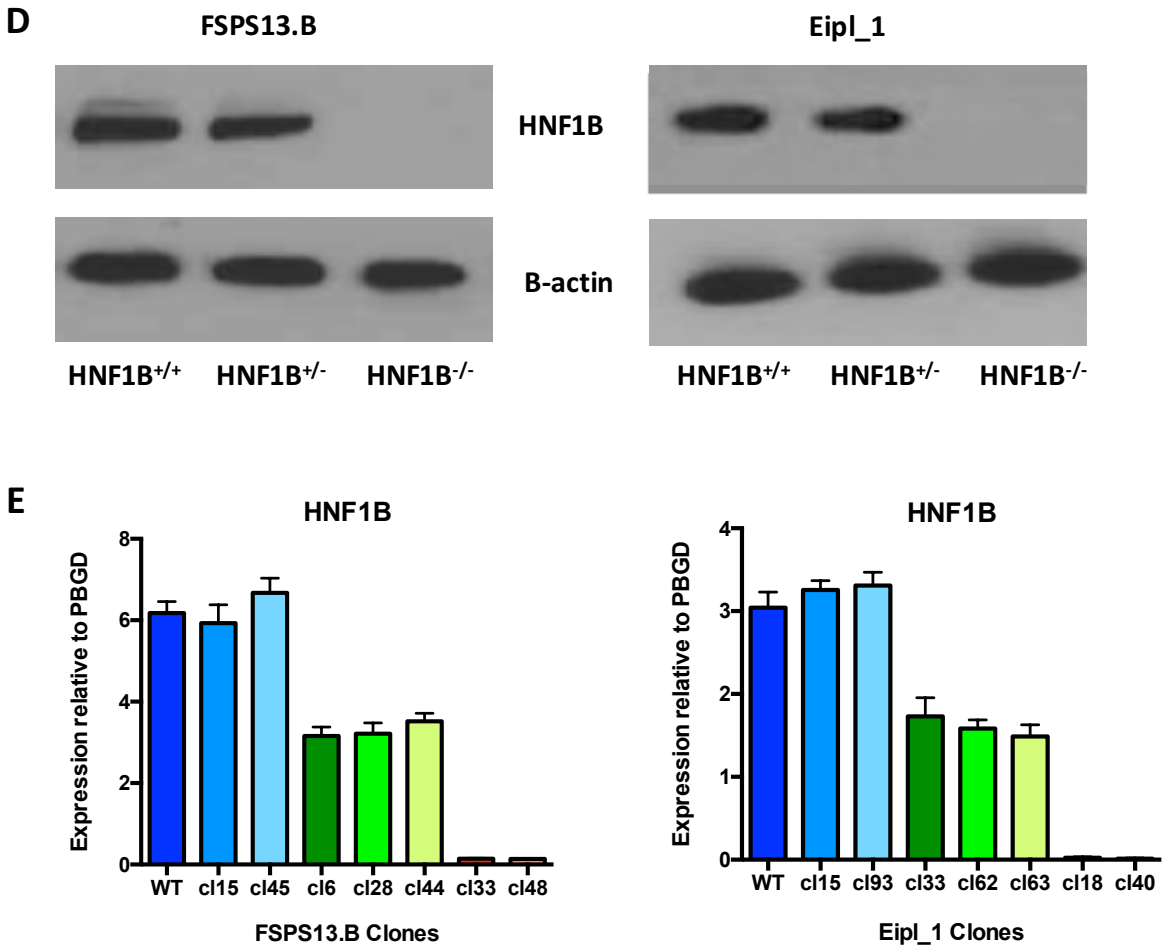


Figure 3.4. Characterisation of targeted clones for FSPS13.B and Eipl_1 hiPSC lines: HNF1B$^{+/+}$ (HNF1B WT and targeted WT), HNF1B$^{+/-}$ (heterozygous knockout) and HNF1B$^{-/-}$ (homozygous knockout) clones.
(A) Representative immunofluorescence images showing expression of pluripotency markers OCT4, NANOG and SOX2 and absence of expression of endoderm, mesoderm and neuroectoderm lineage markers in undifferentiated hiPSCs.
(B) mRNA expression of the pluripotency markers NANOG and OCT4 for undifferentiated HNF1B$^{+/+}$ (blue), HNF1B$^{+/-}$ (green) and HNF1B$^{-/-}$ (red) clones for the FSPS13.B and Eipl_1 hiPSC lines.
(C) Chromosomal analyses showing normal karyotype for FSPS13.B and Eipl_1 WT hiPSCs at passage 71 (E8 41) for the FSPS13.B hiPSC line and at passage 50 (E8 27) for the Eipl_1 hiPSC line (representative images). Chromosomal analyses for the other clones are shown in Appendix 6. Cells were karyotyped every 20 passages.
(D) Western blot showing expression of HNF1B protein at ~65kb against housekeeping control B-actin 42kb at the foregut progenitor stage (Day 6) for HNF1B$^{+/+}$, HNF1B$^{+/-}$ and HNF1B$^{-/-}$ clones for the FSPS13.B and Eipl_1 hiPSC lines.
(E) mRNA expression of HNF1B at the foregut progenitor stage (Day 6) for HNF1B$^{+/+}$ (blue), HNF1B$^{+/-}$ (green) and HNF1B$^{-/-}$ (red) clones for the FSPS13.B and Eipl_1 hiPSC lines.
mRNA levels were measured by qRT-PCR and normalised to the house-keeping gene PBGD. Error bars represent SEM.

3.2.4 Derivation and characterisation of patient derived hiPSC lines

The demographic, genotypic and phenotypic characteristics of the patients with HNF1B-associated disease, from whom skin biopsy samples were collected for this study, was highly variable and are summarised in Table 3.2. Fibroblasts were reprogrammed to hiPSCs by the CGaP facility, Wellcome Trust Sanger Institute and Cambridge BRC hiPSC core facility, using Sendai virus reprogramming. Three independent hiPSC clones were isolated for each patient line. All normal and patient-derived hiPSC lines used were validated by the Cambridge BRC iPSC Core Facility or by the CGaP facility, Wellcome Trust Sanger Institute, for expression of endogenous pluripotency markers versus transgenes by qPCR and by in vitro differentiation into the three germ layers (Figure 3.5).

3.2.5 Correction of mutations in patient iPSC lines

Two patient iPSC lines (R2 and R4) were selected, based on the quality of the iPSC line and ease of maintaining in culture, both with mutations in exon 4 of the HNF1B gene. Patient iPSC line R2 had a 1bp duplication mutation, g.36091625dup resulting in a frameshift (p.His336fs) and patient R4 had nonsense mutation, g.36091805G>A (p.Arg276*). R2 and R4 hiPSC lines were received from the Cambridge BRC iPSC Core Facility at P12 and P10 respectively. Cells were karyotyped at P24 (E8 10) for the R2 hiPSC line and P13 (E8 13) for the R4 hiPSC line (Appendix 6).

As described in Chapter 2, two targeting methods were used to attempt to correct mutations in 2 selected patient iPSC lines, the 1st using the pSpCas9(BB)-2A-Puro and the 2nd using the Cas9 recombinant protein. 96 iPSC clones were picked for further analysis for each patient iPSC line and for each targeting method.

For the R2 patient iPSC line, of 96 iPSC clones targeted using method 1, 68 clones (71%) were not targeted, as evidenced by the presence of both an intact wild-type and an intact mutant allele. Among the targeted clones (n=28), all contained indels (due to NHEJ pathway) near the Cas9 cleavage site either on the same allele or in the 2nd allele. Consequently, none of these clones were successfully repaired by HDR. For the R4 patient iPSC line, of 96 iPSC clones targeted using method 1, 74 clones (77%) were not targeted. Among targeted clones (n=22), all were repaired by NHEJ and contained various indels adjacent to the mutation site on the same

allele. Again, none of the clones were successfully repaired by HDR in the R2 patient patient iPSC line.

For both R2 and R4 patient iPSC lines, none of the 96 iPSC clones targeted using the Cas9 recombinant protein were successfully targeted. The mutation site in the HNF1B gene was not corrected and the clones did not contain any new mutations.

iPSC line	Genotype	Patient characteristics		
		Sex	**Age**	**Clinical features**
R1	c.1-?_1674+?del p.Met1_Trp557del	F	35-39	Mild hydronephrosis of right kidney and a simple cyst in the left kidney, no diabetes abnormal liver function tests, bicornate uterus
R2	c.1006dup p.His336fs	M	50-54	Single kidney, chronic renal disease stage 3, gout, low magnesium, no diabetes
R3	c.1006dup p.His336fs	M	20-24	Single kidney, renal transplant, gout, no diabetes, mixed dyslipidaemia
R4	c.826C>T p.Arg276*	F	5-9	Bilateral renal cysts, no diabetes
R5	c.1-?_1674+?del p.Met1_Trp557del	F	60-64	Bilateral multicystic kidneys, renal transplant, diabetes diagnosed at 56yrs, dyslipidaemia, hypertension, hypothyroidism, learning difficulties
R6	c.1566T>A p.Tyr522Unk	F	60-64	Bilateral renal cysts, right hydronephrosis, chronic renal disease stage 3-4, hypertension, dyslipidaemia, no diabetes
R7	gc.544+3_544+6del p.?	M	25-29	Renal and bladder dysgenesis, renal transplant, diabetes diagnosed aged 19yrs
R8	c.1048dup, p.Val350fs	F	50-54	Diabetes diagnosed at 13yrs, on insulin and tolbutamide, Single left kidney and agenesis of right kidney, renal transplant

Table 3.2. Genotype and phenotypic characteristics of 8 patients, from which fibroblasts and hiPSCs were successfully derived, with *HNF1B* mutations or deletions.
del = deletion; dup = duplication; fs = frameshift; p.? = protein has not been analysed
HbA1c = Glycated Haemoglobin; ALT = Alanine Transaminase; GGT = Gamma Glutamyl Transferase; ALP = Alkaline Phosphatase

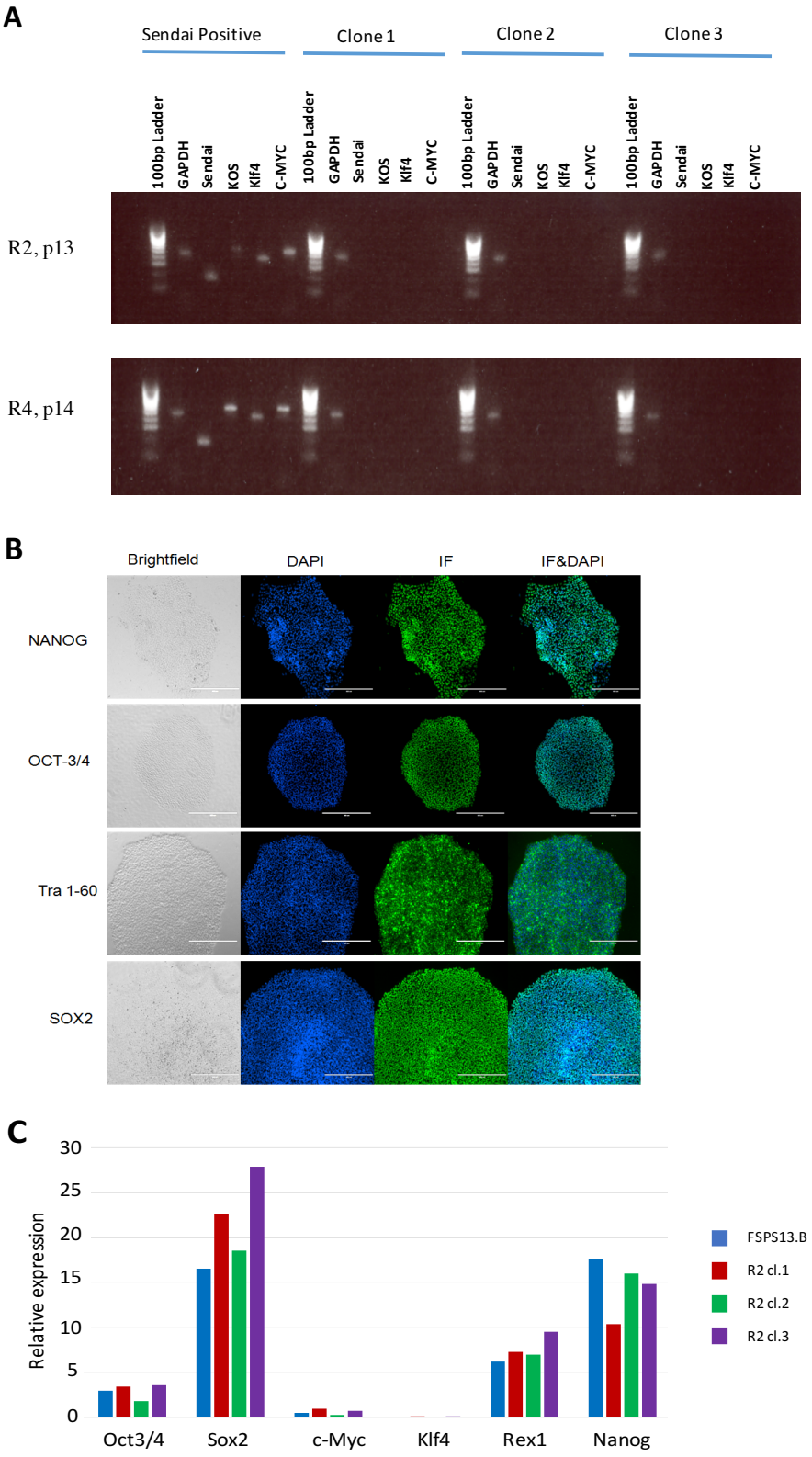


Figure 3.5. Characterization of the R2 and R4 patient derived hiPSC lines
(A) PCR showing absence of transgenes in all 3 clones for the R2 and R4 hiPSC lines
(B) Immunofluorescence showing expression of pluripotency markers in R2 clone 2 (representative images)
(C) qRT-PCR showing expression of pluripotency markers in a control hiPSC line (FSPS13.B) and R2 clones 1, 2 and 3.

3.3 Discussion

Several genome editing strategies are now available to knockout genes or generate small deletions, insertions or correct small mutations in hiPSCs and other cell types. Targeting efficiencies can vary widely depending on cell line, targeted locus, type of repair donor template and location of modification relative to the DSB site (Ran et al., 2013).

The targeting strategy used in this project was successful in generating homozygous and heterozygous HNF1B knockout hiPSCs. The advantages of this targeting strategy include; insertion of a puromycin resistance cassette which can be used to reliably select targeted clones and large disruption in the open reading frame of the gene in one or both alleles. The NHEJ strategy, which only involves cloning of gRNA sequences into a pSpCas9(BB)-2A-Puro vector, is more efficient compared with HDR, particularly for generating homozygous knockout hiPSCs. In most cases, NHEJ gives rise to small indels in the target DNA which result in in-frame amino acid deletions or insertions, or frameshift mutations leading to premature stop codons within the open reading frame (ORF) of the targeted gene. The end result is a loss-of-function mutation within the targeted gene; however, the "strength" of the knockout phenotype for a given mutant cell is ultimately determined by the amount of residual gene function and is unpredictable. Therefore, further validation of loss of the protein for example by Western Blot is required.

Two hiPSC lines with different genetic backgrounds were used to generate homozygous and heterozygous HNF1B knockout hiPSCs in order to ensure that any phenotype seen in subsequent experiments was due to the HNF1B genotype and not the genetic background of the hiPSCs. Thorough genotyping and characterisation of the $HNF1B^{+/+}$, $HNF1B^{+/-}$ and $HNF1B^{-/-}$ hiPSCs was performed to ensure that the cells had no chromosomal abnormalities and that there was no significant difference in the expression of pluripotency markers in the hiPSCs or in their capacity to differentiate to the 3 germ layers. Western blot and qRT-PCR were also used to check for reduction or absence of HNF1B protein in foregut progenitor stage in cells produced from $HNF1B^{+/-}$ and $HNF1B^{-/-}$ hiPSCs compared with $HNF1B^{+/+}$ cells.

Sanger sequencing was used to exclude mutations in the predicted off-target sites in three protein coding genes for the guide RNA used in these targeting experiments. However, in order to ensure that there no off-target effects in other genes or intronic regions, ideally, whole genome sequencing for each cell line should be performed. Chromosome instability is also

common in hiPSCs especially when cultured in E8 media therefore cells used for experiments were karyotyped approximately every 20 massages to ensure no chromosomal abnormalities had developed.

Correction of the mutations in the two patient iPSC lines (R2 and R4) was unsuccessful during the time-course of this project. This was likely due to the limited choice of efficient gRNAs in the region to be targeted. All the gRNAs with a reasonable score (>80) generated from the CRISPR design tool and a low number of off-target effects were located more than 100 bp from the patient mutations, resulting in high efficiency for inducing indels but significantly reduced efficiency for correction of the 1bp duplication in patient R2 and nonsense mutation in patient R4. It is estimated that single-base correction rates drop approximately fourfold at 100 bp away from the DSB site (Ran et al., 2013). For the 2nd targeting method using recombinant Cas9 protein, there was no repair of the targeted mutation and no new indels were introduced, suggesting either that transfection of the recombinant Cas9 protein or one of the other components did not occur, or rapid degradation of the Cas9 protein occurred prior to DNA cleavage. Further optimisation of these experiments using different gRNAs, repair templates and transfection methods is required.

In summary, genetic engineering of two hiPSC lines was successful in generating HNF1B heterozygous and homozygous knockout iPSCs with high efficiency. These cell lines were used in subsequent experiments to study the function of HNF1B in human pancreas development.

Chapter 4 Cellular phenotyping of HNF1B knockout iPSC lines

4.1 Introduction

Studies on human patient tissue samples are extremely challenging and have inherent technical and ethical limitations. Most research on human diseases is therefore based on model organisms and in vitro approaches. As previously described, mouse models are unsuitable to study haploinsufficient forms of monogenic diabetes resulting from mutations in the transcription factors HNF1A, HNF4A, HNF1B or GATA6, since they do not reproduce the phenotype seen in humans.

Human pluripotent stem cells (hPSCs), in combination with genome modification, offer a unique platform to study the genetic basis of diabetes and beta-cell dysfunction. Human PSCs can be easily derived by direct reprogramming of somatic cells using overexpression of transcription factors. The resulting cells can self-renew almost indefinitely in vitro while maintaining their capacity to differentiate into a diversity of cell types. Over the past decade, several in vitro directed differentiation protocols have been developed for making pancreatic cells, particularly insulin-producing β-cells. These PSC-derived pancreatic cells have been used to study monogenic diabetes phenotypes caused by a variety of genes, including HNF1B, GATA6, PDX1 and NEUROG3.

As described in the previous chapter, we have generated a panel of HNF1B heterozygous and homozygous knockout hiPSC lines by performing genome editing using the CRISPR-Cas9 system. We additionally derived hiPSC from two patients with HNF1B heterozygous mutations. Using a pancreatic differentiation protocol developed by our group, we analysed the normal expression pattern of HNF1B at each stage of the pancreatic differentiation and investigated the effect of HNF1B haploinsufficiency and homozygous loss of HNF1B on pancreatic development.

4.2 Results

4.2.1 Pattern of expression of HNF1B and other transcription factors throughout pancreatic differentiation

4.2.1.1 Directed pancreatic differentiation of hiPSCs

Directed differentiation of hiPSCs into pancreatic cells was undertaken, using a protocol developed in our laboratory (Figure 4.1.). This 27-day protocol is a revised version of an 18-day chemically-defined protocol previously published by our group (Cho et al., 2012) (Figure 4.1.). The differentiation protocol was optimised and characterised using two wild-type hiPSC lines; FSPS13.B and Eipl_1.

The differentiation process was characterized by studying the expression of relevant lineage markers at different stages of the pancreatic differentiation process using quantitative real-time PCR (qRT-PCR), immunocytochemistry and flow cytometry.

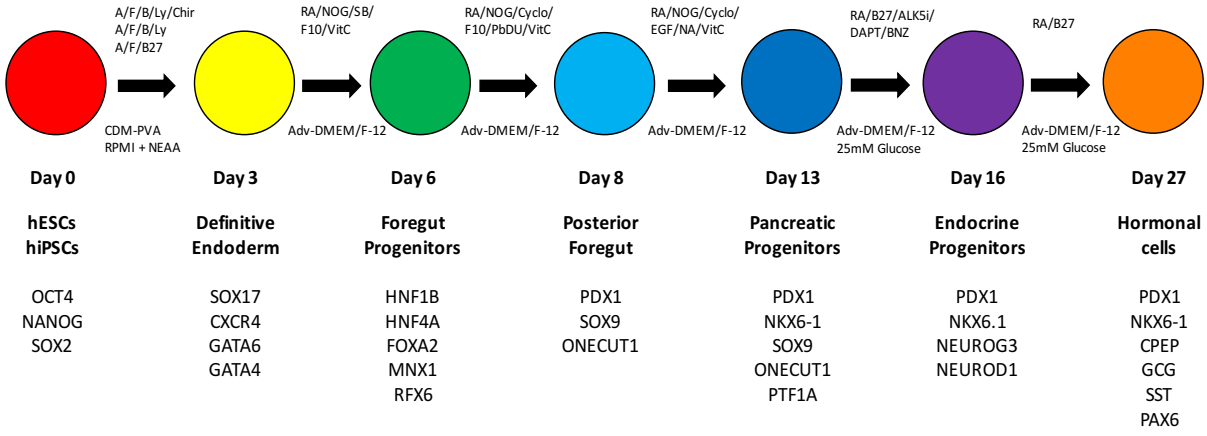


Figure 4.1. Schematic representation depicting the key developmental stages occurring in the embryo during pancreas formation and an overview of the protocol used to generate pancreatic progenitor and endocrine cells from human ESCs and iPSCs.
Successive culture conditions driving differentiation of pluripotent stem cells towards pancreatic endoderm and endocrine cells are shown.
A, activin A; F, fibroblast growth factor 2; B, bone morphogenetic protein; Ly, LY294002; Chir, Chir99021; B27, B-27 Supplement® (ThermoFisher Scientific, Waltham, MA, USA); RA; retinoic acid; NOG, noggin; SB; SB-431542; F10, fibroblast growth factor 10; VitC, Vitamin C; Cyclo, cyclopamine; PdBU, phorbol 12,13-dibutyrate; EGF, epidermal growth factor; NA, nicotinamide; Alk5i, TGFβ type I receptor kinase (Alk5) inhibitor; BNZ, 6-Benzoyladenosine-3',5'-cyclic monophosphate; DAPT, N-(N [3,5-diflurophenylacetyl]-L-alanyl)-S-phenylglycine t-butyl ester. CDM, chemically defined medium; PVA, Polyvinyl Alcohol; RPMI, Roswell Park Memorial Institute medium; Adv-DMEM/F-12; Advanced Dulbecco's Modified Eagle Medium/Ham's F-12.

Undifferentiated iPSCs expressed the pluripotency markers, NANOG, OCT4 and SOX2, and did not show any significant expression of endoderm (SOX17), mesoderm (Brachyury) or neuroectoderm (PAX6) lineage markers. In the first stage of the differentiation process (endoderm specification, days 1-3), resulting cells were homogenously positive for the expression of definitive endoderm markers including SOX17, CXCR4 and GATA6, whilst being negative for expression of pluripotency markers (NANOG, OCT4, SOX2) and foregut progenitor markers (HNF1B, HNF4A) (Figure 4.2A & B). Flow cytometry analysis showed that 85-95% of cells expressed SOX17 on day 3 of the differentiation (Figure 4.2C).

Following the second stage of the protocol (foregut induction, days 4-6), qPCR and immunostaining showed expression of HNF1B, FOXA2 and HNF4A on day 6, all of which are markers of primitive gut tube and foregut progenitors during early mammalian development. (Figure 4.2.).

In the third stage, the foregut progenitor cells were grown for 2 additional days (foregut patterning into pancreatic bud, days 7-8) to generate posterior foregut and early pancreatic progenitor cells. The resulting cells co-expressed a combination of posterior foregut markers (HNF1B, FOXA2, HNF4A) and pancreatic progenitor markers (SOX9, HNF6, and PDX1), however they were negative for the expression of NKX6-1.

In the fourth stage of the protocol (pancreatic progenitor specification, days 9-13), $PDX1^{+}NKX6\text{-}1^{-}$ early pancreatic progenitor cells were further differentiated into $PDX1^{+}/NKX6\text{-}1^{+}$ progenitor cells. Immunostaining analysis on the day 13 pancreatic progenitor cells confirmed that PDX1 was co-expressed in the same cells with NKX6-1 and SOX9. Flow cytometry analysis showed that on day 13, 90-95% of cells expressed PDX1, however, the percentage of cells co-expressing PDX1 and NKX6-1 was quite variable and ranged from around 20 to 70%. The percentage of $PDX1^{+}/NKX6\text{-}1^{+}$ cells could consistently be increased to greater than 95% by addition of higher concentrations of EGF and NA, however this did not improve the efficiency of generating endocrine cells or beta-like cells at day 16 and day 27 respectively.

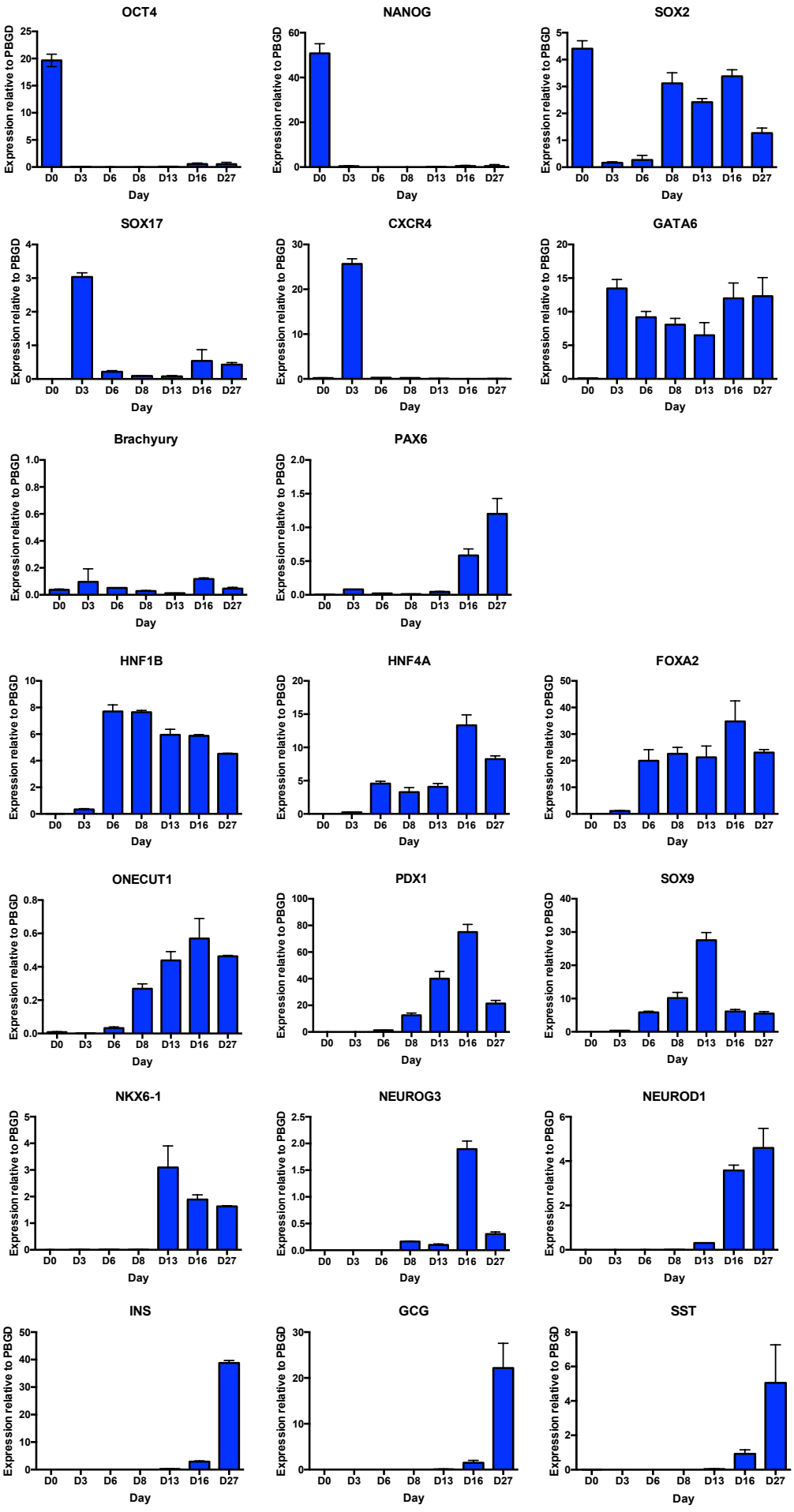
In the fifth stage of the protocol (specification of endocrine progenitors, days 14-16), pancreatic progenitor cells obtained at the end of stage 4 were grown for 3 additional days in culture conditions previously shown to stimulate their differentiation into endocrine progenitor cells. Expression of the endocrine progenitor cell markers NEUROG3 and NEUROD1 peaked at day

16. Flow cytometry analysis showed that around 8-10% of the cells expressed NEUROD1 on day 16.

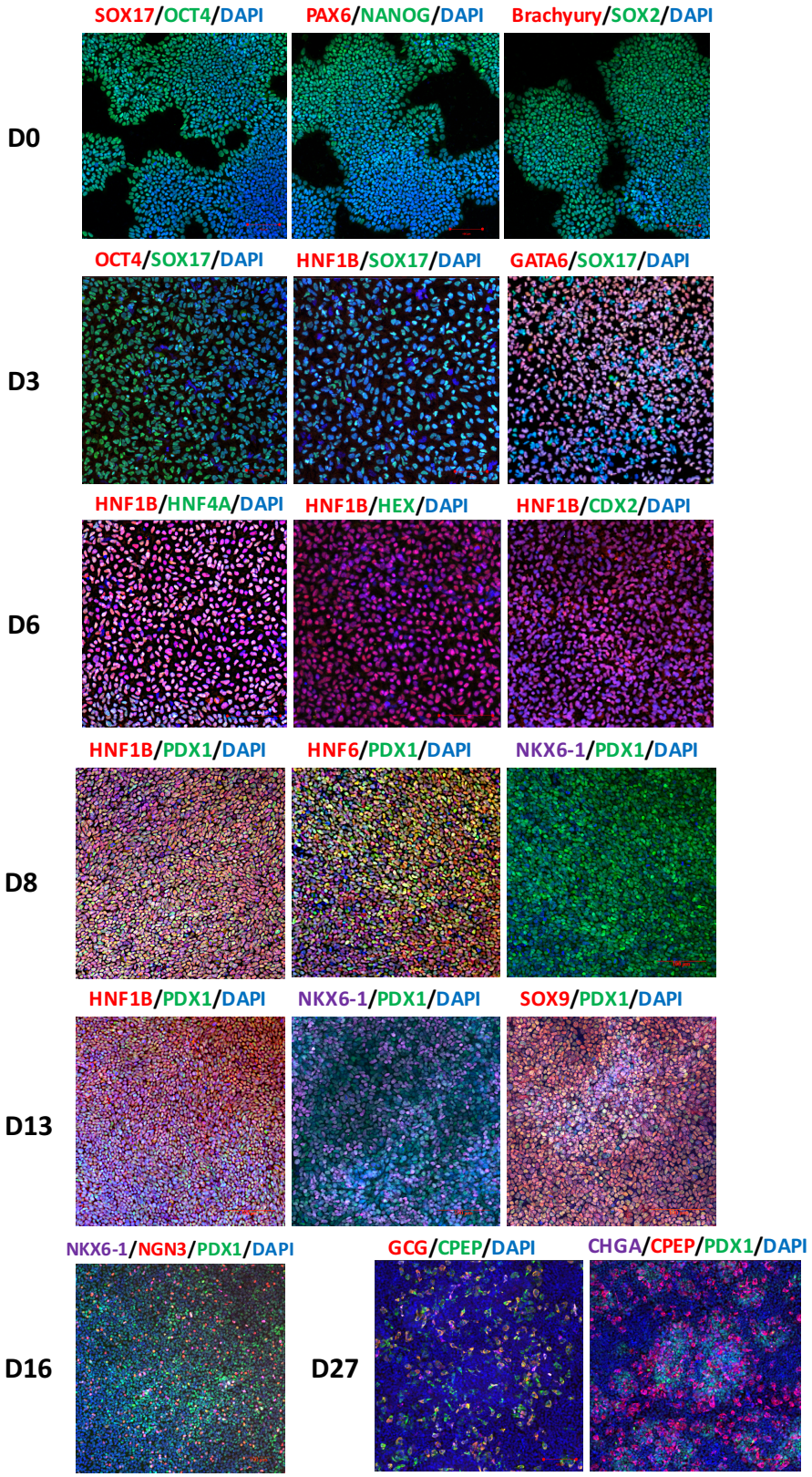
In the sixth stage of the protocol, endocrine progenitor cells were grown and matured for a further 11 days (hormonal cell production, days 16-27). Expression of pancreatic hormonal markers (chromogranin A, insulin, glucagon and somatostatin) was first detected at day 16 and significantly increased until day 27. By day 27, immunostaining was positive for insulin (INS) or C-peptide (CPEP) and to a lesser degree, glucagon (GCG) and somatostatin (SST). A fraction of cells showed multi-hormonal expression of CPEP and GCG and / or SST. To confirm these observations, flow cytometry analysis was performed. At day 27, around 7-8% of the cells expressed CPEP and around 50-60% of these cells were mono-hormonal (4-5% of total cells). Approximately 2-4% of the cells expressed GCG and less than 1% of cells expressed SST (Figure 4.2B).

Considered together, these data demonstrate that hiPSCs can be differentiated into pancreatic progenitor cells and endocrine cells, following a natural path of development.

A


OCT4
NANOG
SOX2
SOX17
CXCR4
GATA6
Brachyury
PAX6
HNF1B
HNF4A
FOXA2
ONECUT1
PDX1
SOX9
NKX6-1
NEUROG3
NEUROD1
INS
GCG
SST
Expression relative to PBGD
Day
D0 D3 D6 D8 D13 D16 D27

B


SOX17/OCT4/DAPI
PAX6/NANOG/DAPI
Brachyury/SOX2/DAPI
D0
OCT4/SOX17/DAPI
HNF1B/SOX17/DAPI
GATA6/SOX17/DAPI
D3
HNF1B/HNF4A/DAPI
HNF1B/HEX/DAPI
HNF1B/CDX2/DAPI
D6
HNF1B/PDX1/DAPI
HNF6/PDX1/DAPI
NKX6-1/PDX1/DAPI
D8
HNF1B/PDX1/DAPI
NKX6-1/PDX1/DAPI
SOX9/PDX1/DAPI
D13
NKX6-1/NGN3/PDX1/DAPI
GCG/CPEP/DAPI
CHGA/CPEP/PDX1/DAPI
D16
D27

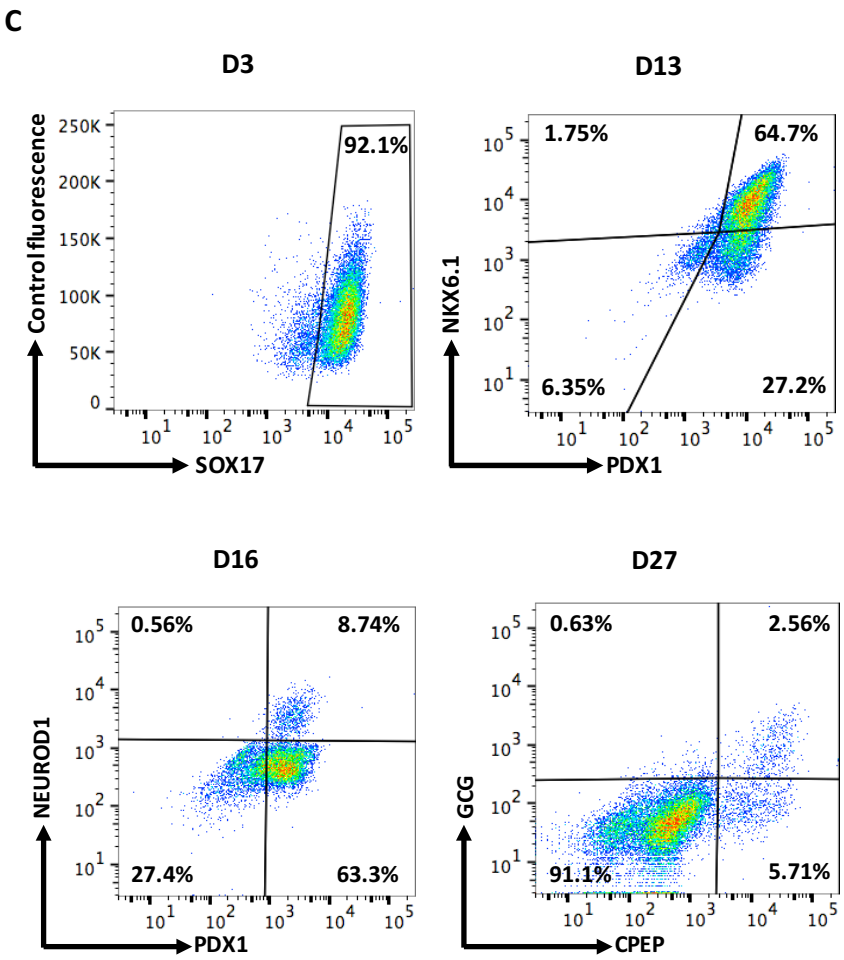


Figure 4.2. Characterization of the pancreatic differentiation process by studying the expression of relevant lineage markers at different stages.
(A) Expression of pluripotency markers (POU5F1, NANOG, SOX2), endoderm, mesoderm and neuroectoderm markers (SOX17, CXCR4, GATA6, Brachyury/T, PAX6), foregut progenitor markers (HNF1B, HNF4A, FOXA2), posterior foregut markers (HNF1B, FOXA2, HNF4A, PDX1, SOX9, ONECUT1), pancreatic progenitor markers (PDX1, SOX9, NKX6-1), endocrine progenitor markers (NEUROG3, NEUROD1, PDX1, NKX6-1) and hormonal cell markers (PDX1, NKX6-1, INS (or CPEP), GCG, SST, PAX6) during the differentiation of hiPSCs into hormonal cells using the 6-step protocol described in Figure 4.1. mRNA levels were measured by qRT-PCR (n = 4 at each stage of differentiation) and normalised to the housekeeping gene porphobilinogen deaminase (PBGD). Data are presented as mean ± SEM unless otherwise indicated.
(B) Representative immunostaining of HNF1B and other stage-specific markers, including pluripotency markers (OCT4, NANOG, SOX2), endoderm markers (GATA6, SOX17), foregut progenitor markers (HNF4A, HNF1B), posterior foregut and pancreatic progenitor markers (HNF6, PDX1, NKX6-1, SOX9) and endocrine progenitor (PDX1, NKX6-1, NEUROG3) and hormonal cell markers (CHGA, CPEP, GCG), Scale bar, 100 mm.
(C) Representative flow cytometry dot plots of cells stained for the stage-specific markers; endoderm; SOX17, pancreatic progenitors; PDX1/NKX6.1, endocrine progenitors; PDX1/NEUROD1 and hormonal cells; CPEP/GCG. The percentage of each cell population is indicated in the corresponding quadrant for all flow cytometry plots.

4.2.1.2 Functional analysis of pancreatic progenitors generated in vitro

Glucose metabolism is the most important physiological event that regulates insulin biosynthesis and secretion. Glucose is taken up into the β-cell via the glucose transporter (GLUT2 or SLC2A2) and phosphorylated by glucokinase (GCK). Further metabolism, especially in the mitochondria, results in generation of ATP at the expense of ADP. This leads to K_{ATP}-channel closure. The increased membrane resistance resulting from K_{ATP} channel closure allows calcium influx through plasmalemmal voltage-gated calcium channels and this increase in intracellular calcium triggers exocytosis of insulin granules.

The functional potential of the endocrine cells generated from hiPSCs was assessed by measuring C-peptide release in response to high and low glucose levels. In functional beta-cells, it is expected that C-peptide should be released in response to high glucose levels and suppressed at low glucose levels. At protein level, C-peptide was measured instead of insulin, as cells can take up insulin from the culture media, which would give false positive signals of insulin content and release in the cells. The proteolytic cleavage of proinsulin prior to secretion produces the insulin molecule and C-peptide. C-peptide is secreted in equimolar quantities to insulin; therefore C-peptide is widely used as an alternative measurement for insulin.

C-peptide-expressing cells derived from wild-type (WT) iPSCs (both FSPS13.B and Eipl_1 backgrounds) were able to release C-peptide in response to high glucose levels 22.5 mmol/l) and C-peptide release was suppressed at low glucose levels (2.25 mmol/l), for two rounds of stimulations (Figure 4.3.). A significant increase in the release of C-peptide was seen following incubation with potassium chloride (30 mmol/l KCl).

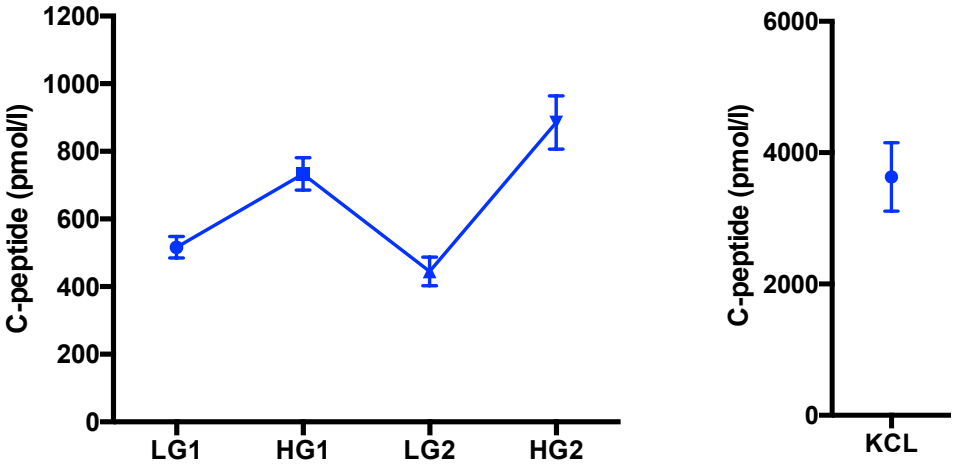


Figure 4.3. C-peptide secretion from hiPSC-derived hormonal cells (day 27). Cells were incubated in high-glucose (22.5 mmol/l) and low-glucose (2.25 mmmol/l) culture medium for 2 rounds of stimulations. Data are presented as the mean of 5 biological replicates and error bars indicate SEM.

4.2.2 HNF1B expression during directed differentiation of human iPSCs into the pancreatic lineage

The pattern of expression of HNF1B was analysed at each stage of the differentiation process in order to determine whether HNF1B followed the normal expression pattern seen during in vivo embryonic development. Consistent with previous studies, qRT-PCR analyses showed that HNF1B expression is significantly up-regulated during differentiation from definitive endoderm to the foregut progenitor stage, reaching a maximum on days 5 to 6 of our differentiation protocol. HNF1B expression persisted throughout the rest of the differentiation up to the hormonal cell stage (day 27) (Figure 4.4.). This is consistent with the known expression pattern and important role of HNF1B in foregut specification and pancreas development observed in mice and humans.

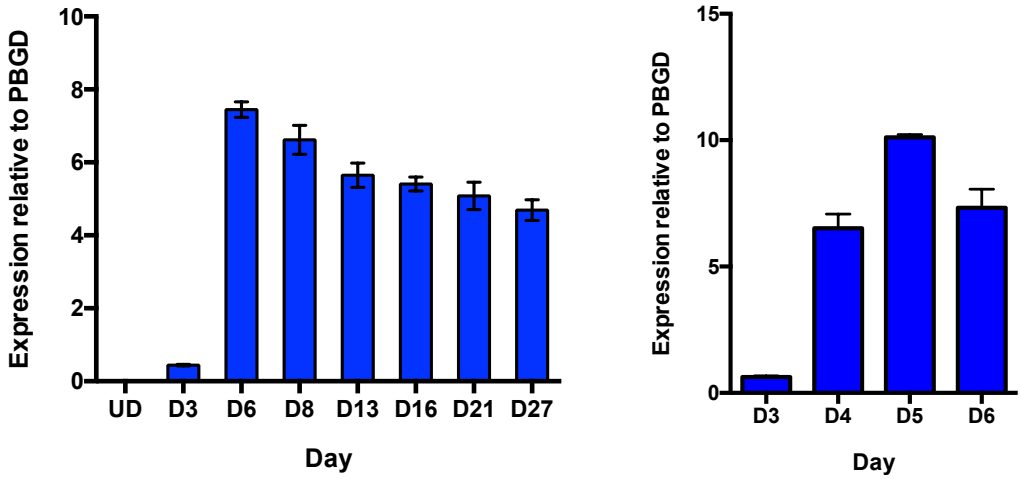


Figure 4.4. Expression pattern of HNF1B during hPSC differentiation (FSPS13.B WT hiPSC line). mRNA levels were measured by qRT-PCR (n = 4 at each stage of differentiation) and normalized to the housekeeping gene porphobilinogen deaminase (PBGD). Data are presented as mean ± SEM unless otherwise indicated.

4.2.3 HNF1B is required for efficient foregut and pancreatic progenitor formation

Pancreatic differentiations were performed using the protocol described previously for two hiPSC lines with different genetic backgrounds (FSPS13.B and Eipl_1). Three HNF1B wild-type, three heterozygous knockout and two homozygous knockout iPSC clones, for each of the hiPSC lines, were differentiated for each experiment. Differentiations were carried out between P73 (E8 43) and P98 (E8 68) for the FSPS13.B hiPSC lines and between P56 (E8 33) and P61 (E8 38) for the Eipl_1 hiPSC lines. At least 3 independent differentiations up to day 27 (hormonal cell stage) were carried out for each cell line. Wild-type, heterozygous and homozygous HNF1B mutant lines will be designated as $HNF1B^{+/+}$, $HNF1B^{+/-}$ and $HNF1B^{-/-}$ genotypes, respectively. HNF1B expression at each stage of the differentiation process was compared at the mRNA and protein level and was found to be reduced in $HNF1B^{+/-}$ mutant lines compared with $HNF1B^{+/+}$ lines and absent in the $HNF1B^{-/-}$ mutant line at each stage of the differentiation process.

$HNF1B^{+/+}$ and $HNF1B^{+/-}$ hiPSCs could be differentiated into DE cells expressing SOX17 and CXCR4 with comparable efficiencies, as determined by qRT-PCR and FACS analysis (Figure 4.5.). Of note, DE cells derived from $HNF1B^{-/-}$ mutant lines showed similar expression of SOX17 compared with $HNF1B^{+/+}$ and $HNF1B^{+/-}$ hiPSC lines however, qRT-PCR showed slightly but not statistically significant reduced expression of CXCR4.

$HNF1B^{-/-}$ mutant lines failed to differentiate to foregut progenitor ($HNF1B^{+}$, $HNF4A^{+}$) cells as determined by qRT-PCR and immunostaining. HNF1B and HNF4A expression levels were also significantly decreased in $HNF1B^{+/-}$ mutant lines compared with $HNF1B^{+/+}$ iPSCs (Figure 4.6.).

Subsequently, posterior foregut ($PDX1^{+}/NKX6\text{-}1^{-}$) and pancreatic progenitor ($PDX1^{+}/NKX6\text{-}1^{+}$) cells failed to form following differentiation of the $HNF1B^{-/-}$ mutant lines for both FSPS13.B and Eipl backgrounds, as determined by qRT-PCR, immunostaining and FACS analysis (Figure 4.7.). This is likely to be due to the earlier requirement for HNF1B at the foregut progenitor and posterior foregut stage (Figure 4.6.). At the pancreatic progenitor stage, HNF1B expression levels were decreased in $HNF1B^{+/-}$ mutant lines compared with $HNF1B^{+/+}$ iPSCs. $HNF1B^{+/-}$ mutant lines produced a lower number of $PDX1^{+}/NKX6\text{-}1^{+}$ pancreatic progenitor cells compared to $HNF1B^{+/+}$ lines, and qRT-PCR analysis showed a reduction in PDX1 and NKX6-1 mRNA expression. Other pancreatic progenitor markers, such as SOX9, PTF1A and HNF6

also showed statistically significant reduced expression levels between HNF1B$^{+/-}$ and HNF1B$^{+/+}$ cells. Similar results were obtained in the Eipl_1 background as there was a significant reduction of PDX1^{+}/NKX6-1^{+} pancreatic progenitor cells formed from HNF1B$^{+/-}$ hiPSCs (5-10% compared to 20-40% in HNF1B$^{+/+}$ cells).

These data demonstrate that HNF1B is essential for the efficient formation of posterior foregut. Thus, homozygous loss of HNF1B protein expression in the human embryo is likely to be lethal due to a primary defect in gut tube / foregut formation. HNF1B haploinsufficiency shows a milder phenotype with impairment of foregut and pancreas progenitor specification.

4.2.4 The effect of HNF1B haploinsufficiency on the formation of functional β-like cells

To investigate the functional consequences of HNF1B haploinsufficiency, we differentiated the cells further to endocrine progenitor (Figure 4.8A and B) and then hormonal or β-like cells (Figure 4.9A, B and C). Immunostaining, FACS analyses and qRT-PCR showed that HNF1B$^{-/-}$ mutant lines failed to form any endocrine progenitor or hormonal cells. Expression of endocrine progenitor cell markers such as NKX6.1, NEUROD1 and GLIS3 was reduced in HNF1B$^{+/-}$ compared with HNF1B$^{+/+}$ cells. HNF1B$^{+/-}$ hiPSCs formed CPEP^{+} monohormonal β-like cells, but the percentage was greatly reduced compared to HNF1B$^{+/+}$ cells (2-3% versus 8%). Consistent with this finding, HNF1B$^{+/-}$ cells showed reduced expression of all endocrine markers, including genes encoding key transcription factors and endocrine hormones important for β cell function. Approximately 50% of CPEP^{+} cells were monohormonal for both HNF1B$^{+/+}$ and HNF1B$^{+/-}$ cells. However, for the HNF1B$^{+/-}$ mutant clones, fewer CPEP^{+} cells co-expressed NKX6-1, which is known to play important roles in maintaining adult β cell function (Taylor et al., 2013).

Functional assays on hPSC-derived β-like cells showed that HNF1B$^{+/-}$ cells exhibited reduced glucose-stimulated insulin secretion (GSIS) compared to HNF1B$^{+/+}$ cells (Figure 4.10.). The total concentration of C-peptide secreted was reduced, but not the ratio of C-peptide secreted in high glucose (22.5 mM) to low glucose (2.25 mM), after correcting for the reduced number of CPEP^{+} β-like cells in HNF1B$^{+/-}$ mutant cells (ratio 1.99 vs. 1.91, p=0.41) This was seen for both FSPS13.B and Eipl_1 hiPSCs. Eipl_1 iPSCs produced reduced numbers of CPEP^{+} cells and GSIS compared with FSPS13.B iPSCs (data not shown).

These data confirm that the absence of HNF1B entirely blocks pancreatic development. On the

other hand, a decrease in HNF1B expression affects the efficiency of differentiation to pancreatic progenitor cells without entirely inhibiting hormonal cell production and the functionality of the β-like cells.

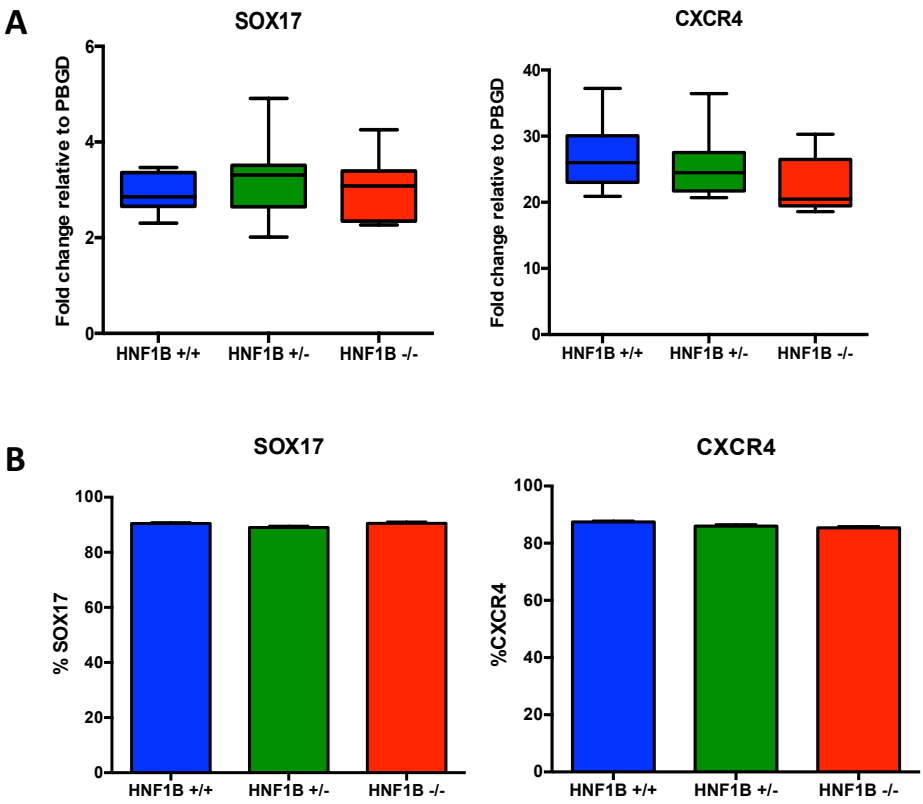


Figure 4.5. Differentiation of HNF1B$^{+/+}$ and HNF1B$^{+/-}$ and HNF1B$^{-/-}$ hiPSCs to produce definitive endoderm (DE) cells (Day 3).
(A) Expression of SOX17 and CXCR4 in DE cells derived from HNF1B$^{+/+}$, HNF1B$^{+/-}$ and HNF1B$^{-/-}$ hiPSC lines. The mRNA levels were measured by qRT-PCR and normalized to the house-keeping gene PBGD.
(B) FACS analysis of cells stained for the DE markers SOX17 and CXCR4 (n=5 independent experiments). There was no significant difference in the number of cells staining for SOX17 or CXCR4 in DE cells derived from HNF1B$^{+/+}$, HNF1B$^{+/-}$ and HNF1B$^{-/-}$ hiPSC lines.
n = 5 independent experiments. Student's t test with two-tailed distribution was used for statistical analysis. All data are presented as mean ± SEM unless otherwise indicated. P-values were not significant.

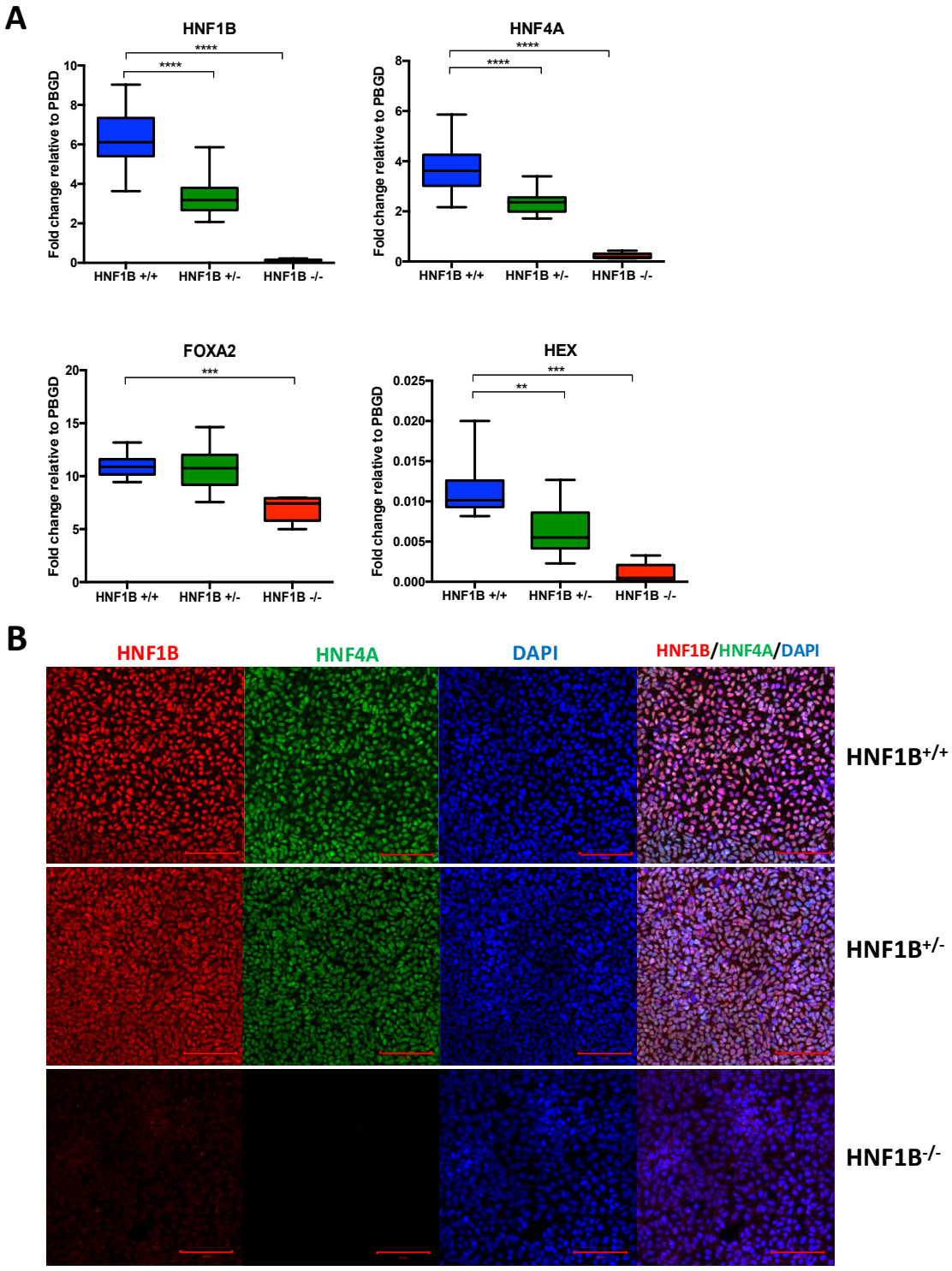
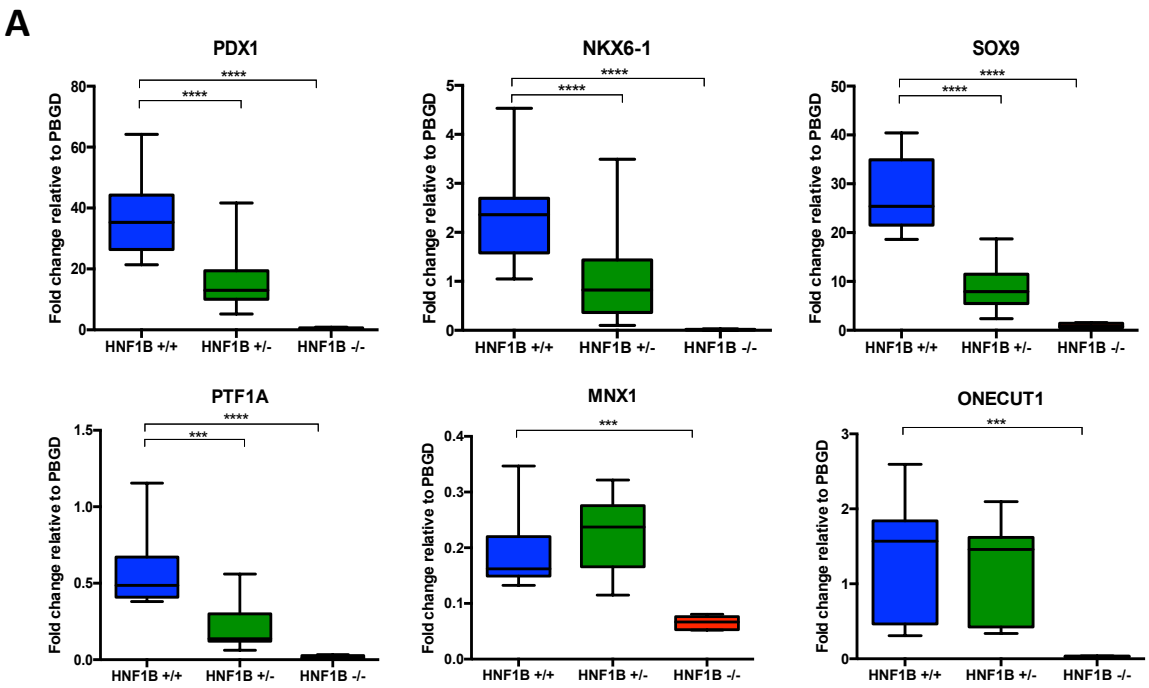


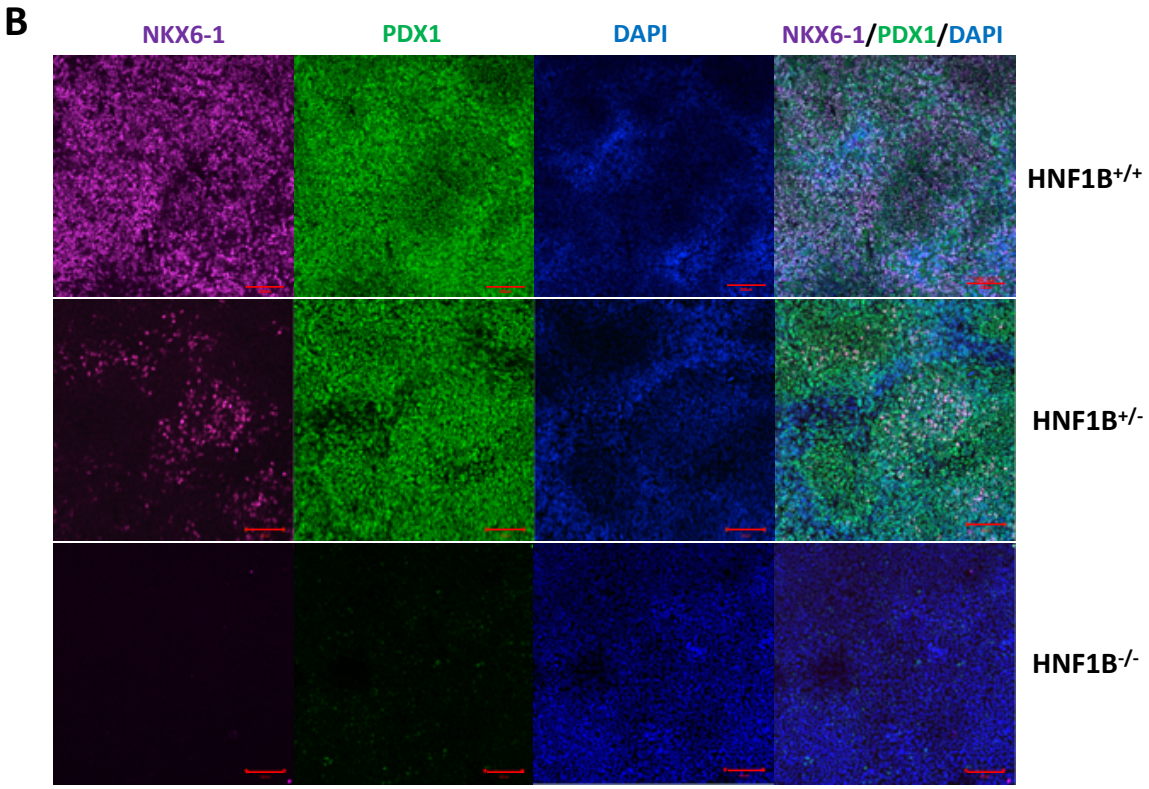
Figure 4.6. Differentiation of HNF1B$^{+/+}$ and HNF1B$^{+/-}$ and HNF1B$^{-/-}$ hiPSCs to produce foregut progenitor (FP) cells (Day 6).

(A) Expression of HNF1B, HNF4A, FOXA2 and HHEX in FP cells derived from HNF1B$^{+/+}$, HNF1B$^{+/-}$ and HNF1B$^{-/-}$ hiPSC lines. The mRNA levels were measured by qRT-PCR and normalized to the house-keeping gene PBGD.

(B) Representative immunofluorescence images showing HNF1B$^{+/+}$, HNF1B$^{+/-}$ and HNF1B$^{-/-}$ mutant cells at the foregut progenitor stage of differentiation. Scale bar – 100µm

n= 5 independent experiments. Student's t test with two-tailed distribution was used for statistical analysis. All data are presented as mean ± SEM unless otherwise indicated. *p < 0.05; **p < 0.01; ***p < 0.001 and ****p < 0.0001.


A
PDX1
NKX6-1
SOX9
PTF1A
MNX1
ONECUT1
Fold change relative to PBGD
HNF1B +/+
HNF1B +/-
HNF1B -/-
****
***


B
NKX6-1
PDX1
DAPI
NKX6-1/PDX1/DAPI
HNF1B+/+
HNF1B+/-
HNF1B-/-

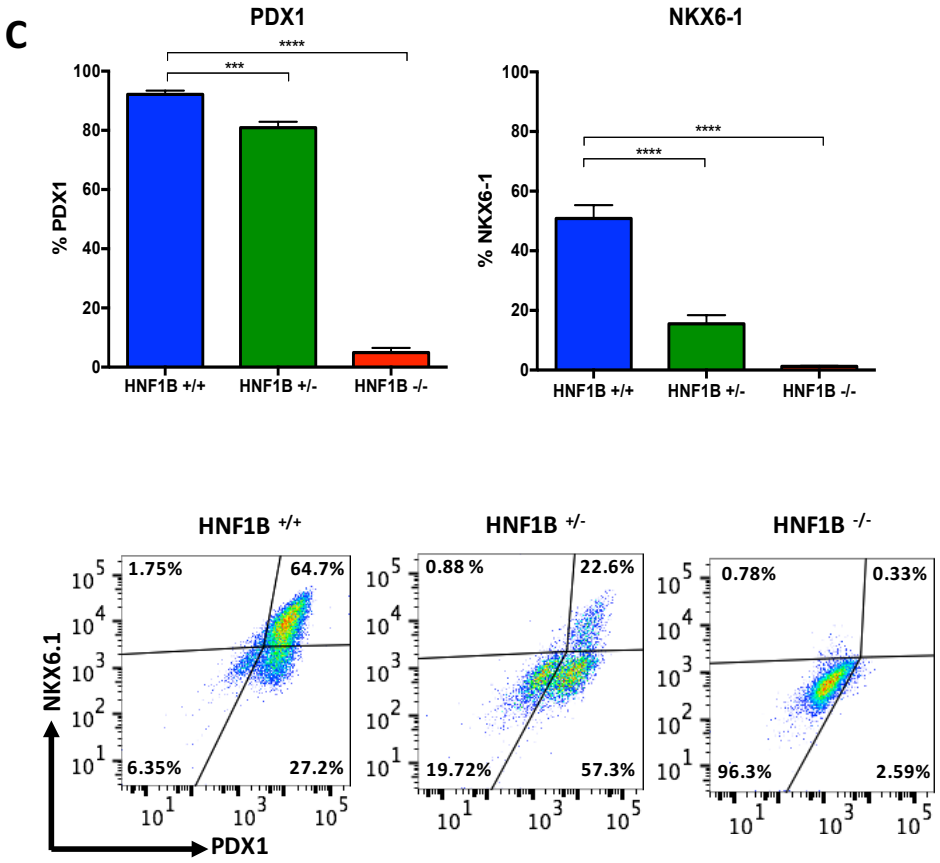


Figure 4.7. Differentiation of HNF1B$^{+/+}$ and HNF1B$^{+/-}$ and HNF1B$^{-/-}$ hiPSCs to produce pancreatic progenitor (PP) cells (Day 13).
(A) Expression of PDX1, NKX6-1, SOX9, PTF1A, MNX1 and HNF6 in PP cells derived from HNF1B$^{+/+}$, HNF1B$^{+/-}$ and HNF1B$^{-/-}$ hiPSC lines. mRNA levels were measured by qRT-PCR and normalized to the house-keeping gene PBGD.
(B) Representative immunofluorescence images showing PDX1 and NKX6-1 co-expression at the pancreatic progenitor stage of differentiation in HNF1B$^{+/+}$, HNF1B$^{+/-}$ and HNF1B$^{-/-}$ cells. Scale bar - 100µm
(C) Percentage of cells expressing PDX1 and NKX6-1 and representative FACS dot plots of cells stained for PDX1 and NKX6-1. The percentage of each cell population is indicated in the corresponding quadrant for all FACS plots.
n= 5 independent experiments. Student's t test with two-tailed distribution was used for statistical analysis. All data are presented as mean ± SEM unless otherwise indicated. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$ and **** $p < 0.0001$.

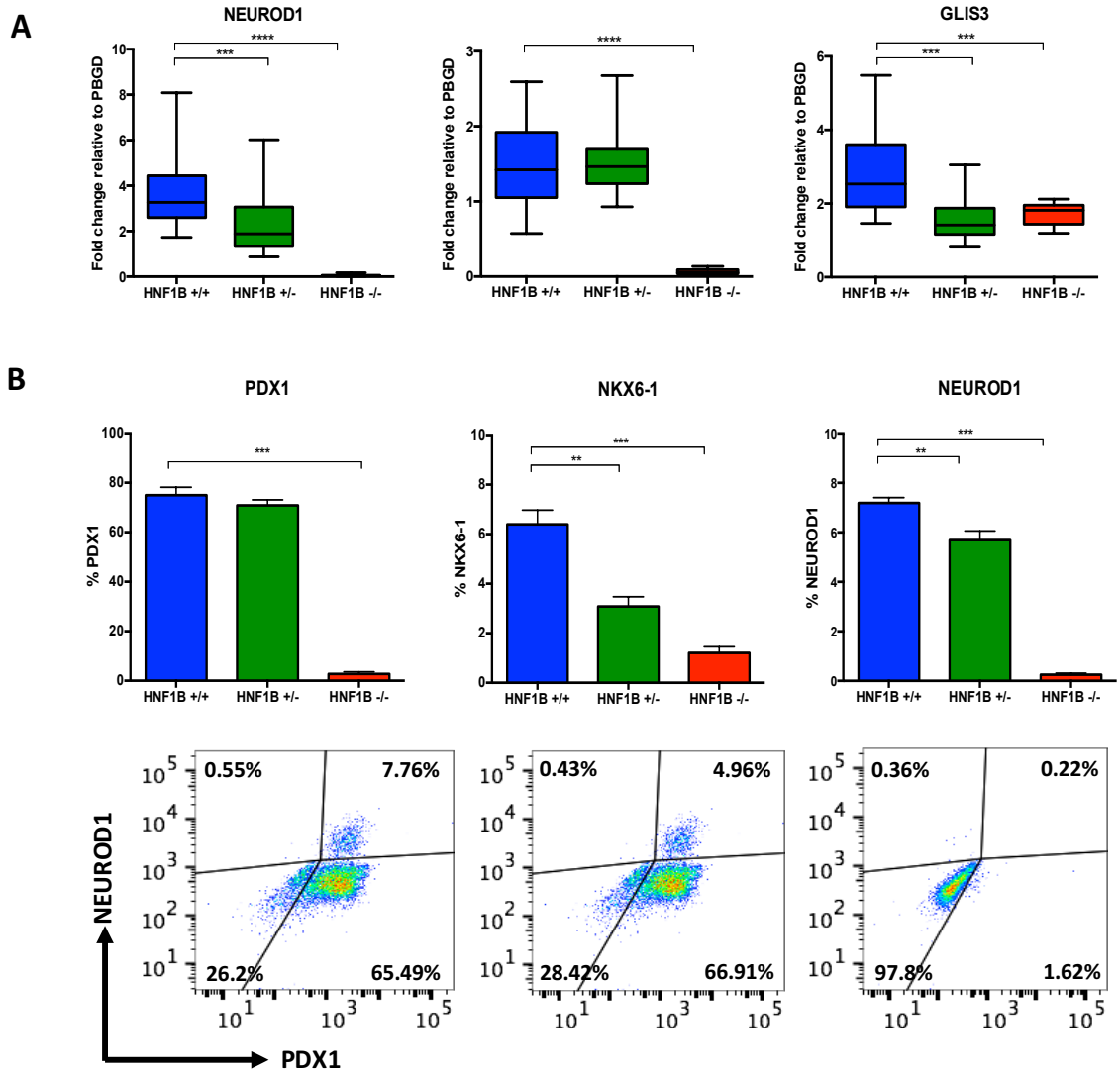


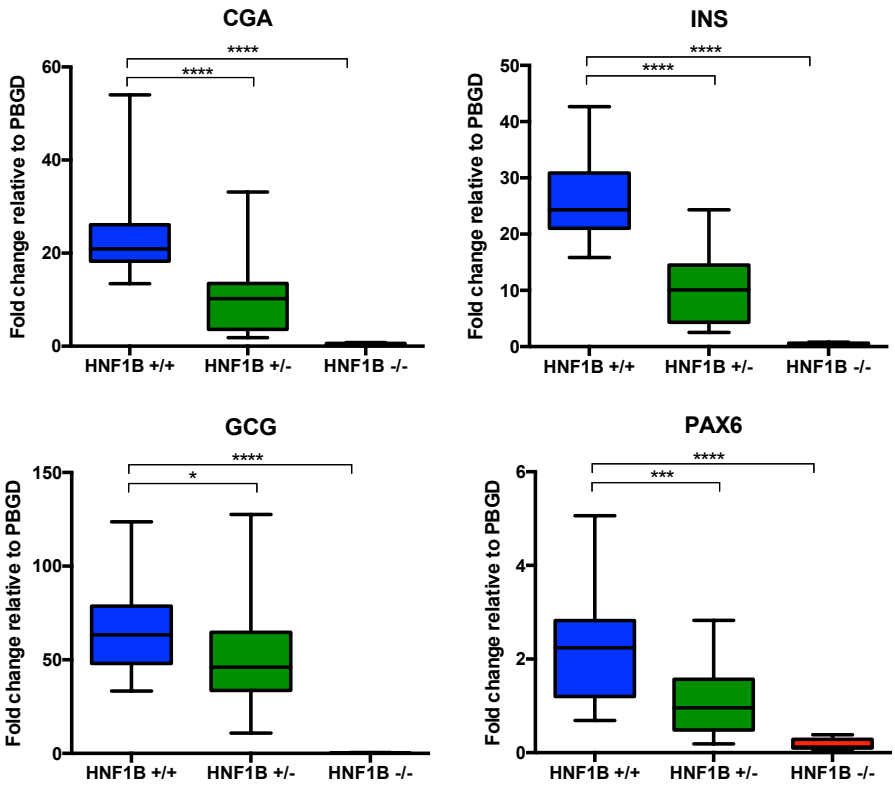
Figure 4.8. Differentiation of HNF1B$^{+/+}$ and HNF1B$^{+/-}$ and HNF1B$^{-/-}$ hiPSCs to produce endocrine progenitor (EP) cells (Day 16).

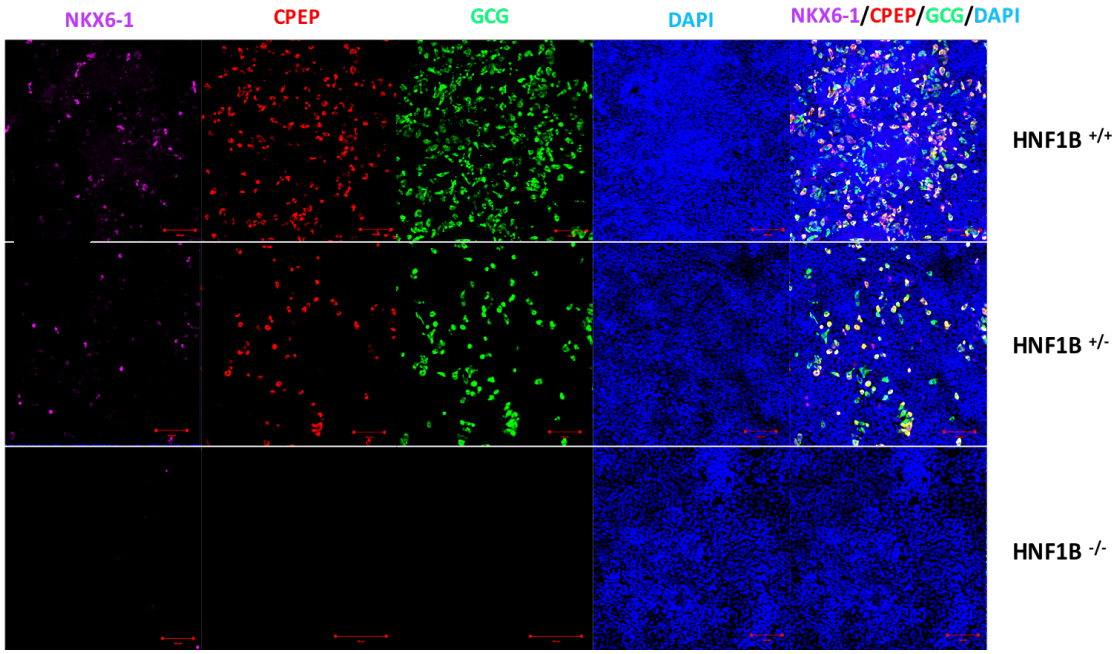
(A) Expression of GLIS3, NEUROD1 and NEUROG3 in PP cells derived from HNF1B$^{+/+}$, HNF1B$^{+/-}$ and HNF1B$^{-/-}$ hiPSC lines. mRNA levels were measured by qRT-PCR and normalized to the house-keeping gene PBGD.

(B) Percentage of cells expressing PDX1, NKX6-1 and NEUROD1 and representative FACS dot plots of cells stained for PDX1 and NEUROD1. The percentage of each cell population is indicated in the corresponding quadrant for all FACS plots.

n = 5 independent experiments Student's t test with two-tailed distribution was used for statistical analysis. All data are presented as mean ± SEM unless otherwise indicated. *p < 0.05; **p < 0.01; ***p < 0.001 and **** p < 0.0001.

A


CGA
INS
GCG
PAX6
Fold change relative to PBGD
HNF1B +/+
HNF1B +/-
HNF1B -/-
****
***
*


NKX6-1
CPEP
GCG
DAPI
NKX6-1/CPEP/GCG/DAPI
HNF1B +/+
HNF1B +/-
HNF1B -/-

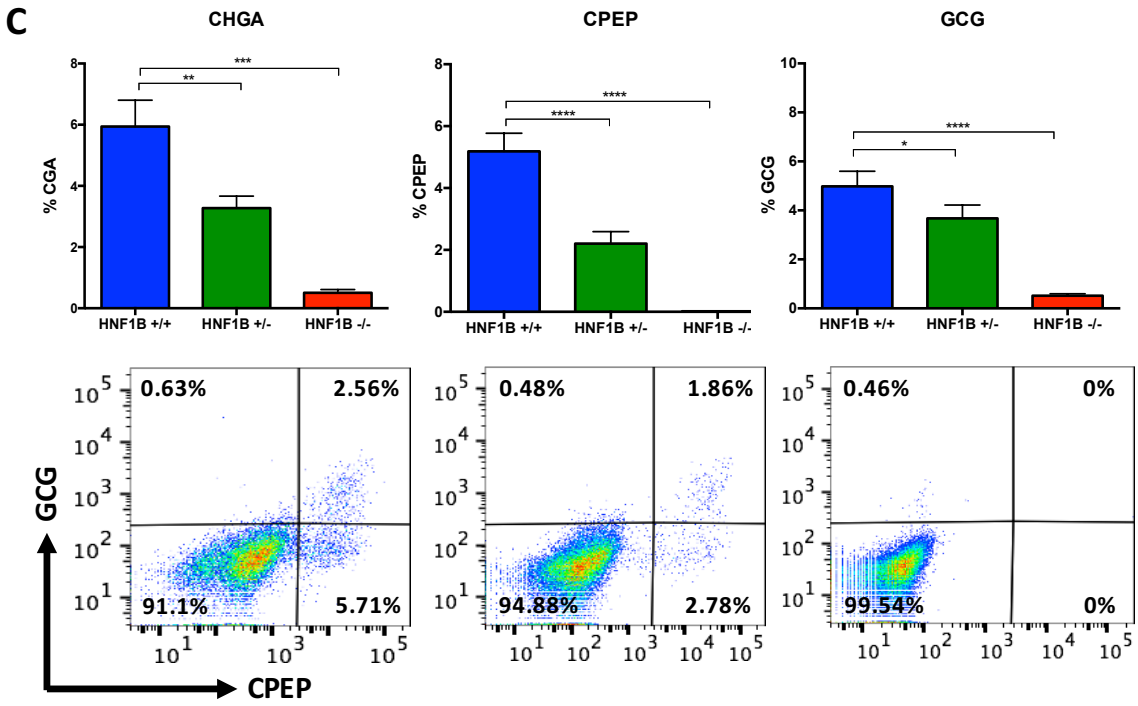


Figure 4.9. Differentiation of HNF1B$^{+/+}$ and HNF1B$^{+/-}$ and HNF1B$^{-/-}$ hiPSCs to produce hormonal β-like cells (βLC; Day 27).

(A) Expression of CHGA, INS, GCG and PAX6 in βLCs derived from HNF1B$^{+/+}$, HNF1B$^{+/-}$ and HNF1B$^{-/-}$ hiPSC lines. mRNA levels were measured by qRT-PCR and normalized to the house-keeping gene PBGD.

(B) Representative immunofluorescence images showing NKX6-1, CPEP and GCG co-expression at the βLC stage of differentiation in HNF1B$^{+/+}$, HNF1B$^{+/-}$ and HNF1B$^{-/-}$ cells. Scale bar - 100μm

(C) Percentage of cells expressing CHGA, CPEP and GCG and representative FACS dot plots of cells stained for CPEP and GCG. The percentage of each cell population is indicated in the corresponding quadrant for all FACS plots.

n = 5 independent experiments. Student's t test with two-tailed distribution was used for statistical analysis. All data are presented as mean ± SEM unless otherwise indicated. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$ and **** $p < 0.0001$.

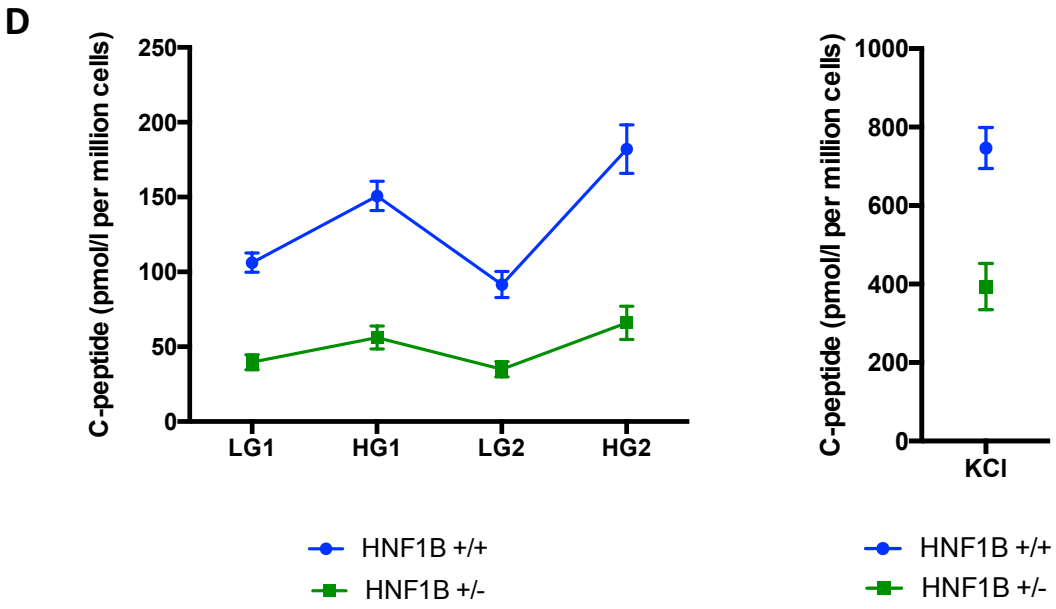


Figure 4.10. C-peptide secretion from β-like cells (day 27) derived from HNF1B$^{+/+}$ and HNF1B$^{+/-}$ hiPSCs. Cells were incubated in high-glucose (22.5 mmol/l) and low-glucose (2.25 mmol/l) culture medium for 2 rounds of stimulations. Data are presented as average of 5 biological replicates and error bars indicate SEM.

4.2.5 The effect of HNF1B genotype on cell apoptosis and proliferation

During the course of these experiments, we observed that the number of cells generated at the pancreatic progenitor stage was lower in HNF1B mutant cell lines when compared to wild type for both the FSPS13.B and Eipl_1 hiPSC lines. In agreement, the number of cells harvested, from 1 well of a 12-well plate was significantly lower in $HNF1B^{+/-}$ and $HNF1B^{-/-}$ compared with $HNF1B^{+/+}$ cells, starting at the foregut progenitor stage ($2.37x10^6$ vs $2.2x10^6$ vs $1.46x10^6$ cells, $p=0.24$, $p<0.05$). This difference was greatest at the pancreatic progenitor stage (day 13) ($7.33x10^6$ vs $5.38x10^6$ vs $3.78x10^6$, $p<0.05$, $p<0.05$) (Figure 4.11A). We therefore hypothesised that increased cell death and/or decreased proliferation could be the cause of the decrease in the number of cells in the $HNF1B^{+/-}$ and $HNF1B^{-/-}$ mutant lines. The following experiments were performed using the FSPS13.B hiPSC lines and 3 independent experiments were performed.

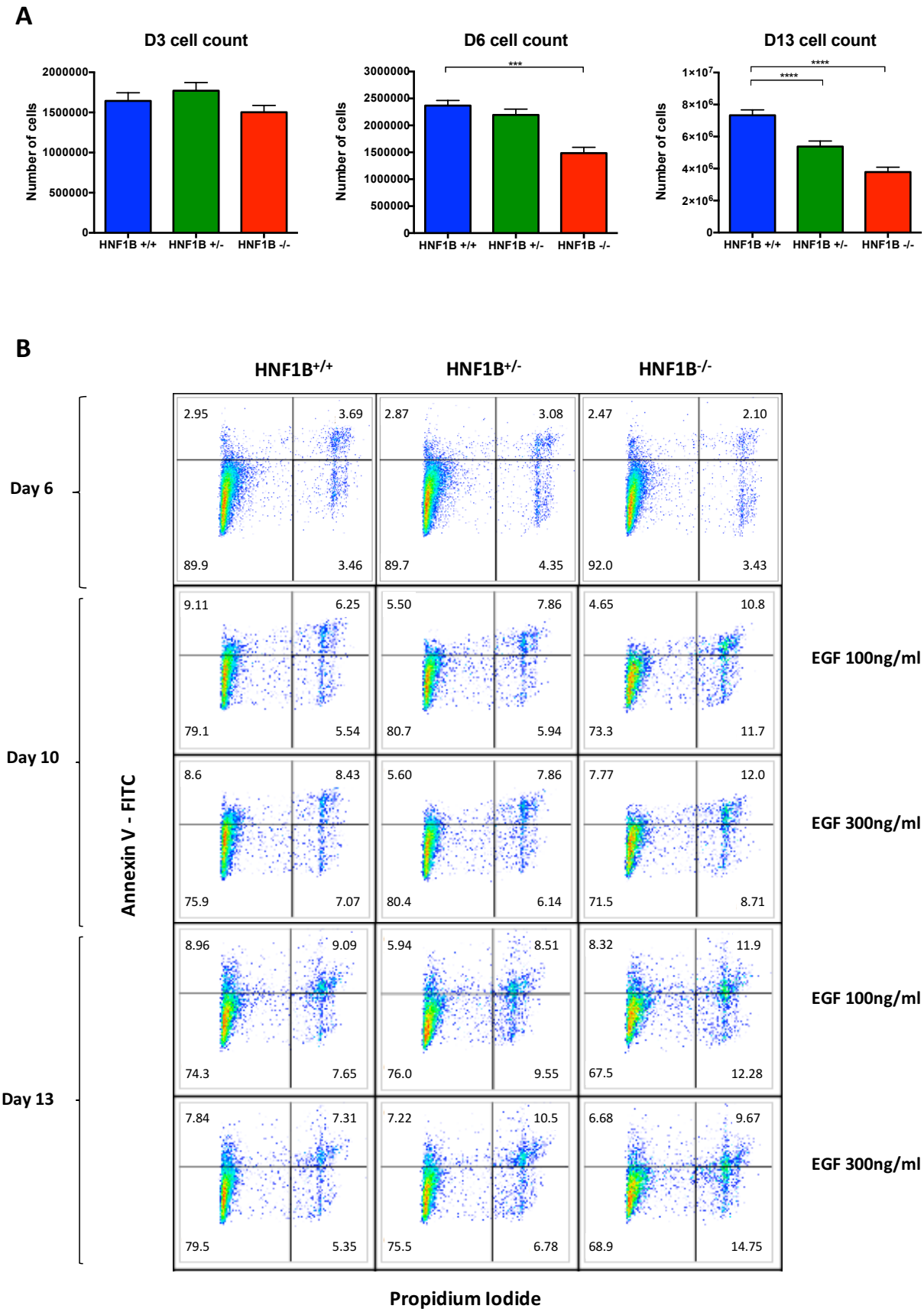
To determine whether increased cell death was the cause for the decrease in cell number, propidium iodide (PI) and Annexin V staining was performed at day 6, 10 and 13 of differentiation. No significant difference in the number of cells in early or late apoptosis was observed between $HNF1B^{+/+}$, $HNF1B^{+/-}$ and $HNF1B^{-/-}$ cell lines (n=3 independent experiments) (Figure 4.11B and Table 4.1A).

To determine whether the decrease in cell number in $HNF1B^{+/-}$ and $HNF1B^{-/-}$ mutant lines was due to reduced cell proliferation, DNA synthesis and DNA content analysis using EdU and FxCycle Far Red staining respectively, was performed at day 6, 10 and 13 of differentiation (Figure 4.11C and Table 4.1B). EdU staining showed a significant reduction in cell proliferation in $HNF1B^{+/-}$ and $HNF1B^{-/-}$ cells at day 6 (i.e. the foregut progenitor stage when HNF1B expression is up-regulated), compared with $HNF1B^{+/+}$ cells. There was a significant decrease in the number of cells in S-phase (EdU^+ cells), in $HNF1B^{+/-}$ and $HNF1B^{-/-}$ cells compared with $HNF1B^{+/+}$ cells at day 6 and in $HNF1B^{+/-}$ compared with $HNF1B^{+/+}$ cells at days 10 and 13. Correspondingly, the percentage of cells at G1 and G2/M phase increased as non-proliferating cells accumulated at these stages. At day 10 and day 13 there was no consistent decrease in proliferation in the $HNF1B^{-/-}$ cells suggesting that by this stage $HNF1B^{-/-}$ cells have a completely different identity and respond differently to external stimuli.

During embryogenesis, development of pancreatic tissue requires the proliferation of progenitor cells, followed by their differentiation into endocrine or acinar cells. epidermal growth factor (EGF) stimulates proliferation of pancreatic progenitor cells resulting in

increased PDX1^{+}/NKX6.1^{+} cells. A study by Verdeguer at al. (Verdeguer et al., 2010) showed that Hnf1b only caused polycystic kidney disease if it was inactivated (haploinsufficient) in rapidly proliferating tubular cells due to its role as a bookmarking factor. In order to investigate if a similar effect is seen in human pancreas differentiation, we increased the concentration of EGF, from 100ng/ml to 300ng/ml, to induce cells to proliferate more rapidly and to see if this exacerbates the phenotype. Higher concentrations of EGF resulted in an increase in the number of proliferating or EdU positive cells in both HNF1B$^{+/+}$ and HNF1B$^{+/-}$ cell lines. At each concentration however, the percentage of proliferating cells was greater in HNF1B$^{+/+}$ compared with HNF1B$^{+/-}$ cell lines and there was no exacerbation of the phenotype in HNF1B$^{+/-}$ cells at the higher concentration of EGF.

Taken together, these results suggest that loss of one functional HNF1B allele results in impairment of cell proliferation which impairs the production of pancreatic progenitor cells from foregut cells. This subsequently will result in the reduced production of CPEP^{+} β-like cells.


A
D3 cell count
D6 cell count
D13 cell count
Number of cells
HNF1B +/+
HNF1B +/-
HNF1B -/-
***
****
B
HNF1B+/+
HNF1B+/-
HNF1B-/-
Day 6
Day 10
Day 13
EGF 100ng/ml
EGF 300ng/ml
Annexin V - FITC
Propidium Iodide
2.95
3.69
89.9
3.46
2.87
3.08
89.7
4.35
2.47
2.10
92.0
3.43
9.11
6.25
79.1
5.54
5.50
7.86
80.7
5.94
4.65
10.8
73.3
11.7
8.6
8.43
75.9
7.07
5.60
7.86
80.4
6.14
7.77
12.0
71.5
8.71
8.96
9.09
74.3
7.65
5.94
8.51
76.0
9.55
8.32
11.9
67.5
12.28
7.84
7.31
79.5
5.35
7.22
10.5
75.5
6.78
6.68
9.67
68.9
14.75

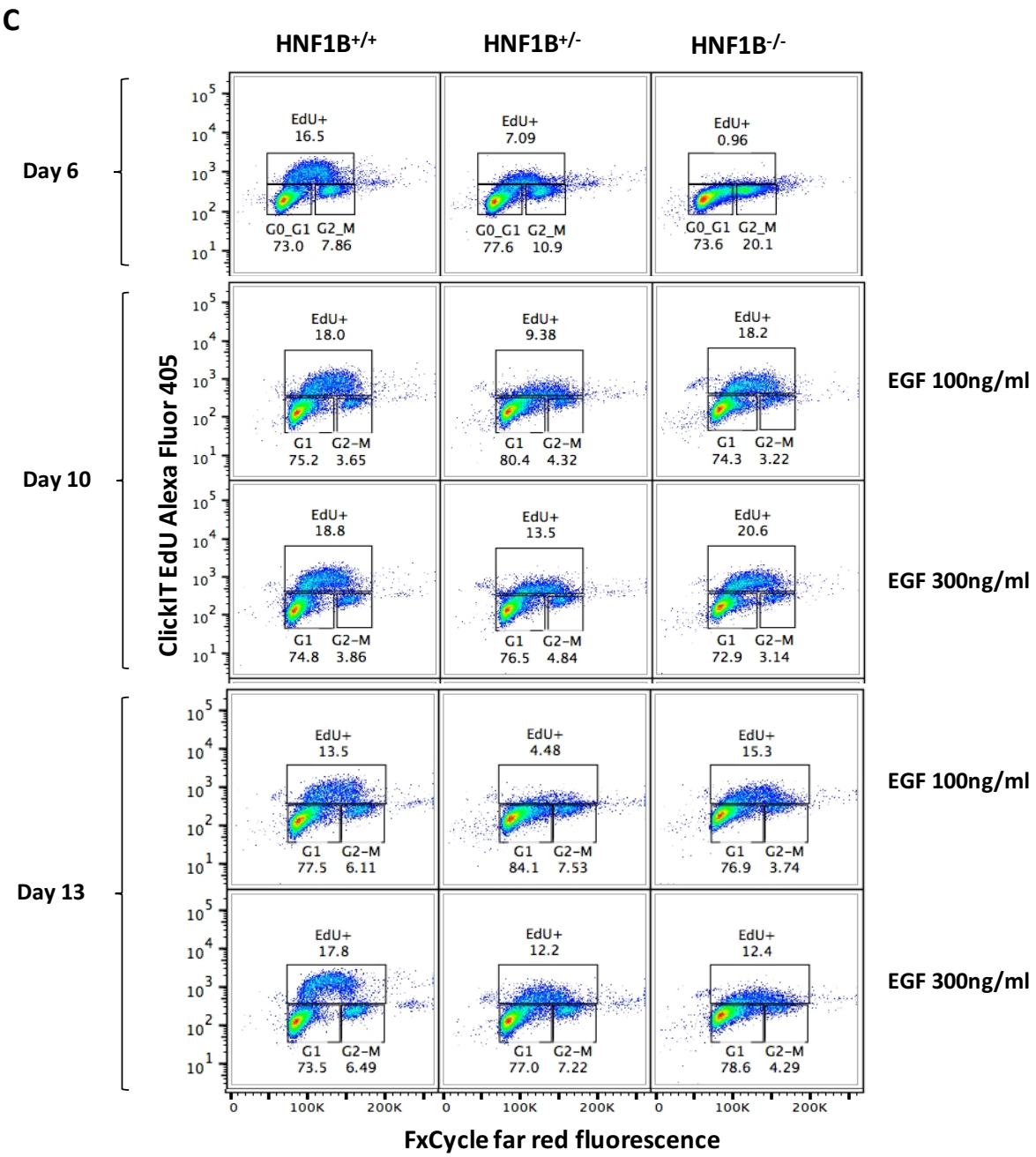


Figure 4.11. Effect of HNF1B genotype on apoptosis and cell proliferation from the foregut progenitor to the pancreatic progenitor stages of differentiation.
(A) Cell count determined by number of cells harvested, from 1 well of a 12-well plate in HNF1B$^{+/-}$, HNF1B$^{+/-}$ and HNF1B$^{-/-}$ cells at the definitive endoderm (day 3), foregut progenitor stage (day 6) and pancreatic progenitor stage (day 13) of differentiation. n=3 independent experiments. Student's t test with two-tailed distribution was used for statistical analysis. Data is presented as mean ± SEM unless otherwise indicated. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$ and **** $p < 0.0001$. n=3 independent experiments.
(B) FACS plots showing Annexin V and PI staining and the percentage of cells in early (Annexin V^{+}/PI^{-}) and late apoptosis / necrosis (Annexin V^{+}/PI^{+}) for HNF1B$^{+/+}$, HNF1B$^{+/-}$ and HNF1B$^{-/-}$ genotypes at days 6, 10 and 13 for one representative experiment.
(C) FACS plots showing EdU staining (EdU^{+} cells = percentage of cells in S-phase) and the percentage of cells in G1, S and G2M phase for HNF1B$^{+/+}$, HNF1B$^{+/-}$ and HNF1B$^{-/-}$ genotypes at days 6, 10 and 13 for one representative experiment.

A

		HNF1B$^{+/+}$	HNF1B$^{+/-}$	HNF1B$^{-/-}$
Day 6	-	7.81 (1.37)	7.00 (1.45)	4.75 (1.20)
Day 10	**EGF 100ng/ml**	14.56 (1.09)	13.31 (0.61)	15.69 (1.27)
	EGF 300ng/ml	15.02 (1.74)	14.04 (1.62)	20.34 (2.73)
Day 13	**EGF 100ng/ml**	18.05 (1.96)	14.70 (2.09)	20.67 (2.81)
	EGF 300ng/ml	15.65 (1.13)	16.68 (3.33)	16.56 (1.72)

B

		HNF1B$^{+/+}$	HNF1B$^{+/-}$	HNF1B$^{-/-}$
Day 6	-	14.20 (1.67)	6.39 (1.84)	1.95 (0.75)
Day 10	**EGF 100ng/ml**	19.57 (2.90)	9.12 (2.38)	14.57 (2.55)
	EGF 300ng/ml	20.70 (2.76)	12.15 (2.38)	14.70 (1.72)
Day 13	**EGF 100ng/ml**	10.30 (1.66)	5.34 (1.82)	7.57 (2.48)
	EGF 300ng/ml	16.83 (1.49)	11.46 (1.67)	10.92 (2.93)

Table 4.1. The effect of HNF1B genotype on apoptosis and cell proliferation from the foregut progenitor to the pancreatic progenitor stages of differentiation.
(A) Percentage of early and late apoptotic cells for HNF1B$^{+/+}$, HNF1B$^{+/-}$ and HNF1B$^{-/-}$ genotypes at days 6, 10 and 13.
(B) Percentage of proliferating cells (EdU^{+}) HNF1B$^{+/+}$, HNF1B$^{+/-}$ and HNF1B$^{-/-}$ genotypes at days 6, 10 and 13.
Data is shown as mean (standard deviation), n=3 independent experiments.

4.2.6 Differentiation of patient iPSCs

Two patient iPSC lines R2 and R4 were differentiated to the pancreatic progenitor stage (Day 13). Three clones for each patient line were differentiated and 3 differentiations were carried out for each patient-derived iPSC clone. Since isogenic corrected cell lines for both patients R2 and R4 were not available, the differentiation capacity was compared to the two non-targeted WT hiPSC lines used previously, FSPS13.B and Eipl_1. Differentiations were carried out at passages: P27 (E8 13), P34 (E8 20) and P37 (E8 23) for the R2 hiPSC line; P17 (E8 17), P25 (E8 25) and P28 (E8 28) for the R4 hiPSC line; P69 (E8 39), P75 (E8 45) and P78 (E8 48) for the FSPS13.B WT hiPSC line; and P56 (E8 33), P62 (E8 39) and P65 (E8 42) for the Eipl_1 WT hiPSC line.

At the DE stage, the percentage of $SOX17^+$ cells was 78% for the R2 hiPSC line, 82% for the R4 hiPSC line, 92% for the FSPS13.B hiPSC line and 93% for the Eipl_1 hiPSC line (Figure 4.12A). At the pancreatic progenitor stage, the percentage of $PDX1^+$ and $PDX1^+/NKX6\text{-}1^+$ cells was 79% and 4% respectively for the R2 iPSC line, 87% and 31% for the R4 iPSC line compared with 97% $PDX1^+$ cells and 42% $PDX1^+/NKX6\text{-}1^+$ for the FSPS13.B iPSC line and 96% $PDX1^+$ cells and 32% $PDX1^+/NKX6\text{-}1^+$ for the Eipl_1 iPSC line (Figure 4.12B and C). These results suggest that iPSCs derived from patients R2 and R4 can be differentiated into pancreatic progenitor cells. Significantly reduced efficiency to differentiate to $PDX1^+/NKX6\text{-}1^+$ cells was seen for the R2 hiPSC line but not the hiPSC R4 line, compared with the wild-type FSPS13.B and Eipl_1 iPSC lines. However, without an isogenic corrected iPSC line, it is unclear whether this difference is due to the HNF1B mutation or the capacity of differentiation determined by the overall genetic background of the hiPSC lines.

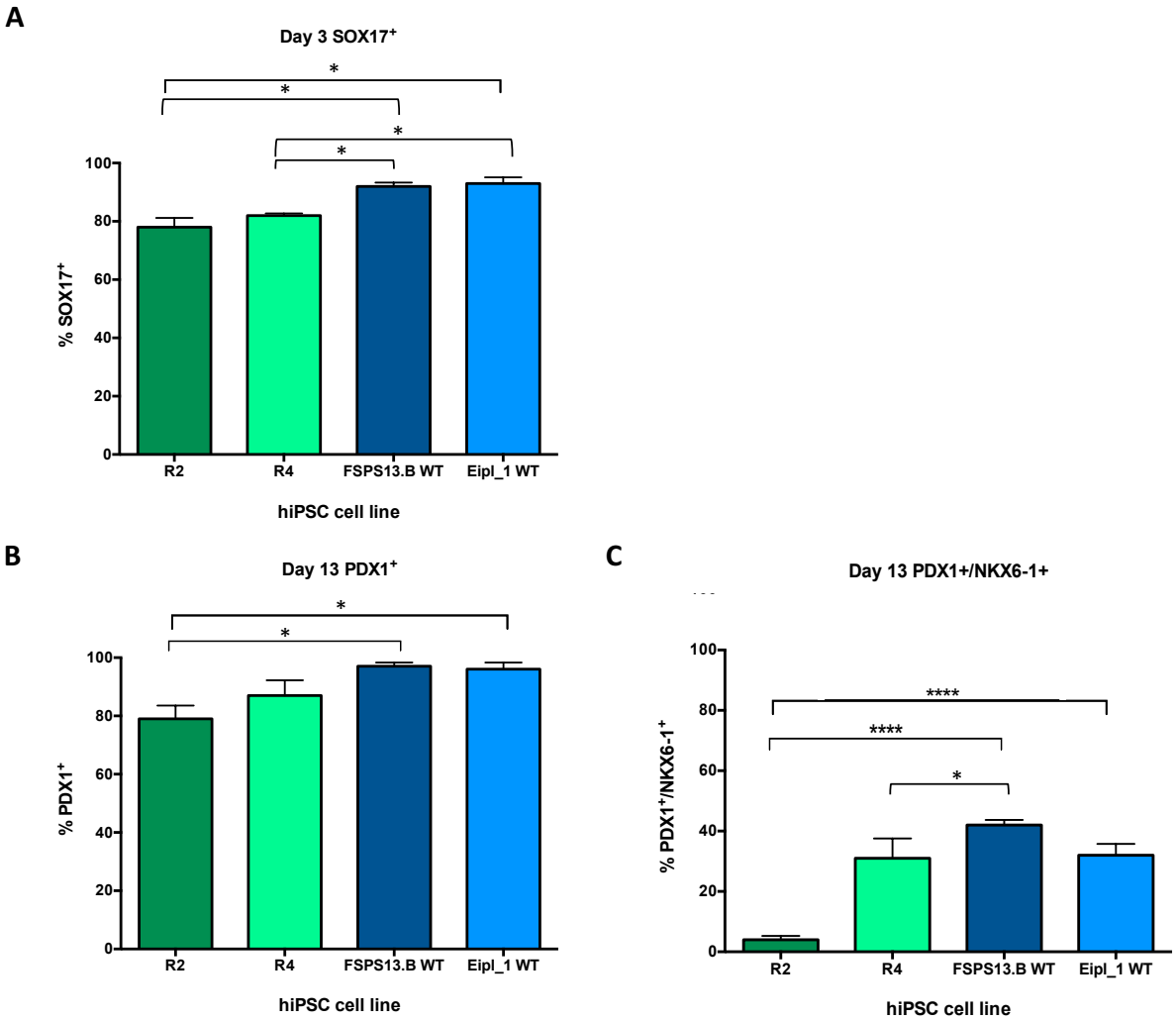


Figure 4.12. Differentiation of 2 patient hiPSC lines; R2 and R4 to **(A)** definitive endoderm (day 3) and **(B)** pancreatic progenitor (day 13) cells. 3 clones for each patient line were differentiated and 3 differentiations were carried out for each patient-derived iPSC clone. Differentiation capacity was compared to two WT hiPSC lines, FSPS13.B and Eipl_1. Data are presented as mean ± SEM unless otherwise indicated. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$ and **** $p < 0.0001$. n = 3 independent experiments.

4.2.7 In vivo animal transplantation studies

One of the objectives of this project is to study the function of HNF1B in adult-like islet cells. Current protocols do not allow the generation of functional adult-like islet cells in vitro. Therefore, an alternative approach involves injecting pancreatic progenitors under the kidney capsule of immunodeficient mice, which provides an in vivo environment to allow the progenitor cells to develop into functionally mature adult-like islet cells (Cho et al., 2012, Kroon et al., 2008, Nostro et al., 2015).

In order to study the effect of HNF1B heterozygous and homozygous genotypes on differentiation of pancreatic progenitor cells to islet cells, hiPSC-derived pancreatic progenitor (day 13) cell aggregates in Matrigel were loaded into a syringe for cell delivery under the kidney capsule of NSG mice (5 million cells per animal). Approximately 80% of cells were still viable at 4 hours post collection of cell aggregates. Three sets of transplantation experiments were performed, and for each experiment, animals received pancreatic progenitor cells derived from $HNF1B^{+/+}$, $HNF1B^{+/-}$ or $HNF1B^{-/-}$ hIPSCs. These experiments were performed using the FSPS13.B hiPSC lines. Each group in each set of experiments contained 4 animals which were matched for age and sex.

Out of the total of 36 mice who received pancreatic progenitor cell transplants, 8 animals died intraoperatively or within two days following the surgery due to complications. Intra-operative complications included cardiac arrest, likely due to the anaesthetic or emboli formed from the injected cell aggregates (n=5). Initial post-operative complications included biting of sutures resulting in large open wounds and subsequent culling of the animals (n=5). Four further animals were culled prior to 18 weeks due to development of large intra-abdominal masses at the site of the grafts suggesting tumorigenic growth. Three out of the four animals who developed masses received $HNF1B^{-/-}$ pancreatic progenitor cells and one received $HNF1B^{+/-}$ pancreatic progenitor cells. In total, at 18 weeks, 8 mice remained in the $HNF1B^{+/+}$ group, 9 mice in the $HNF1B^{+/-}$ group and 7 mice in the $HNF1B^{-/-}$ group.

Immunohistochemistry showed that $HNF1B^{+/+}$ pancreatic progenitor cells were able to generate all three pancreatic lineages: ductal (cytokeratin 7 (CK7) positive), islet cells (chromogranin A (CHGA) positive) and to a lesser extent acinar cells (PASd positive), following transplantation under the kidney capsule in mice (Representative images are shown in Figure 4.13.). H&E staining showed that pancreatic structures could only be found in 4 out of the 8 mice who

received HNF1B$^{+/+}$ pancreatic progenitor cells, 2 out of the 9 mice who received HNF1B$^{+/-}$ and none of the mice who received HNF1B$^{-/-}$ (Representative images are shown in Figure 4.14.). Immunofluorescence staining showed presence of CK19^{+} ductal cells and INS^{+} cells in mice who received HNF1B$^{+/+}$ or HNF1B$^{+/-}$ pancreatic progenitor cells (representative images are shown in Figure 4.15.). No significant differences in the development of pancreatic structures were observed between mice who received HNF1B$^{+/+}$ or HNF1B$^{+/-}$ pancreatic progenitor cells. Immunostaining was attempted for other ductal, endocrine and exocrine cell markers; however, these experiments were not successful and require further optimisation.

In vivo glucose stimulated insulin secretion (GSIS) was determined at 18 weeks following transplantation of hiPSC-dervied pancreatic progenitor cells. An increase in glucose concentration post-glucose administration was seen in 3 out of 4 of the mice who received HNF1B$^{+/+}$ pancreatic progenitor cells and in whom pancreatic structures were seen, and in 1 of the 2 mice who received HNF1B$^{+/-}$ pancreatic progenitor cells and in whom pancreatic structures were seen (Figure 4.16.). However, due to the very low numbers of mice no conclusions can be made regarding the differences in GSIS between the HNF1B$^{+/+}$ and HNF1B$^{+/-}$ groups. In addition, very low concentrations of C-peptide, at almost the lower limit of detection of the kit, were detected pre and post-glucose administration. Further optimisation of these in vivo experiments, with larger numbers of mice is required.

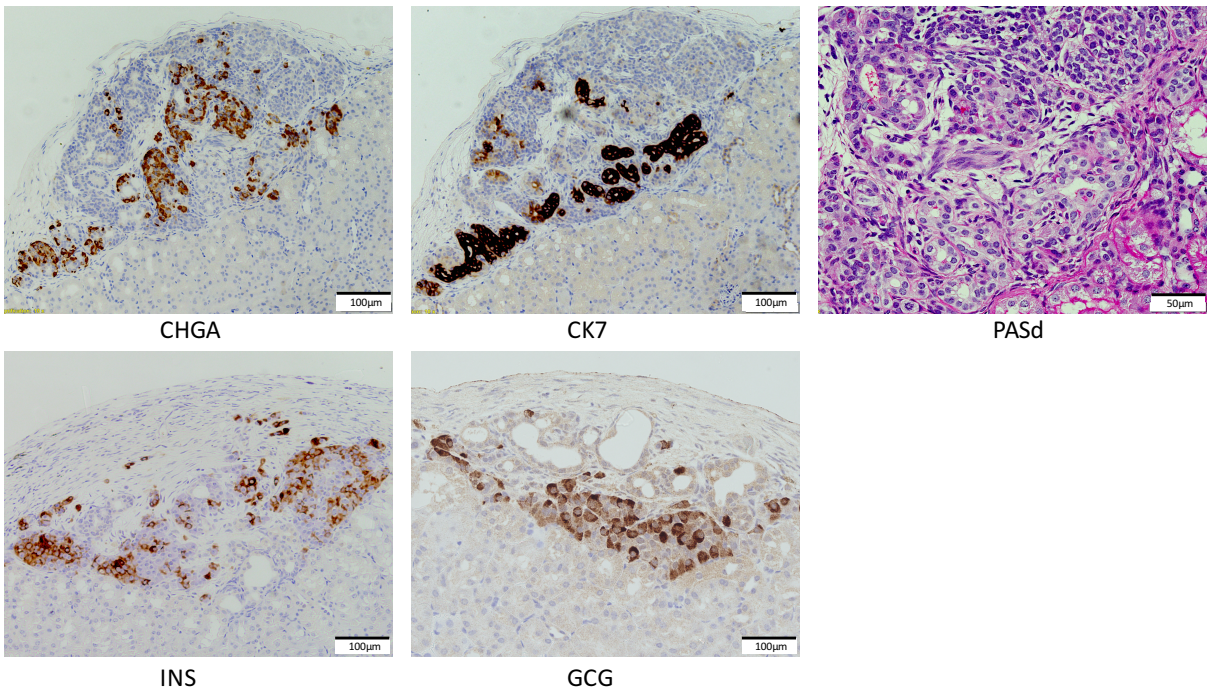


Figure 4.13. Immunohistochemistry showing that HNF1B$^{+/+}$ pancreatic progenitor cells were able to generate all three pancreatic lineages: ductal (cytokeratin 7 (CK7) positive), islet cells (chromogranin A (CHGA) positive) and to a lesser extent acinar cells (PASd positive), following transplantation under the kidney capsule in mice. Representative images from 1 experiment are shown.

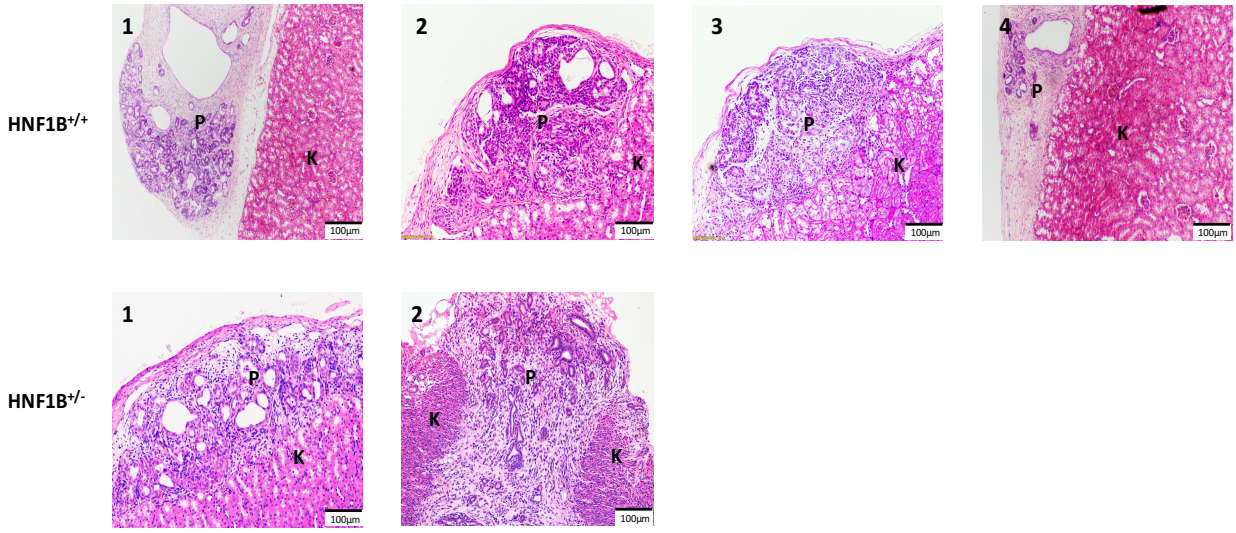


Figure 4.14. H&E staining showing development of pancreatic structures following transplantation of HNF1B$^{+/+}$ and HNF1B$^{+/-}$ pancreatic progenitor cells under the kidney capsule in mice. Images from 4 mice who received HNF1B$^{+/+}$ pancreatic progenitor cells and 2 mice who HNF1B$^{+/-}$ pancreatic progenitor cells are shown.
P = pancreatic tissue, K = kidney tissue

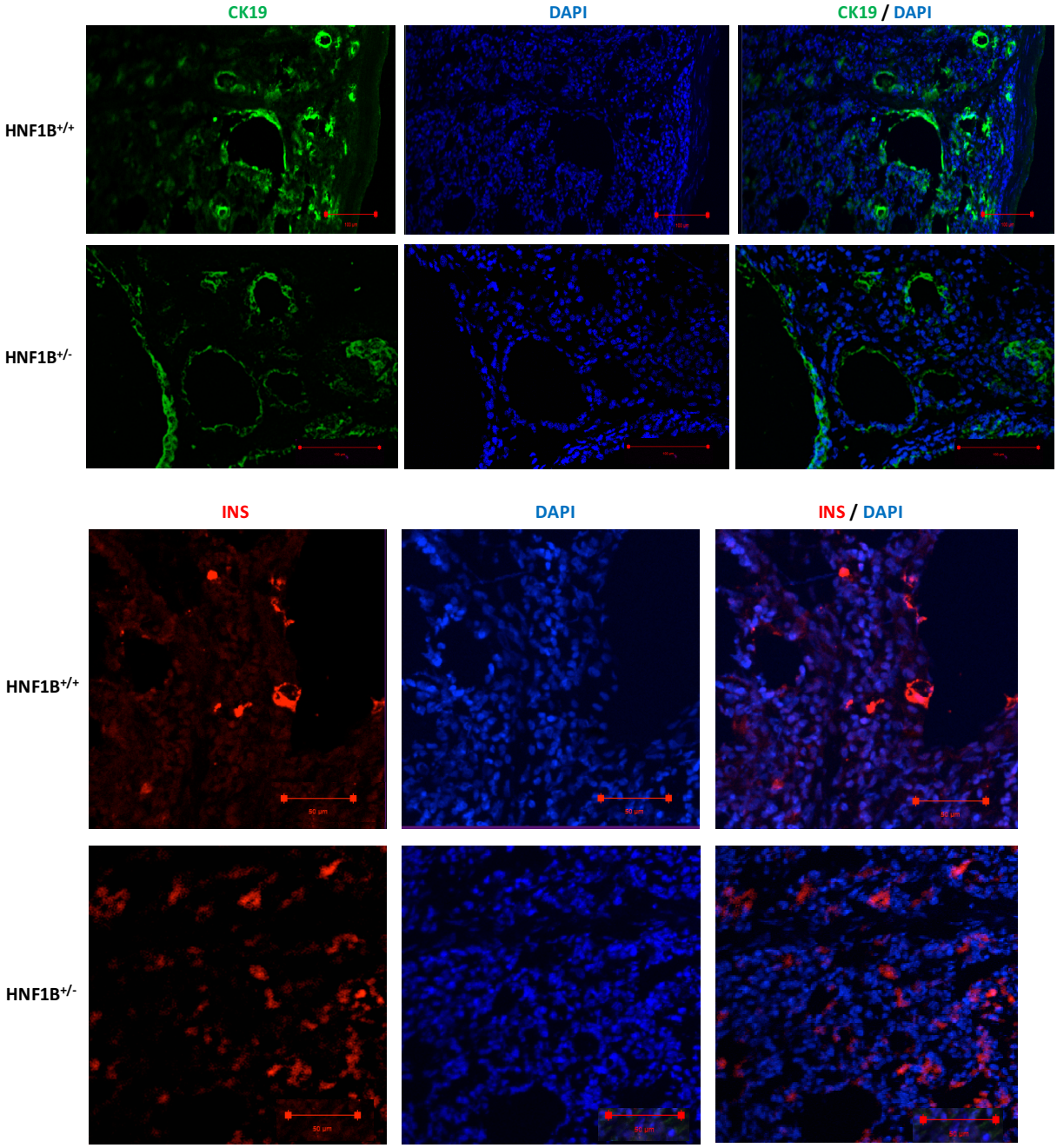


Figure 4.15. Immunofluorescence staining showing presence of CK19^{+} ductal (green) and INS^{+} endocrine cells (red) following transplantation of HNF1B$^{+/+}$ and HNF1B$^{+/-}$ pancreatic progenitor cells under the kidney capsule in mice. Representative images from 1 mouse for each genotype are shown. Scale bar - 100µm

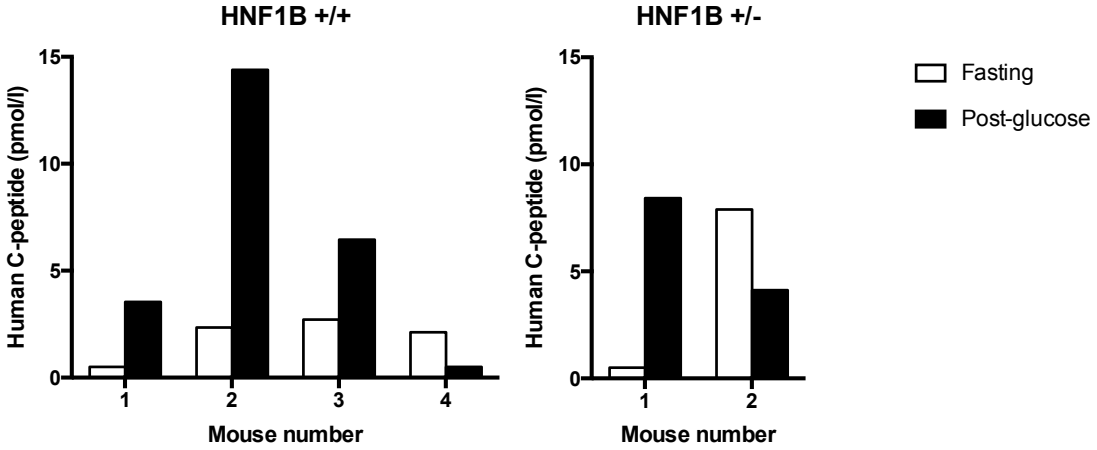


Figure 4.16. In vivo GSIS at 18 weeks following transplantation of pancreatic progenitor cells. Blood samples were collected after a 6-8 hour fast and 60 min post glucose injection, respectively. Human C-peptide was measured for each mouse transplanted with either HNF1B$^{+/+}$ or HNF1B$^{+/-}$ pancreatic progenitor cells in whom pancreatic structures were seen on H&E staining.

4.3 Discussion

Using the 27-day pancreatic differentiation protocol developed in our laboratory, we were able to differentiate hiPSCs along the pancreatic lineage to produce β-like cells with similar efficiency to that seen with other differentiation protocols in the literature using a 2D culture system. Expression of HNF1B was up-regulated at the foregut progenitor stage and remained high at the pancreatic progenitor stage and throughout the rest of the differentiation up to the hormonal or β-like cell stage.

HNF1B$^{-/-}$ hiPSCs were able to differentiate to definitive endoderm cells with the same efficiency as HNF1B$^{+/+}$ and HNF1B$^{+/-}$ cells; however, there was a complete defect of foregut and pancreatic development *in vitro*. A dose-sensitive effect of HNF1B loss was observed since heterozygous knockout of HNF1B in hiPSCs resulted in significant impairment, but not complete loss, of pancreatic progenitor and endocrine cell development. There was a significant reduction in the number of C-peptide positive β-like cells in HNF1B$^{+/-}$ compared with HNF1B$^{+/+}$ cells, although the produced β-like cells were able to secrete C-peptide, with similar GSIS in proportion to the number of β-like cells. Importantly, these observations are in agreement with the known expression pattern and function of HNF1B in human foregut and pancreatic development. The findings support a predisposition of patients with heterozygous HNF1B mutations to diabetes due to a decrease in β-cell number however, some β-cell dysfunction cannot be completely excluded.

A significant reduction in cell proliferation was seen in HNF1B$^{+/-}$ and HNF1B$^{-/-}$ cells at the foregut progenitor stage and in HNF1B$^{+/-}$ at the pancreatic progenitor stage compared with HNF1B$^{+/+}$ cells. No significant differences in cell apoptosis was seen between the genotypes. Therefore, HNF1B haploinsufficiency seems to impair the expansion and maintenance of pancreatic progenitor cells during pancreas development, resulting in reduced β-cell numbers.

Two hiPSC lines with different genetic backgrounds were used for the differentiation experiments to determine the differences in the expression of stage-specific markers throughout the pancreatic differentiation process between the HNF1B$^{+/+}$, HNF1B$^{+/-}$ and HNF1B$^{-/-}$ genotypes. Three HNF1B$^{+/+}$, three HNF1B$^{+/-}$ and two HNF1B$^{-/-}$ clones for each of the hiPSC lines, were differentiated for each experiment. Ideally, these experiments should be repeated using a greater number of hiPSC lines including iPSCs derived from patients with HNF1B-associated diabetes in order to minimize or exclude the effect of the genetic background of the

hiPSC lines and any off-target effects. The differentiations should also be repeated using different pancreatic differentiation protocols in order to ensure that the phenotype seen is not specific to our protocol.

In addition, only one cell line (FSPS13.B) was used for the cell proliferation and apoptosis experiments. The experiments should be repeated in at least one further cell line to ensure that the phenotype is reproducible in another cell line with a different genetic background. Another limitation of the experiments used to determine cell apoptosis is that dead cells are removed during media changes prior to staining with the apoptosis markers, and therefore, the number of cells in early and late apoptosis detected will only reflect those which are still attached to the culture plate.

In vivo experiments were performed in order to study the function of HNF1B in functionally mature adult-like islet cells, since current differentiation protocols do not allow the generation of functional adult-like islet cells in vitro. The in vivo experiments showed that $HNF1B^{+/+}$ and $HNF1B^{+/-}$ pancreatic progenitor cells can differentiate to all three pancreatic lineages: ductal, endocrine and exocrine cells following transplantation under the kidney capsule in mice. Transplantation of $HNF1B^{-/-}$ cells in vivo did not result in the development of any pancreatic structures. However, in order to determine whether there are any significant differences in the development of endocrine, exocrine and ductal pancreatic cells and insulin secretion in vivo between $HNF1B^{+/+}$ and $HNF1B^{+/-}$ genotypes, further optimisation of the in vivo experiments is required. Technical challenges that need to be optimised include: number of pancreatic progenitor cells transplanted under the kidney capsule, different techniques for delivering the cells under the kidney capsule, timing of GSIS post-transplantation of pancreatic progenitor cells and timing of blood samples collections post glucose injection. Optimization of immunostaining protocols for endocrine, exocrine and ductal cell markers is also required.

Another approach for these experiments would be to render the mice diabetic using streptozotocin and then perform the GSIS experiments. This would allow us to determine whether the beta cells which develop in vivo after transplantation of the pancreatic progenitor cells would be able to produce enough insulin to reverse the diabetes in these mice.

In summary, we have shown that HNF1B has an essential role in foregut and pancreas development; homozygous loss of HNF1B results in a complete defect of foregut and pancreatic development *in vitro*, while heterozygous loss results in significant impairment of pancreatic

progenitor and endocrine cell development. Further understanding of the molecular mechanisms underlying how the loss of function of one or both alleles of HNF1B impairs pancreatic progenitor development is required and is explored in the next chapter.

Chapter 5 Exploring the transcriptional network controlled by HNF1B

5.1 Introduction

5.1.1 Regulation of gene expression

The transcriptional regulatory system plays a central part in controlling many biological processes, ranging from cell cycle progression, maintenance of intracellular metabolic and physiological balance, to cellular differentiation and developmental time courses (Voss and Hager, 2014). Transcription factors (TFs) are proteins that bind to cis-regulatory DNA sequences and are responsible for regulating the transcription of specific genes to mRNA. They control gene expression by binding to *cis*-regulatory DNA sequences or motifs in promotor and/or enhancer regions, interacting with each other and with cofactors to form complexes, and recruiting RNA polymerase II.

Enhancers are regulatory elements located remotely from gene promoters and can be several kilobases away from the TSS. Multiple transcription factors may bind to individual enhancers and regulate transcription from the core promoters of nearby or distant genes through physical contacts that involve looping of the DNA between enhancers and the core promoters (Lee and Young, 2013). Enhancer interactions are highly specific to cell type and state. Hence, a gene may be regulated by different enhancers in different cell types, at different developmental stages, and in response to different signals (Andersson, 2015).

The processes of development and differentiation in mammalian systems are regulated by constantly changing cohorts of transcription factors (TFs), that direct cell-specific transcriptional programmes. However, simple site-specific binding is insufficient to govern gene regulatory networks given the large sizes and the complexity of the mammalian genomes (Voss and Hager, 2014) and transcriptional regulation usually involves two important elements; transcription factors and the transcription apparatus and epigenetic modifications.

5.1.2 Chromatin structure and function

DNA in cells is packaged into chromatin to allow long DNA molecules to fit into the nucleus, prevent DNA damage and to regulate gene expression and DNA replication. The basic

structural unit of chromatin is the nucleosome, which consists of 147bp of DNA wrapped around 8 histone molecules, 2 copies each of the core histones H2A, H2B, H3, and H4, separated by linker DNA and another histone (H1) (Kornberg, 1974). Histone tails are post-translationaly modified to effect their interaction with DNA and other proteins. The most common modifications or "marks" include methylation, acetylation, phosphorylation and ubiquitination, but other modifications can also occur (Bannister and Kouzarides, 2011). Chromatin state is very important in determining whether a particular gene is activated, poised or repressed and is crucial in regulating the transcriptional identity of the cell.

Cellular lineage commitments are implemented through remodelling of chromatin accessibility that allows transcription factor binding of promoter and enhancer cis-regulatory modules (CRMs) across the genome. Positioning of nucleosomes throughout a genome has a significant regulatory function by modifying the availability of binding sites to transcription factors (TFs) and the general transcription machinery and thus affecting DNA-dependent processes such as transcription, DNA repair, replication and recombination (Radman-Livaja and Rando, 2010). Densely positioned nucleosomes or "closed" chromatin can restrict access for transcription factors (both activators and repressors), CCCTC-binding factor (CTCF), RNA polymerase II (Pol II) and other proteins. Accessible nucleosome-depleted regions or "open" chromatin can be bound by these proteins and are identified as regulatory elements e.g. active enhancers, repressors or core promoters. The transition between 'open' and 'closed' chromatin, is determined by regulatory proteins, including pioneer transcription factors (Shlyueva et al., 2014). Insulator proteins (for example, CTCF) and other architectural proteins also bind to open regions, and make up a substantial proportion of sites that are accessible across multiple cell types (e.g. 10%) (Xi et al., 2007).

5.1.3 Use of NGS to study chromatin accessibility

To successfully interact with genomic regulatory elements, transcription factors must induce the reorganization of local nucleosome structures. These disturbances of the ordered nucleosome positioning can be mapped by detecting DNA sequences that have increased nuclease accessibility i.e. nuclease hypersensitivity. Major insights into the epigenetic information encoded within the nucleoprotein structure of chromatin have come from high-throughput, genome-wide methods for separately assaying chromatin accessibility ('open chromatin'), nucleosome positioning, and transcription factor (TF) occupancy. Large collaborative projects such as ENCODE (The Encode Project Consortium, 2012) and the NIH

epigenomic roadmap (Bernstein et al., 2010) are part of a major effort to identifying various functional elements of the genome. These include regions of DNA that bind proteins that control transcription including enhancers and promoters.

Chromatin accessibility assays separate the genome by enzymatic or chemical means and isolate either the accessible or protected locations. Isolated DNA is then quantified using a next-generation sequencing (NGS) platform. These approaches directly measure the effect of chromatin structure modifications on gene transcription and unlike ChIP-seq do not require antibodies or epitope tags that can introduce potential bias (Tsompana and Buck, 2014). Chromatin accessibility assays that isolate accessible locations across the whole genome include; deoxyribonuclease (DNase) -seq (Boyle et al., 2008), which identifies DNase I hypersensitivity sites by capturing DNase-digested fragments; formaldehyde-assisted isolation of regulatory elements (FAIRE) –seq (Giresi et al., 2007), which involves the crosslinking of chromatin with formaldehyde to capture protein-DNA interactions, and subsequent isolation of sheared chromatin using phenol-chloroform; and assay for transposase-accessible chromatin (ATAC) –seq (Buenrostro et al., 2013), which is based on the ability of hyperactive Tn5 transposase to fragment DNA and integrate into active regulatory regions; and micrococcal nuclease (MNase) –seq (Gaffney et al., 2012, Ponts et al., 2010) which indirectly evaluates chromatin accessibility by digestion of chromatin followed by recovery and sequencing of nucleosome-associated DNA, which in insensitive to digestion by MNase (Figure 5.1).

ATAC-seq uses the enzyme Tn5 transposase (Goryshin and Reznikoff, 1998; Adey et al., 2010) to simultaneously cleave DNA and integrate adapters into the cleaved genomic DNA i.e. tagmentation for high-throughput sequencing at regions of increased accessibility. This method can be used to uncover open chromatin, nucleosome positioning and TF footprints genome-wide. If regions of open chromatin are sequenced to great depths (>1x10^6 reads/sample), classic footprints that correspond to potential transcription factor-binding sites can be appreciated. ATAC-seq has a higher sensitivity than other methods with a significantly lower starting cell number (500 to 50,000 cells) and multiple aspects of chromatin architecture can be studied simultaneously at high resolution.

Figure 5.1. Schematic diagram of current chromatin accessibility assays performed with typical experimental conditions. Representative DNA fragments generated by each assay are shown, with end locations within chromatin defined by colored arrows. Bar diagrams represent data signal obtained from each assay across the entire region. The footprint created by a transcription factor (TF) is shown for ATAC-seq and DNase-seq experiments (Chi, 2016).

5.1.4 Use of PSCs to study stage-specific transcriptional programs during pancreatic lineage progression

Numerous diseases arise from a breakdown in the regulatory system; a third of human developmental disorders have been attributed to dysfunctional TFs (Boyadjiev and Jabs, 2000, Vaquerizas et al., 2009). One obvious limitation to studying the genomic regulation of human organogenesis and the role of a particular gene in a developmental process lies in the restricted access to normal and disease human embryonic tissue.

Classic mouse knockout models and human genetics have uncovered multiple transcription factors (TFs) that regulate development of the pancreas. However, little is known with regards to how particular pancreatic TFs function within regulatory networks. One approach to study transcriptional networks involves using human pluripotent stem cells (hPSCs) to derive cellular populations that express organ-specific progenitor markers. Human pluripotent stem cells are an ideal model for studying the dynamic epigenetic changes that accompany the development of cellular diversity (Boland et al., 2014). High-resolution global analyses of epigenetic dynamics during hPSC differentiation can be performed and have greatly advanced our understanding of the mechanisms that regulate aspects of human embryonic development. The use of in-vitro pancreatic progenitor cells as a model to study gene regulation in early pancreas development has been validated in a study by Cebola et al. (Cebola et al., 2015).

In this chapter, an in-depth comparison of the global transcriptional profile of HNF1B WT ($HNF1B^{+/+}$), HNF1B heterozygous knockout ($HNF1B^{+/-}$) and HNF1B homozygous knockout ($HNF1B^{-/-}$) hiPSC cells lines at each stage of the differentiation process was performed using RNA-seq. In addition, ATAC-seq was used to determine chromatin accessibility in $HNF1B^{+/+}$ compared with $HNF1B^{+/-}$ and $HNF1B^{-/-}$ cells at day 6 and day 13 of the pancreatic differentiation process. The RNA-seq, ChIP-seq and ATAC-seq data was analysed with the assistance of a bioinformatician; Dr Pedro Madrigal (WTSI).

5.2 Results

5.2.1 Differential gene expression analysis using RNA-sequencing

5.2.1.1 RNA-sequencing data quality assessment and control

RNA-seq was performed on HNF1B$^{+/+}$, HNF1B$^{+/-}$ and HNF1B$^{-/-}$ cells at the foregut progenitor stage (day 6), when HNF1B starts to be significantly expressed, posterior foregut stage (day 8), pancreatic progenitor stage (day 13), endocrine progenitor stage (day 16) and hormonal cell or β-like cell stage (day 27). Sample information together with the total number of aligned fragments and mapped reads with quality score >10 are shown in Table 5.1. One HNF1B$^{+/-}$, one HNF1B$^{-/-}$ and one HNF1B$^{+/+}$ (targeted wild-type) clone from the FSPS13.B hiPSC line were differentiated along the pancreatic lineage to produce cells for these experiments. Three independent experiments (biological triplicates) were sequenced for each clone at each stage of differentiation.

Data quality assessment and quality control were performed, preceding differential gene analysis. A heatmap of the distance matrix is shown in Figure 5.1A. and gives an overview of the similarities and dissimilarities between samples. Principal Componant Analysis (PCA) was also used to generate a genome wide overview of the global patterns and sources of variation between HNF1B$^{+/+}$, HNF1B$^{+/-}$ and HNF1B$^{-/-}$ cells at each stage of differentiation (Figure 5.1B). The first principal component (PC1) explained 53% of the variance and separated HNF1B$^{+/+}$, HNF1B$^{+/-}$ and HNF1B$^{-/-}$ cells by day of differentiation. The second principal component explained 20% of the variance and separated HNF1B$^{+/+}$, HNF1B$^{+/-}$ and HNF1B$^{-/-}$ cells by HNF1B level or genotype. As expected, HNF1B$^{+/+}$ and HNF1B$^{+/-}$ cell lines were transcriptionally more similar to each other than to HNF1B$^{-/-}$ cells. PCA analyses therefore confirms that absence of HNF1B diverts pancreas specification of hPSCs from the foregut lineage, while heterozygous knockouts have a milder but still detectable effect.

Sample name	Day	HNF1B Level	Library size	Mapped reads Quality score >10
13B Cl.45 D6_1	06	WT	118384519	88543640
13B Cl.6 D6_1	06	Het	116131701	86113194
13B Cl.48 D6_1	06	Hom	111558545	85740614
13B Cl.45 D6_2	06	WT	111655388	84139726
13B Cl.6 D6_2	06	Het	129777107	98644498
13B Cl.48 D6_2	06	Hom	111826972	88110998
13B Cl.45 D6_3	06	WT	139606811	103703665
13B Cl.6 D6_3	06	Het	106679668	80614687
13B Cl.48 D6_3	06	Hom	128714446	98328972
13B Cl.45 D8_1	08	WT	123888095	93134642
13B Cl.6 D8_1	08	Het	147013538	112050531
13B Cl.48 D8_1	08	Hom	123679567	93162048
13B Cl.45 D8_2	08	WT	127493326	98438445
13B Cl.6 D8_2	08	Het	127798804	95737326
13B Cl.48 D8_2	08	Hom	122936103	92031674
13B Cl.45 D8_3	08	WT	117085741	86900383
13B Cl.6 D8_3	08	Het	119824087	90038548
13B Cl.48 D8_3	08	Hom	119466266	90660554
13B Cl.45 D13_1	13	WT	142169329	102528677
13B Cl.6 D13_1	13	Het	135094009	93279477
13B Cl.48 D13_1	13	Hom	112511990	85968763
13B Cl.45 D13_2	13	WT	118693241	83580395
13B Cl.6 D13_2	13	Het	132360151	94662742
13B Cl.48 D13_2	13	Hom	140083454	103631461
13B Cl.45 D13_3	13	WT	127000896	96077132
13B Cl.6 D13_3	13	Het	117594841	84714110
13B Cl.48 D13_3	13	Hom	122503320	92967061
13B Cl.45 D16_1	16	WT	117505611	81310336
13B Cl.6 D16_1	16	Het	130164511	91700486
13B Cl.48 D16_1	16	Hom	141466373	100263082
13B Cl.45 D16_2	16	WT	147170171	96305978
13B Cl.6 D16_2	16	Het	147452466	87898580
13B Cl.48 D16_2	16	Hom	133070182	90936195
13B Cl.45 D16_3	16	WT	113804780	84558461
13B Cl.6 D16_3	16	Het	134215663	93391591
13B Cl.48 D16_3	16	Hom	120543917	81094469
13B Cl.45 D27_1	27	WT	125249986	90666868
13B Cl.6 D27_1	27	Het	116555579	87162835
13B Cl.48 D27_1	27	Hom	116216729	85354846
13B Cl.45 D27_2	27	WT	143787152	94199999
13B Cl.6 D27_2	27	Het	118236331	82906196
13B Cl.48 D27_2	27	Hom	127138325	80287366
13B Cl.45 D27_3	27	WT	133940466	92711641
13B Cl.6 D27_3	27	Het	130352438	93655003
13B Cl.48 D27_3	27	Hom	137048178	95732442

Table 5.1. General information about RNA-seq samples.
Sample descriptions as well as library size and number of mapped reads with quality score >10 are shown for each sample.

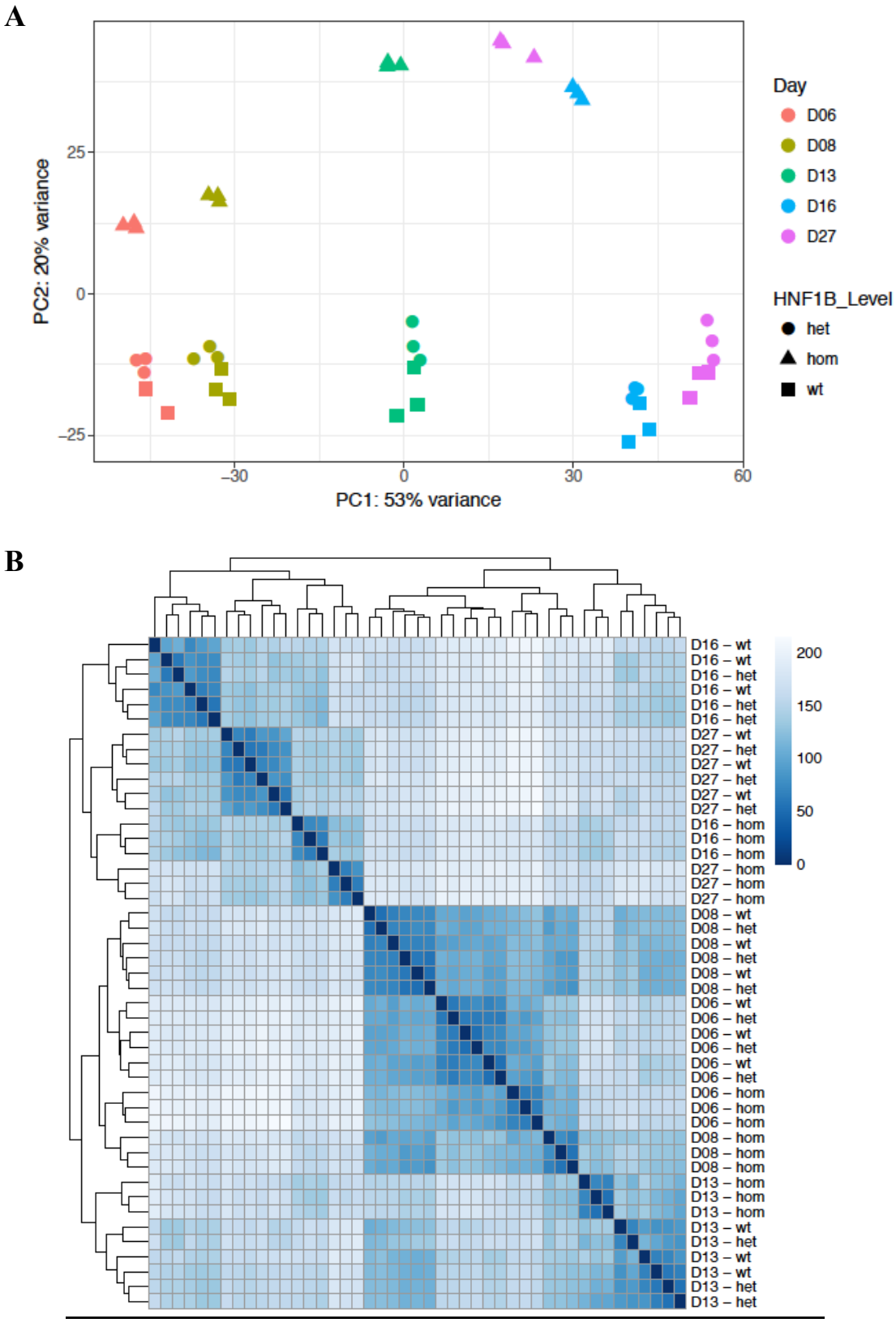


Figure 5.2. Gene expression variation between HNF1B$^{+/+}$, HNF1B$^{+/-}$ and HNF1B$^{-/-}$ cells. n= 3 biological replicates per sample. **(A)** Principal component analysis of expressed genes (counts >1) in HNF1B$^{+/+}$ (wt), HNF1B$^{+/-}$ (het) and HNF1B$^{-/-}$ (hom) cells. **(B)** Heatmap of sample-to-sample distances using rlog-transformed values. Rectangles correspond to measurements from individual biological replicates.

Transcriptional changes at the individual gene level were identified using the DESeq2 package. The normalised expression level of individual genes was calculated using the fragments per kilobase of transcript per million mapped reads (fpkm) function. Significantly differentially expressed genes were defined as a fold change of greater than 2 between two conditions with a false discovery rate threshold of 0.05.

5.2.1.2 Expression of differentiation marker genes

In Chapter 4, qRT-PCR was used to determine the expression levels of several differentiation marker genes at each stage of the differentiation protocol; at day 0 (undifferentiated hiPSCs), day 3 (definitive endoderm), day 6 (foregut progenitors), day 8 (posterior foregut), day 13 (pancreatic progenitors), day 16 (endocrine progenitors) and day 27 (β-like cells) in $HNF1B^{-/-}$, $HNF1B^{+/-}$ and $HNF1B^{+/+}$ cells.

The RNA-sequencing data was used to compliment this data and the expression levels of a selection of important differentiation marker genes from day 6 to day 27 in $HNF1B^{-/-}$, $HNF1B^{+/-}$ and $HNF1B^{+/+}$ cells is shown in Figure 5.3. Most of the genes showed similar patterns of expression to that seen from our qRT-PCR data at the relevant stage of differentiation.

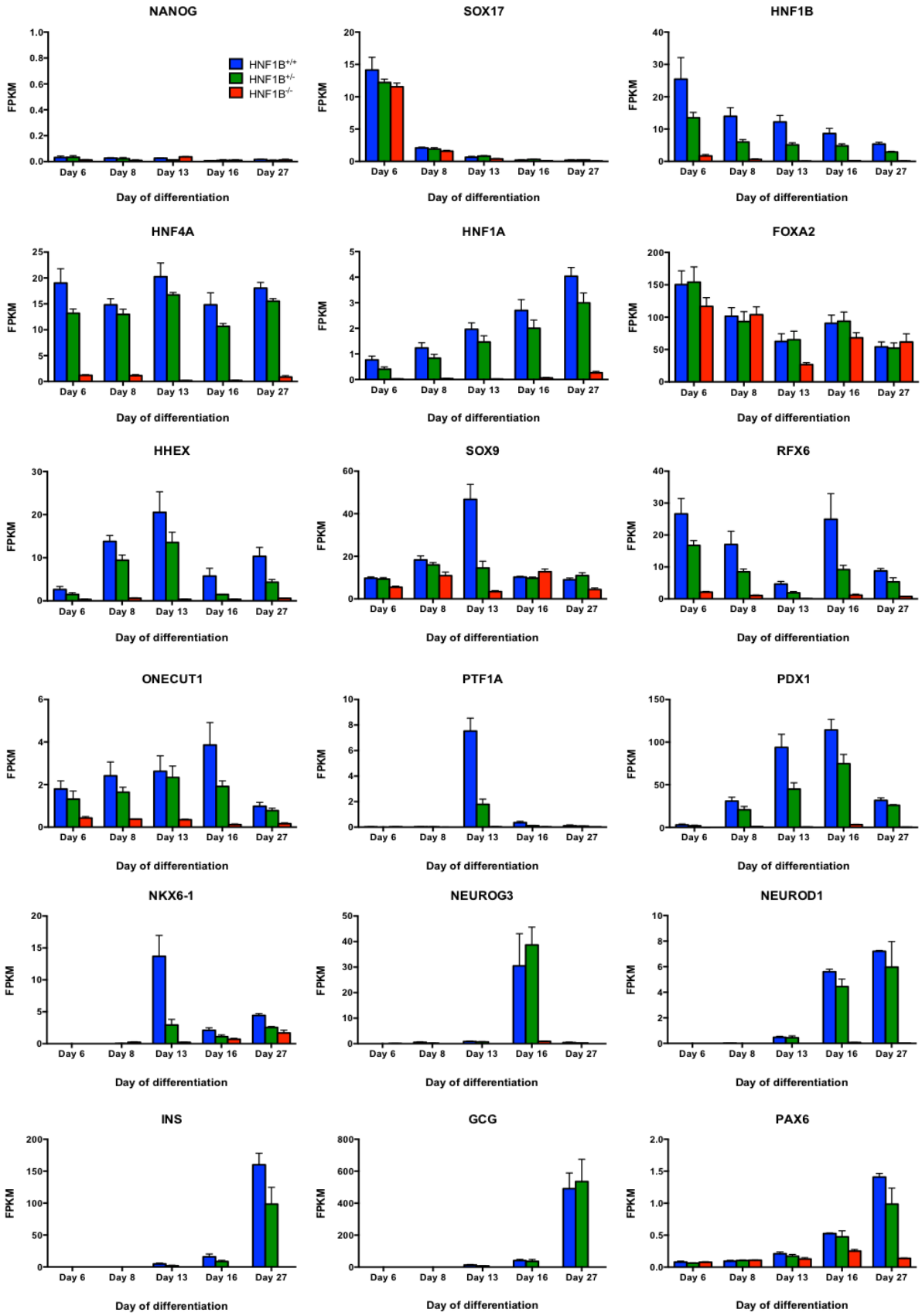


Figure 5.3. Differential expression of key pancreatic differentiation markers between HNF1B$^{+/+}$, HNF1B$^{+/-}$ and HNF1B$^{-/-}$ cells using RNA-seq data.
Gene expression was measured using RNA sequencing (n = 3 at each stage of differentiation and for each genotype) and counts were normalized using the fragments per kilobase of transcript per million mapped reads (fpkm) function of the DESeq2 package. Data are presented as mean ± SEM unless otherwise indicated.

5.2.1.3 Global patterns of gene expression

Using the significance parameters described previously, 1555 genes were differentially expressed in HNF1B$^{-/-}$ cells compared with HNF1B$^{+/+}$ cells at day 6, 1459 genes at day 8, 2998 genes at day 13, 3485 genes at day 16 and 4287 genes at day 27. Only 49 genes at day 6 and 23 genes at day 8 were differentially expressed in HNF1B$^{+/-}$ cells compared with HNF1B$^{+/+}$ cells. 374 genes were differentially expressed in HNF1B$^{+/-}$ cells compared with HNF1B$^{+/+}$ cells at day 13, 274 genes at day 16 and 187 genes at day 27. The total number of up-regulated or down-regulated genes in HNF1B$^{+/+}$ compared with HNF1B$^{+/-}$ and HNF1B$^{-/-}$ cells and the number of up and down-regulated genes filtered for protein coding genes with average expression fpkm >1 in at least one condition are shown in Table 5.2. Lists of all the down-regulated and up-regulated genes between HNF1B$^{+/+}$ and HNF1B$^{+/-}$ and HNF1B$^{+/+}$ and HNF1B$^{-/-}$ cells at the foregut progenitor stage (day 6), posterior foregut stage (day 8), pancreatic progenitor stage (day 13), endocrine progenitor stage (day 16) and hormonal cell or β-like cell stage (day 27) is shown in Appendix 7.

Heatmaps of the most significant differentially expressed genes in HNF1B$^{+/+}$, HNF1B$^{+/-}$ and HNF1B$^{-/-}$ cells at day 6, day 13, day 16 and day 27 are shown in Figure 5.4. The Molecular Signature Database (MSigDb) in the Gene Set Enrichment Analysis (GSEA) tool (Mootha et al., 2003, Subramanian et al., 2005) was used to compute overlaps between the top 50 down-regulated genes in HNF1B$^{+/-}$ and/or HNF1B$^{-/-}$ cells compared with HNF1B$^{+/+}$ cells and other curated gene sets. Apart from genes known to be involved in pancreas development and β-cell function (HNF4A, HNF1B, PDX1, HNF1A, SOX9, CPA1, CPA2, SPINK1, CHGA, CHGB, ABCC8, GCG), several genes which are also involved in liver development and metabolic function including the HNF4A, HNF1A and FOXA2 gene targets; SERPINA1, AFP, ASGR2, FGB, HBP2, HMGCS2, DPYS, PRODH2; were significantly differentially expressed at all stages of the differentiation process.

A

All	Het vs WT		Hom vs WT	
	Down	Up	Down	Up
Day 6	48	1	988	567
Day 8	20	3	765	694
Day 13	263	111	1664	263
Day 16	204	70	1819	1666
Day 27	38	149	2325	981

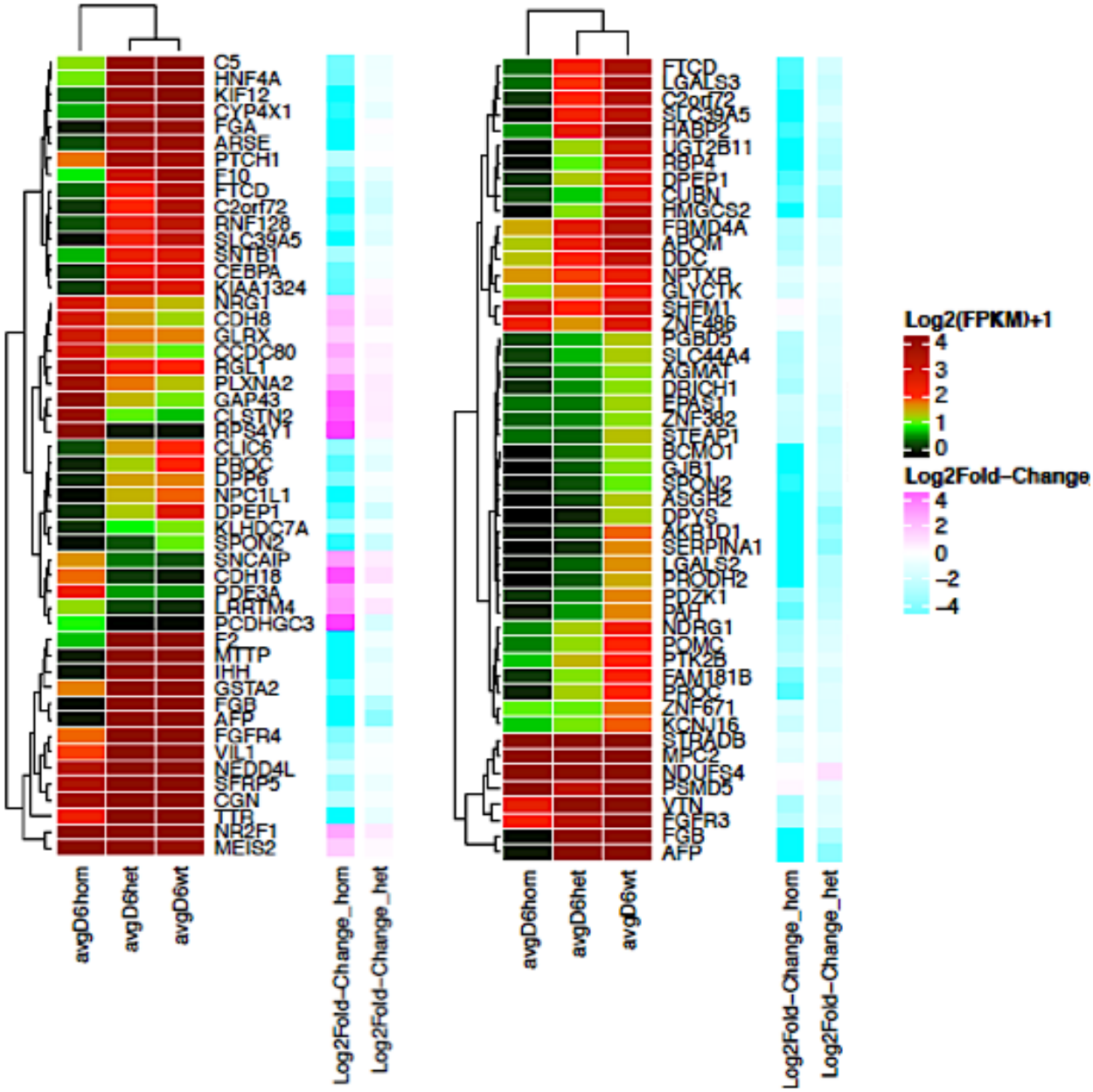
B

Protein coding with average expression >1	Het vs WT		Hom vs WT	
	Down	Up	Down	Up
Day 6	34	0	399	242
Day 8	10	1	326	320
Day 13	147	73	596	670
Day 16	108	35	863	776
Day 27	22	98	1045	326

Table 5.2. Number of differentially expressed genes (log2 fold change ≥1; adjusted p-value ≤ 0.05) for the RNA-seq experiments; The **(A)** total number and **(B)** number of protein coding genes with fpkm>1 that are up-regulated or down-regulated between $HNF1B^{+/+}$, $HNF1B^{+/-}$ and $HNF1B^{-/-}$ cells at each stage of pancreatic differentiation is shown. n=3 biological replicates for each sample.
Lists of all the down-regulated and up-regulated genes between $HNF1B^{+/+}$ and $HNF1B^{+/-}$ and $HNF1B^{+/+}$ and $HNF1B^{-/-}$ cells at the foregut progenitor stage (day 6), posterior foregut stage (day 8), pancreatic progenitor stage (day 13), endocrine progenitor stage (day 16) and hormonal cell or β-like cell stage (day 27) is shown in Appendix 7.

A

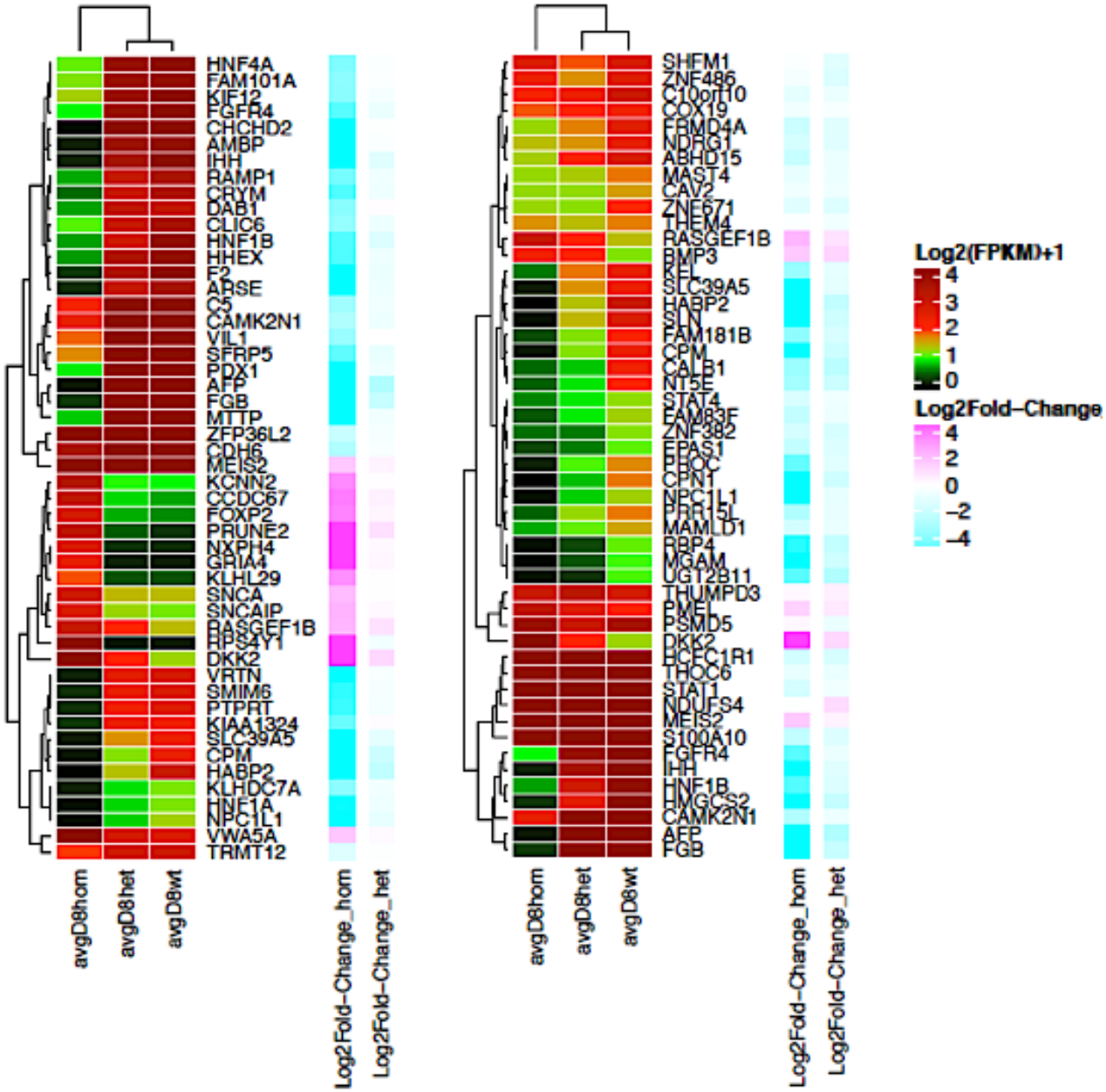
Foregut Progenitors - Day 6



HNF1B$^{-/-}$ vs HNF1B$^{+/+}$ **HNF1B$^{+/-}$ vs HNF1B$^{+/+}$**

B

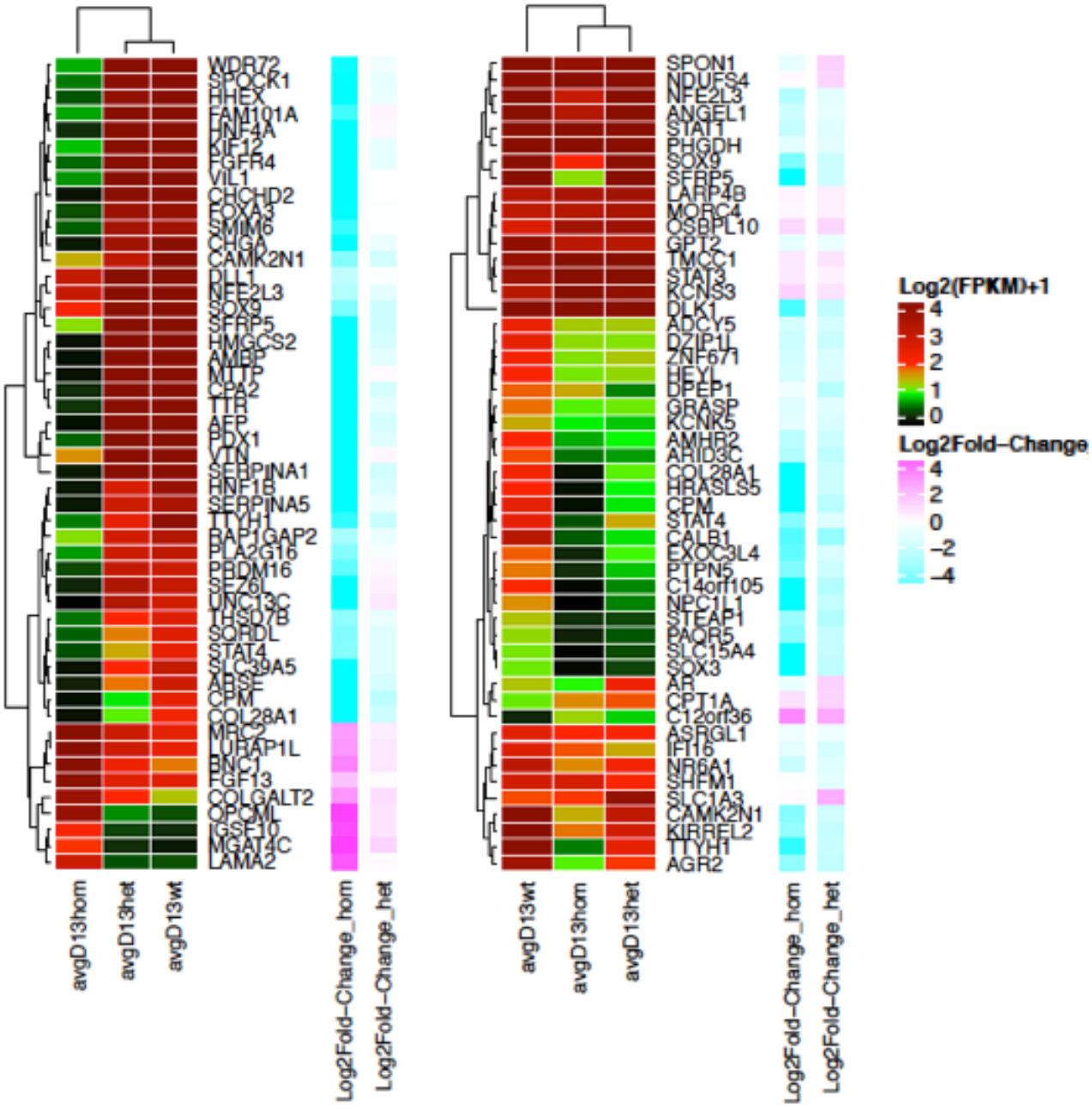
Posterior Foregut - Day 8



HNF1B$^{-/-}$ vs HNF1B$^{+/+}$ **HNF1B$^{+/-}$ vs HNF1B$^{+/+}$**

C

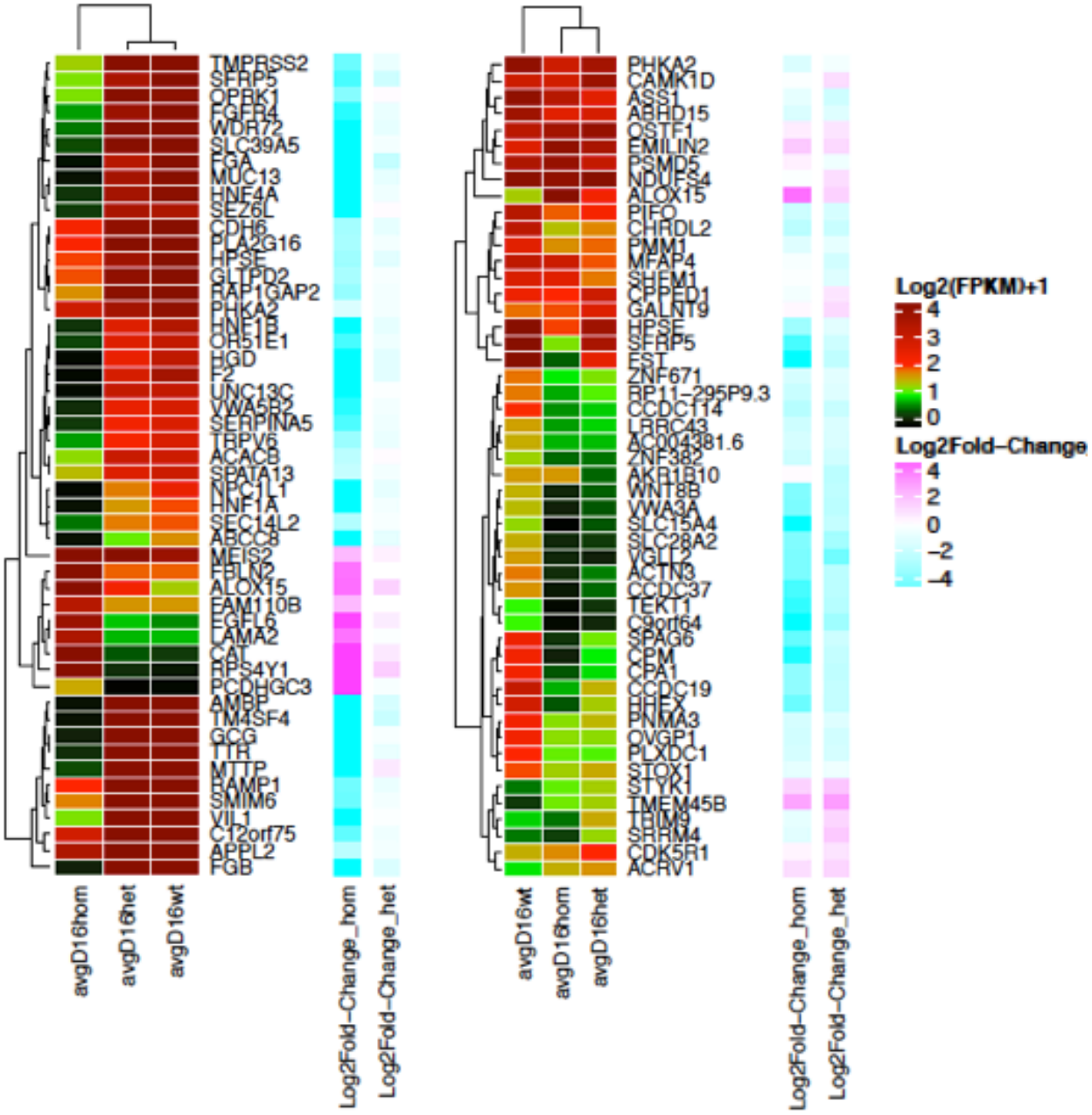
Pancreatic Progenitors - Day 13



HNF1B$^{-/-}$ vs HNF1B$^{+/+}$ **HNF1B$^{+/-}$ vs HNF1B$^{+/+}$**

D

Endocrine Progenitors - Day 16



HNF1B$^{-/-}$ vs HNF1B$^{+/+}$

HNF1B$^{+/-}$ vs HNF1B$^{+/+}$

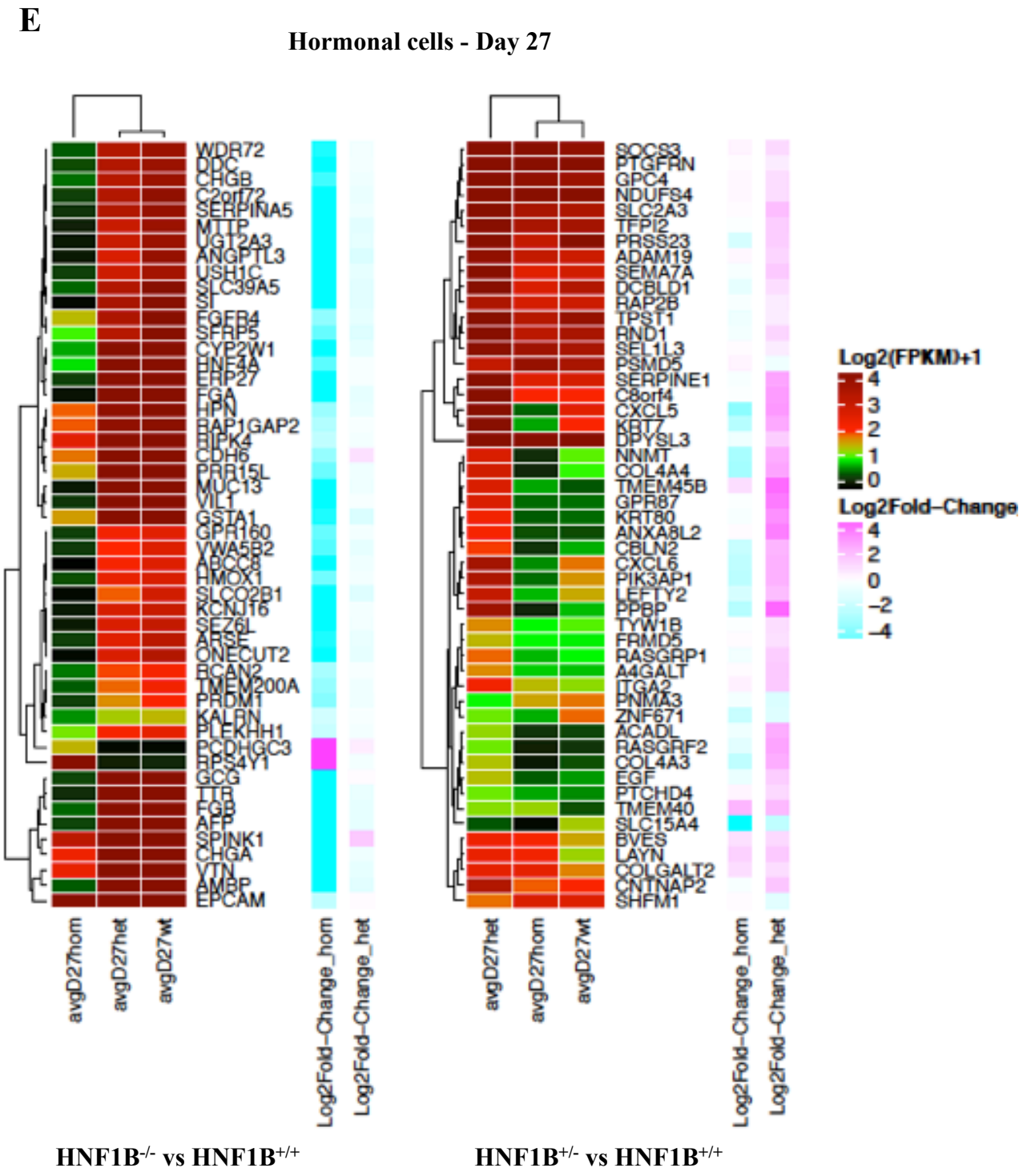


Figure 5.4. Heatmaps showing the 50 most significantly differentially expressed genes (up-regulated or down-regulated) between HNF1B$^{+/+}$, HNF1B$^{+/-}$ and HNF1B$^{-/-}$ cells at **(A)** day 6 (foregut progenitors), **(B)** day 8 (posterior foregut), **(C)** day 13 (pancreatic progenitors), **(D)** day 16 (endocrine progenitors) and **(E)** day 27 (hormonal cells / β-like cells) of differentiation. For heatmaps on left; genes are ranked based adjusted p-value for HNF1B$^{-/-}$ vs HNF1B$^{+/+}$ cells and for heatmaps on right; genes are ranked based adjusted p-value for HNF1B$^{+/-}$ vs HNF1B$^{+/+}$ cells. The mean of 3 biological replicates for each sample is shown.

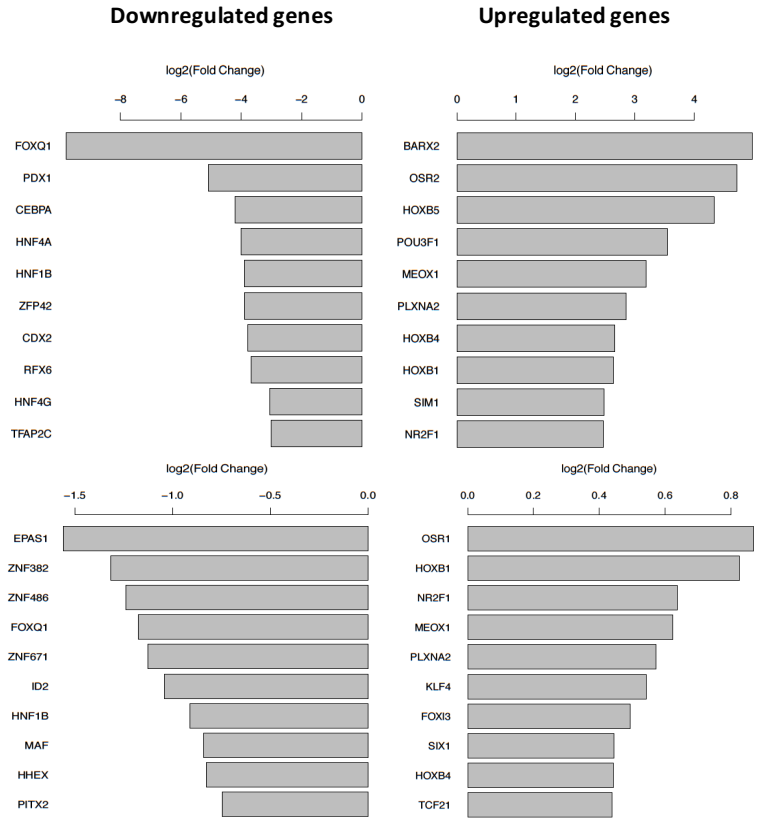
5.2.1.4 Differentially expressed transcription factor genes

In vivo and our in vitro studies have shown that HNF1B plays an important role in the pancreatic transcriptional network during development. Therefore, transcription factor genes that were significantly up or down regulated between HNF1B$^{+/+}$, HNF1B$^{+/-}$ and HNF1B$^{-/-}$ cells were identified using a curated list of 1908 human transcription factors (downloaded from http://www.tfcheckpoint.org/index.php/download-data) (Chawla et al., 2013, Vaquerizas et al., 2009). Consistent with the qRT-PCR data in chapter 4, HNF1B$^{-/-}$ cells and HNF1B$^{+/-}$ cells generally showed significantly decreased expression of transcription factor genes expressed in foregut and pancreatic progenitor cells (Figure 5.5). A smaller effect i.e. dose dependent effect was seen in HNF1B$^{+/-}$ cells compared with HNF1B$^{-/-}$ cells. At days 6 and 8, transcription factor genes known to be involved in development of foregut and foregut derivatives were down-regulated in HNF1B$^{+/-}$ and/or HNF1B$^{-/-}$ cells including HNF1B, HNF4A, HNF4G, HHEX, RFX6 and PDX1. CDX2 which is required for patterning of posterior endoderm also showed decreased expression in mutant cells.

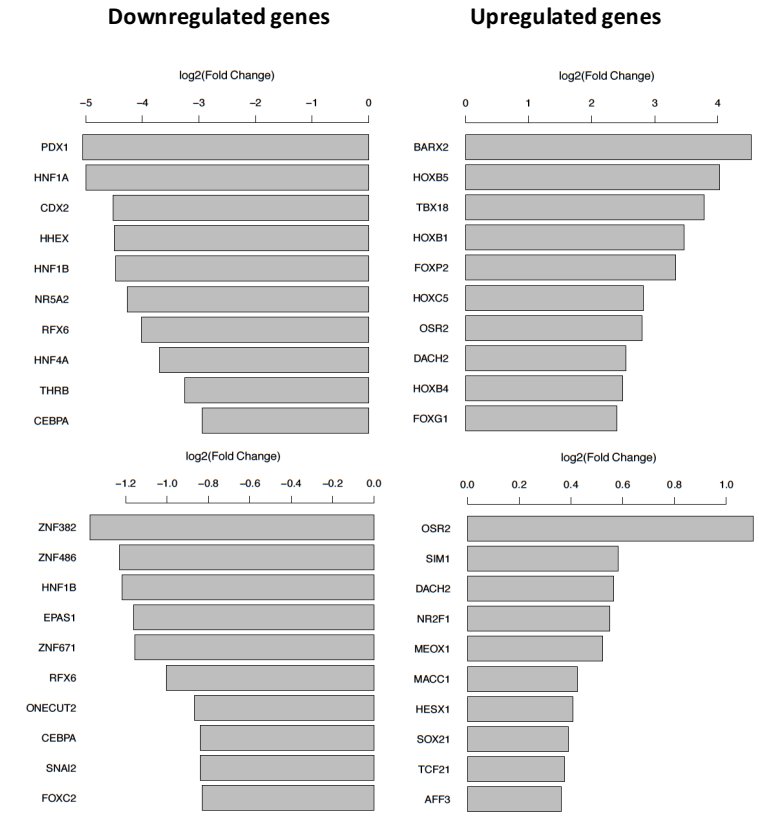
Up-regulated transcription factor genes include those involved in a variety of developmental processes such as; MEOX1, which is involved in somite development, HOXB1, HOXB4, HOXB5 and HOXC1, part of the HOX family of genes which play an important role in regional patterning of the gut tube / neuronal, neural crest, endodermal and mesodermal derivatives, TCF21, which is involved in lung, gut and kidney development and PLXNA2 which is involved in axon and cell guidance during neuronal development.

At the pancreatic progenitor stage (day 13), consistent with RT-qPCR and FACS results, HNF1B$^{-/-}$ and HNF1B$^{+/-}$ cells showed decreased expression of pancreatic signature genes i.e. down-regulation of PDX1, SOX9, PTF1A, NKX6-1, RFX6, ONECUT2 (HNF6B), HNF1B, HNF1A and HNF4A. At the endocrine progenitor (day 16) and hormonal cell / β-like cell stage (D27), transcription factors known to be important in regulation of endocrine cell differentiation and β-cell development were down-regulated including NEUROD1, PAX4, ARX, PDX1 and NR5A2 as well as HNF1B, HNF1A and HNF4A.

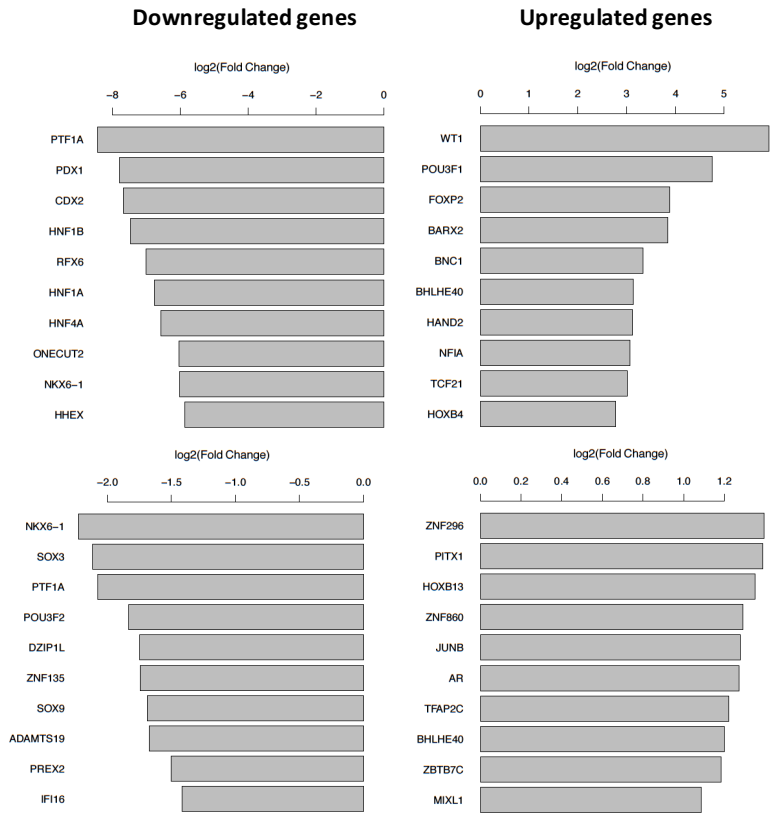
A **Day 6**


Downregulated genes
Upregulated genes
log2(Fold Change)
−8 −6 −4 −2 0
FOXQ1
PDX1
CEBPA
HNF4A
HNF1B
ZFP42
CDX2
RFX6
HNF4G
TFAP2C
0 1 2 3 4
BARX2
OSR2
HOXB5
POU3F1
MEOX1
PLXNA2
HOXB4
HOXB1
SIM1
NR2F1
HNF1B-/-
−1.5 −1.0 −0.5 0.0
EPAS1
ZNF382
ZNF486
FOXQ1
ZNF671
ID2
HNF1B
MAF
HHEX
PITX2
0.0 0.2 0.4 0.6 0.8
OSR1
HOXB1
NR2F1
MEOX1
PLXNA2
KLF4
FOXI3
SIX1
HOXB4
TCF21
HNF1B+/-

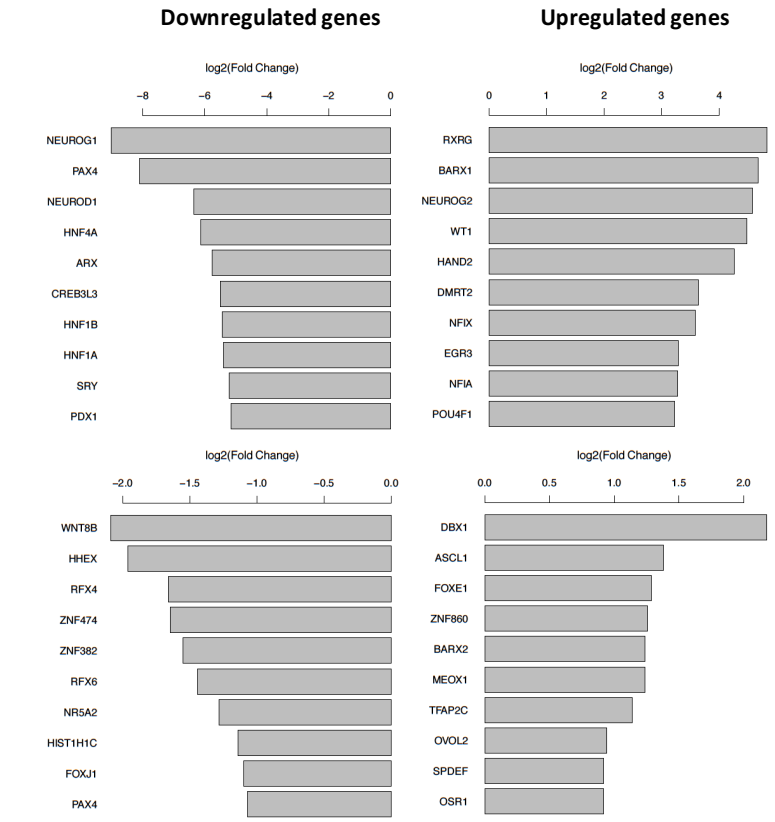
B **Day 8**


Downregulated genes
Upregulated genes
log2(Fold Change)
−5 −4 −3 −2 −1 0
PDX1
HNF1A
CDX2
HHEX
HNF1B
NR5A2
RFX6
HNF4A
THRB
CEBPA
0 1 2 3 4
BARX2
HOXB5
TBX18
HOXB1
FOXP2
HOXC5
OSR2
DACH2
HOXB4
FOXG1
HNF1B-/-
−1.2 −1.0 −0.8 −0.6 −0.4 −0.2 0.0
ZNF382
ZNF486
HNF1B
EPAS1
ZNF671
RFX6
ONECUT2
CEBPA
SNAI2
FOXC2
0.0 0.2 0.4 0.6 0.8 1.0
OSR2
SIM1
DACH2
NR2F1
MEOX1
MACC1
HESX1
SOX21
TCF21
AFF3
HNF1B+/-

C **Day 13**



D **Day 16**



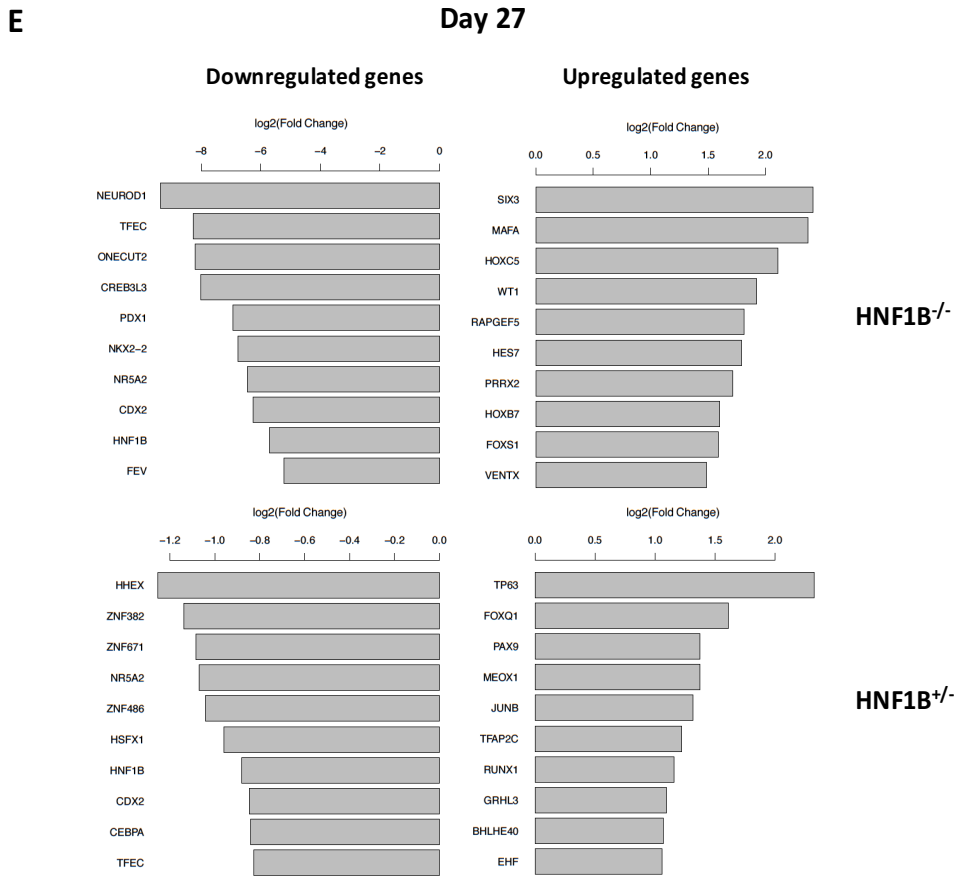


Figure 5.5. Barplots showing the top 10 transcription factors which are down- or up-regulated in HNF1B$^{-/-}$ vs HNF1B$^{+/+}$ and HNF1B$^{+/-}$ vs HNF1B$^{+/+}$ cells at **(A)** day 6 (foregut progenitors), **(B)** day 8 (posterior foregut), **(C)** day 13 (pancreatic progenitors), **(D)** day 16 (endocrine progenitors) and **(E)** day 27 (hormonal cells / β-like cells) of differentiation.

5.2.1.5 Gene ontology and Kyoto Encyclopaedia of Genes and Genomes pathway analysis of differentially expressed genes

Gene ontology (GO) for biological processes (BP) and Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway analysis was performed using the edgeR limma package. The top 15 enriched GO(BP) terms and KEGG pathway for down and up-regulated genes at each stage of the differentiation process between $HNF1B^{-/-}$ and $HNF1B^{+/+}$ cells and $HNF1B^{+/-}$ and $HNF1B^{+/+}$ cells is shown in (Figure 5.6). There were not enough up-regulated genes at day 6 or day 8 or down-regulated genes at day 8 or day 27 between $HNF1B^{+/-}$ and $HNF1B^{+/+}$ cells for GO or KEGG pathway analysis to be performed.

As expected, GO and KEGG pathway analysis of down-regulated genes in $HNF1B^{-/-}$ cells compared with $HNF1B^{+/+}$ cells revealed significant enrichment for terms associated with tube and digestive tract development at day 6. GO and KEGG pathway analysis of down-regulated genes at days 8 and 13 (pancreatic progenitor stages) revealed significant enrichment for terms associated with pancreas development, endocrine pancreatic development, pancreatic secretion, and maturity onset diabetes of the young. Other closely related biological processes including hepatobiliary and liver development were also significantly down-regulated in $HNF1B^{-/-}$ compared with $HNF1B^{+/+}$ cells. HNF1B also plays an important role in embryonic kidney development and at day 13 down-regulated genes showed significant enrichment for the GO(BP) terms and KEGG pathways; kidney development, renal system development, urogenital system development, kidney epithelium development, renin-angiotensin system and renin secretion. At the endocrine cell stages of differentiation (days 16 and days 27), GO and KEGG pathway analysis of down-regulated genes showed significant enrichment for terms associated with pancreatic development and hormonal cell development and function including; maturity onset diabetes of the young, endocrine pancreas development, hormone secretion, regulation of secretion, pancreatic secretion and insulin secretion.

GO and KEGG pathway analysis revealed up-regulation of genes involved in nervous system development in $HNF1B^{-/-}$ cells including; nervous system development, generation of neurons, neurogenesis and axon guidance; but did not reveal up-regulation of genes involved in any particular developmental lineage in $HNF1B^{+/-}$ cells. This suggests that absence of HNF1B appears to affect foregut patterning, resulting in cell types with a mainly neuronal identity, and subsequent impairment of pancreas development. This effect may however be masked by the cell culture conditions which direct cells to differentiate along pancreatic lineage and would affect survival of other cell types. Use of alternative culture conditions could help to further

reveal the identity of the cells generated by the absence of HNF1B. GO and KEGG pathway analysis did not reveal any up-regulation of pluripotency genes or pathways in $HNF1B^{-/-}$ compared with $HNF1B^{+/+}$ cells.

Genes associated with biological process and pathways which are generally important for cell structure, motility and adhesion were significantly up or down-regulated in $HNF1B^{-/-}$ and $HNF1B^{+/-}$ cells compared with $HNF1B^{+/+}$ cells at each stage of the differentiation process. Of particular interest down-regulated genes showed enrichment for the GO terms and KEGG pathways; extracellular structure organisation, cell junction and cell adhesion molecules (CAMs) and up-regulated genes showed enrichment for the GO terms and KEGG pathways; regulation of cell adhesion, regulation of cell motility, regulation of locomotion, extracellular structure organisation, extracellular matrix organisation, biological adhesion, cell adhesion, regulation of cell migration, regulation of cell motility, axon guidance and focal adhesion. Up-regulated genes included in the KEGG pathways axon guidance and focal adhesion include; SEMA3A, SEMA3C, SEMA6D, SEMA3E, EFNA5, EPHA7, PLXNA2, DPYSL5, SLIT2, NTN1, THBS4, COL1A2, COL3A1 and TNC.

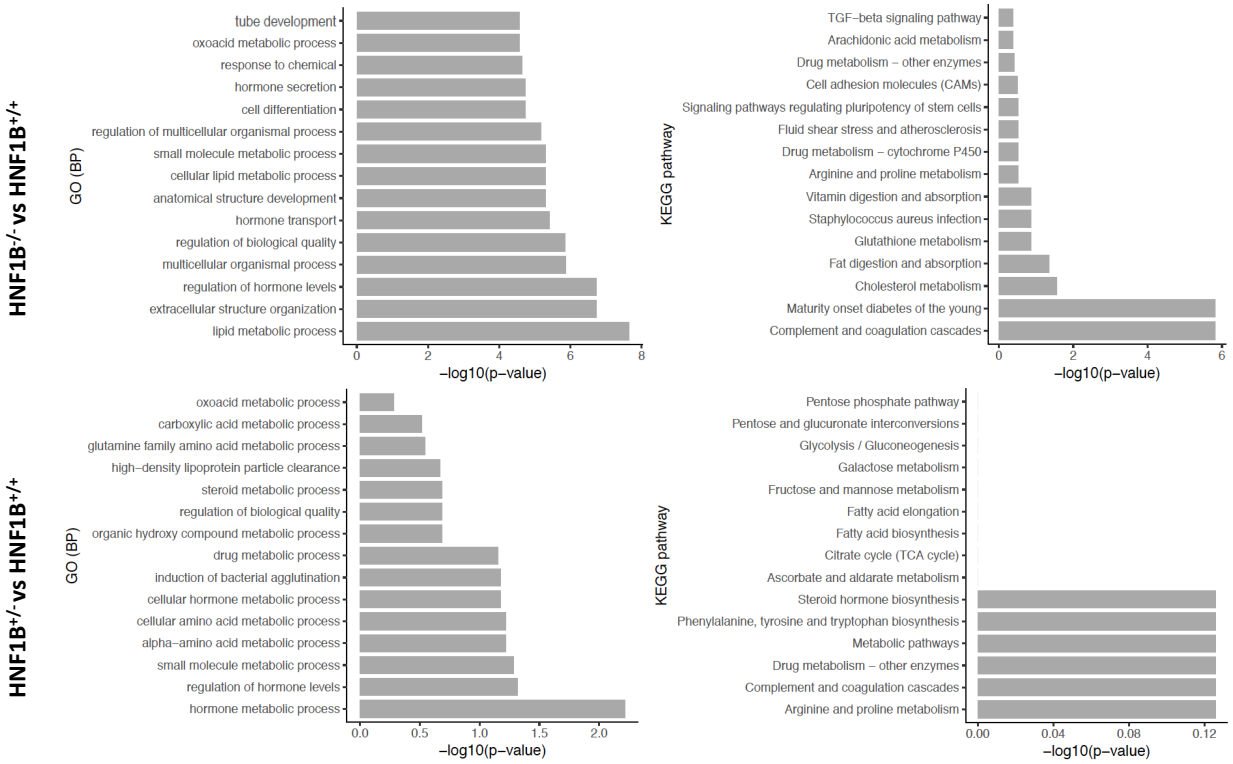
Up and down-regulated genes in $HNF1B^{-/-}$ and $HNF1B^{+/-}$ cells compared with $HNF1B^{+/+}$ cells were also significantly enriched for GO terms and KEGG pathways associated with various cellular metabolic processes and pathways. Interestingly, KEGG pathway analysis showed that genes associated with the peroxisome proliferator-activated receptor (PPAR) signalling pathway were significantly down-regulated at days 13, 16 and 27 of the differentiation process, particularly in $HNF1B^{-/-}$ cells. Peroxisome proliferator-activated receptors (PPARs) compose a family of three nuclear receptors (PPARalpha, PPARbeta/delta, PPARgamma) which act as lipid sensors to modulate gene expression. PPAR pathways are known to play an important role in lipid metabolism and cell proliferation (Feige et al., 2006). In the pancreas, PPAR-γ has been shown to directly regulate key β cell genes involved in glucose sensing, insulin secretion and insulin gene transcription (Gupta et al., 2010).

Previous experiments described in chapter 4 revealed significant reductions in cell proliferation in $HNF1B^{+/-}$ and $HNF1B^{-/-}$ cells at day 6 and day 13 of the differentiation process. Genes associated with the GO terms and KEGG pathways; cell proliferation, negative regulation of cell proliferation, mitotic cell cycle, chromosomal segregation and cell cycle were significantly differentially expressed in $HNF1B^{-/-}$ and $HNF1B^{+/-}$ cells compared with $HNF1B^{+/+}$ cells at days 6, 13 and 16. Genes involved in cell cycle regulation and proliferation which are differentially

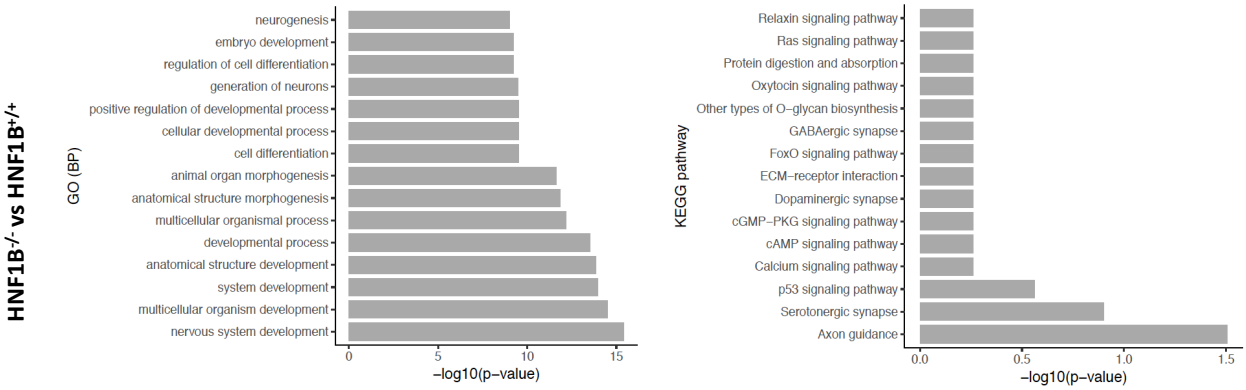
expressed (log2FC > 1, adj p-value < 0.01) between HNF1B$^{+/-}$ and HNF1B$^{-/-}$ compared with HNF1B$^{+/+}$ cells, were also determined by comparing the list of up-regulated and down-regulated genes for each condition, with a list of cell cycle genes obtained from the Cyclebase 3.0 database (Santos et al., 2015). 45 genes were significantly up-or down-regulated in HNF1B$^{+/-}$ and HNF1B$^{-/-}$ at day 6 and 99 genes were significantly up-or down-regulated in HNF1B$^{+/-}$ and HNF1B$^{-/-}$ at day 13. At day 6, 12 genes were associated with the GO terms regulation of cell proliferation (EGR1, NGFR, ATOH8, CNTFR, PGF, B4GALT1, LIF, FOSL1, FOSL2, AHR, CLU and ASPH), however their exact function in the developing foregut is unknown. At day 13, 12 genes were associated with the GO terms regulation of cell proliferation (LIF, EGR1, CYR61, MYC, CTSH, NFIB, NOX4, ITGAV, CDKN2A, CLU, ASPH, CNTFR, TNFRSF21 and NDFIP1).

A

Day 6 – down-regulated GO biological processes and KEGG pathways

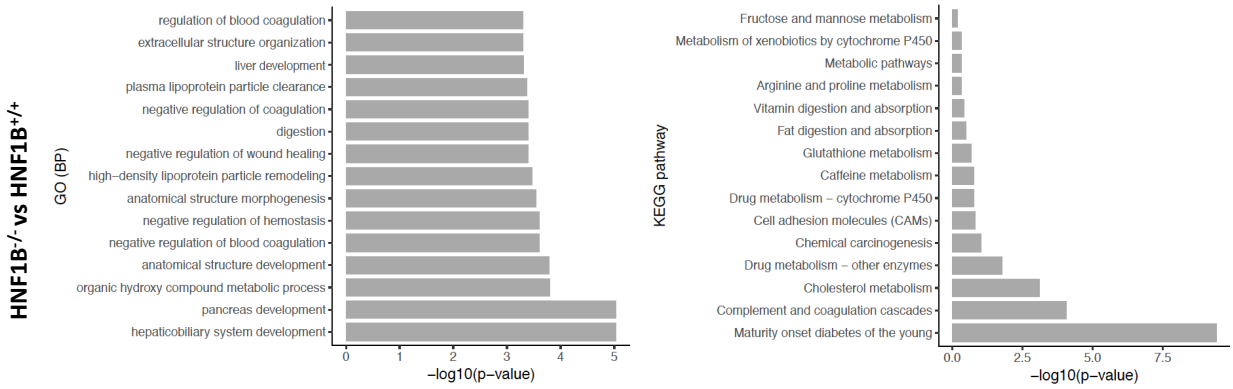


Day 6 – up-regulated GO biological processes and KEGG pathways

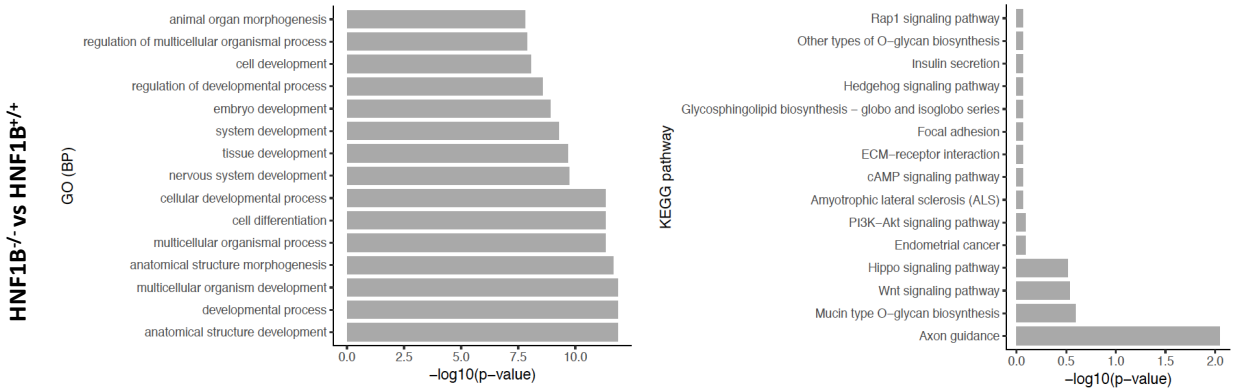


B

Day 8 – down-regulated GO biological processes and KEGG pathways

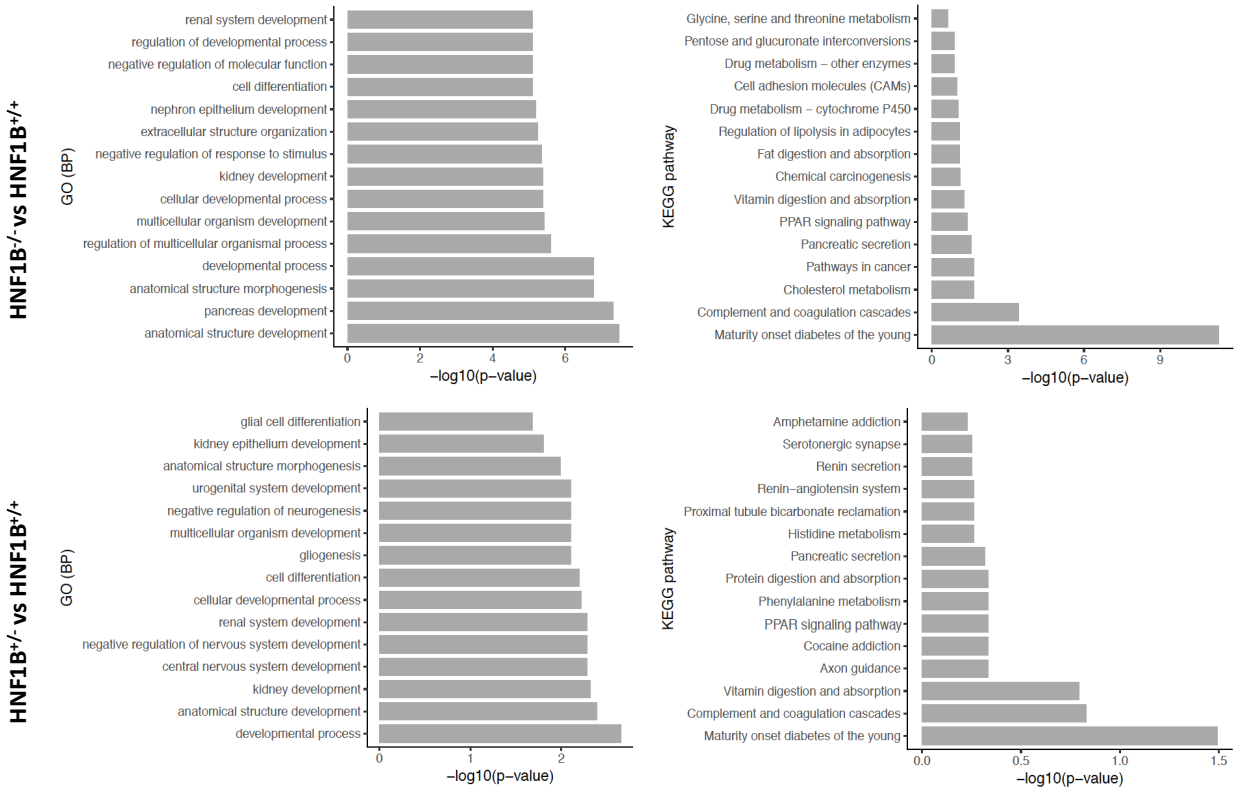


Day 8 – up-regulated GO biological processes and KEGG pathways

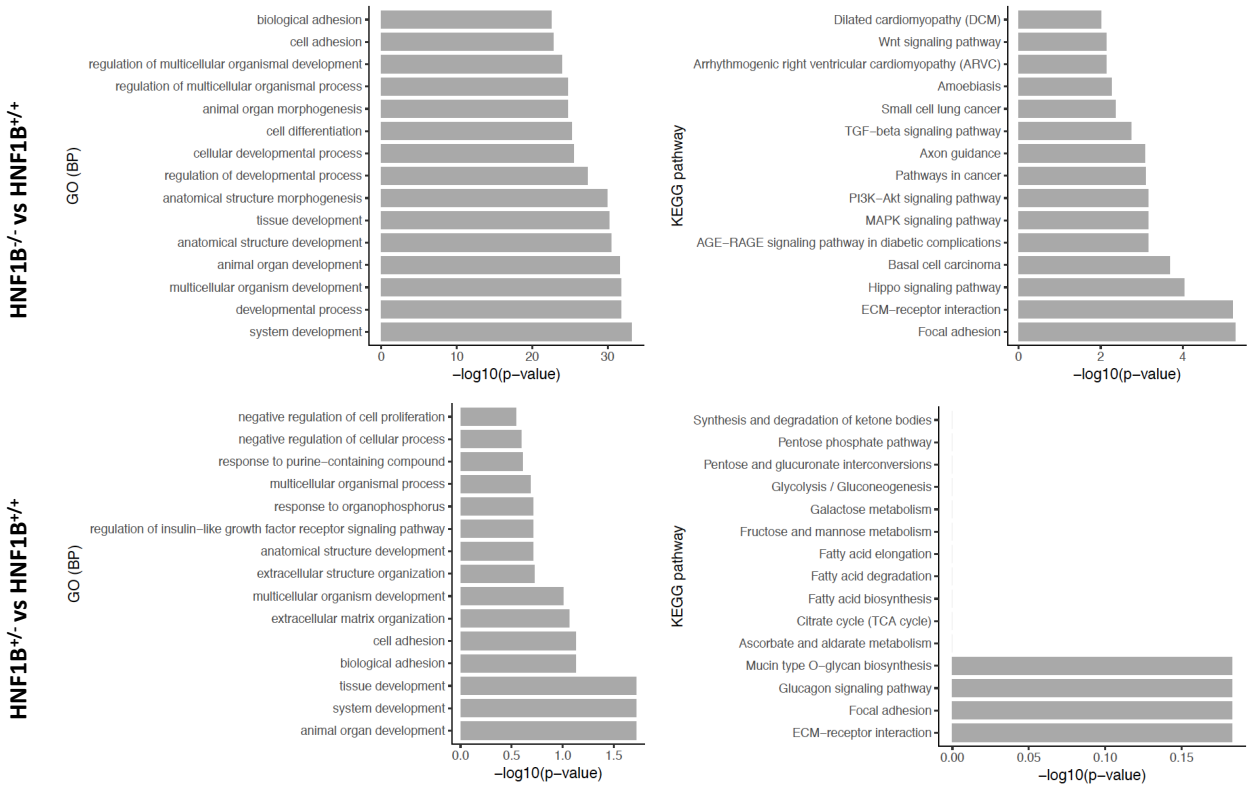


C

Day 13 – down-regulated GO biological processes and KEGG pathways

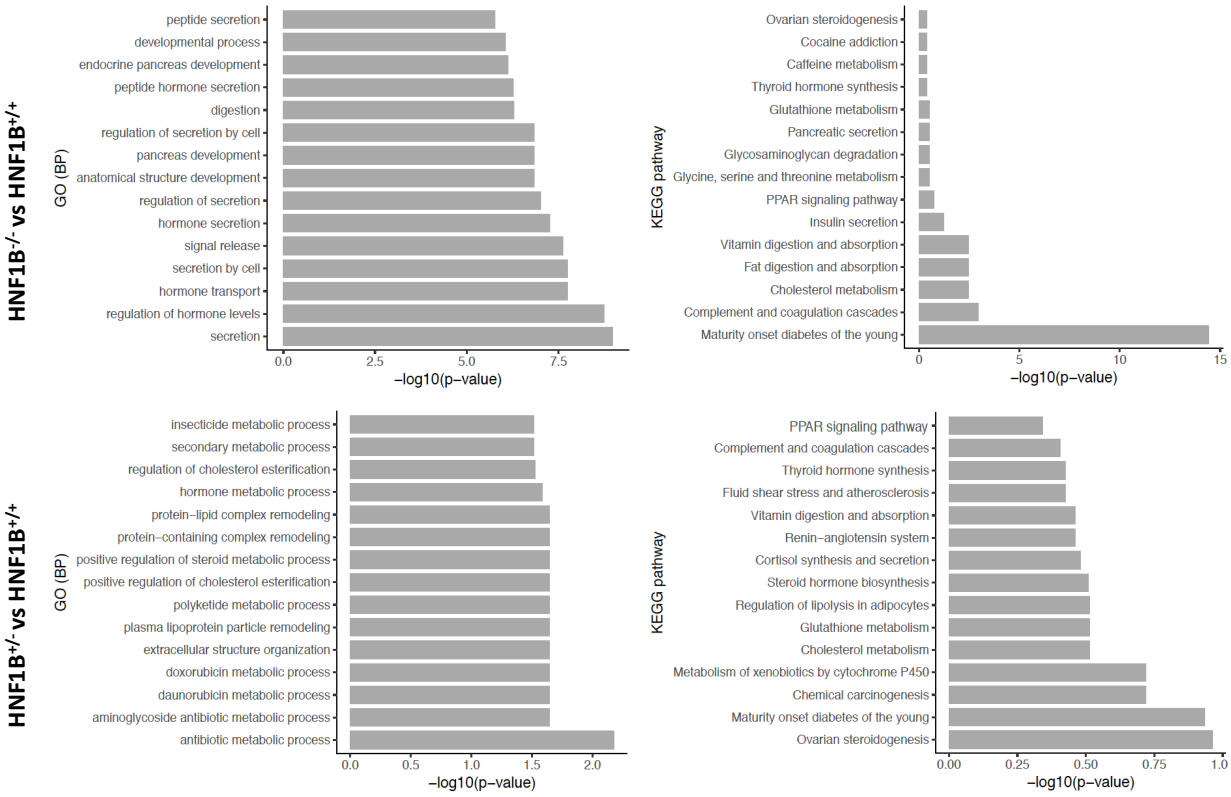


Day 13 – up-regulated GO biological processes and KEGG pathways

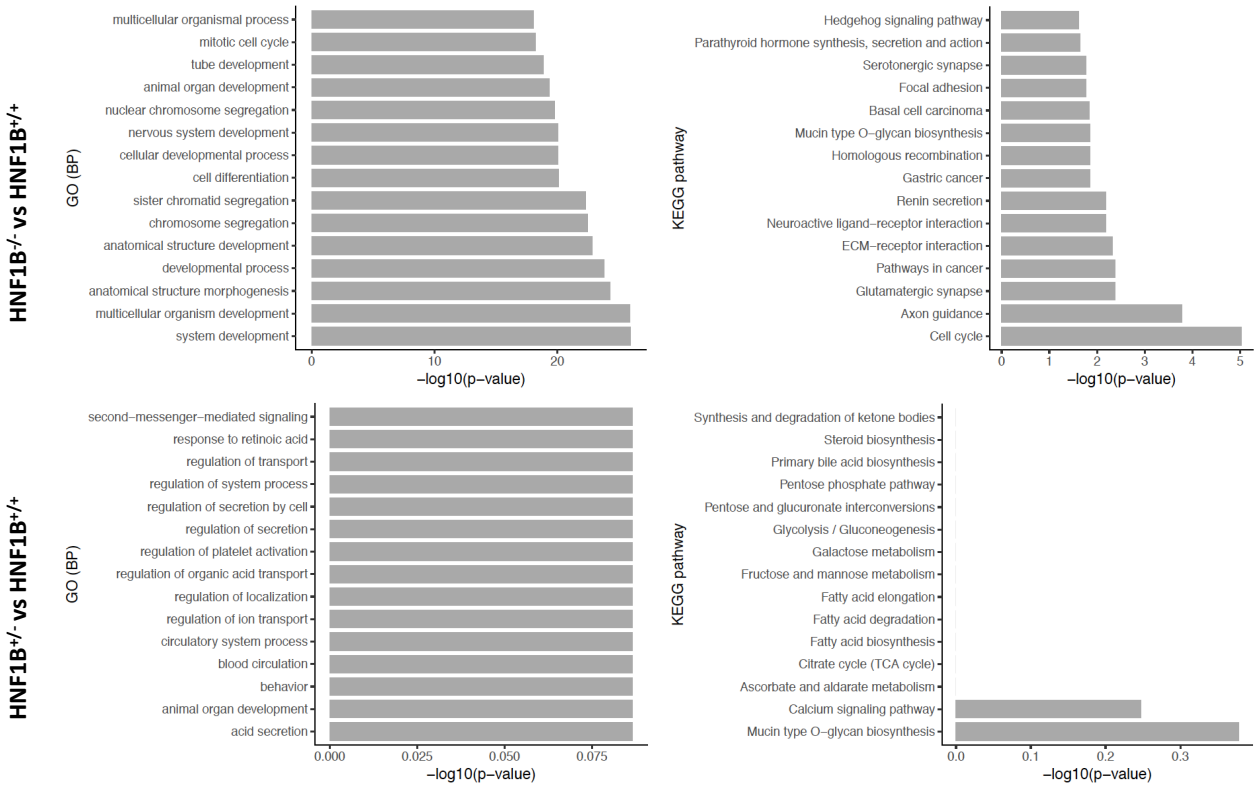


D

Day 16 – down-regulated GO biological processes and KEGG pathways



Day 16 – up-regulated GO biological processes and KEGG pathways



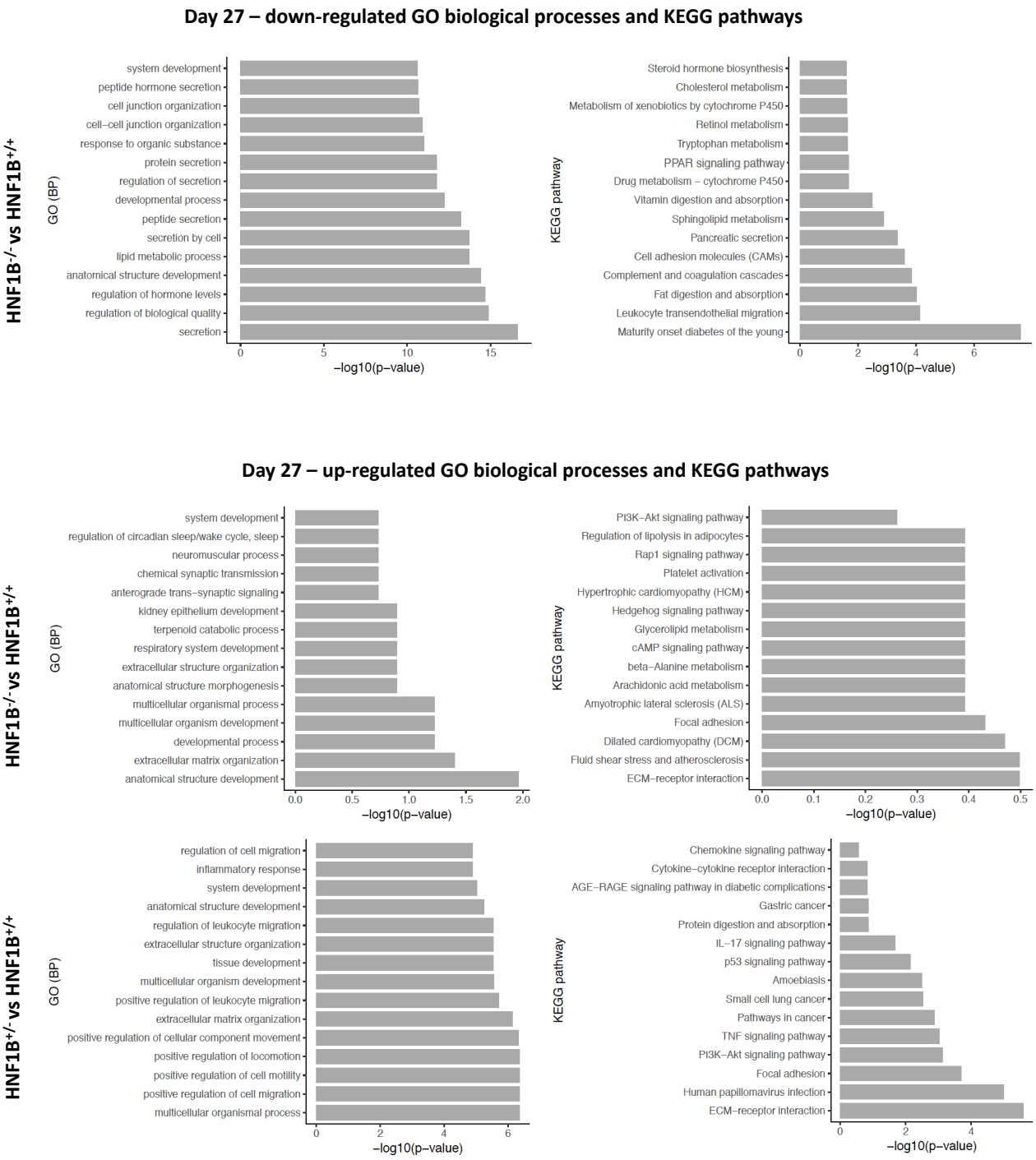


Figure 5.6. Barplots showing the top 15 down and up-regulated gene ontology biological process (GO (BP)) and KEGG pathways in HNF1B$^{-/-}$ vs. HNF1B$^{+/+}$ and HNF1B$^{+/-}$ vs. HNF1B$^{+/+}$ cells at: (A) day 6, (B) day 8, (C) day 13, (D) day 16 and (E) day 27.

5.2.2 Identification of gene targets directly regulated by HNF1B

Ideally, at this stage I would have liked to integrate gene expression data from RNA-seq with HNF1B binding regions from ChIP-seq experiments from $HNF1B^{+/+}$, $HNF1B^{+/-}$ and $HNF1B^{-/-}$ cells at day 6 and day 13 in order to identify genes which are bound and regulated by HNF1B. However, unfortunately the ChIP-seq experiments I performed were not successful. The HNF1B ChIP samples were of poor quality with high percentage of PCR duplicates (> 60%), resulting in low numbers of good quality mapped reads (< $50x10^6$ per sample) and poor reproducibility between samples.

Instead, HNF1B ChIP-seq data from day 12 pancreatic progenitor cells derived from hESCs was obtained through our collaboration with the Ferrer group. The unprocessed data in FASTQ format is publicly accessible through ArrayExpress E-MATB-1990 (Cebola et al., 2015). 2332 peaks were found, associated with 1965 unique gene names. Motif finding was performed using HOMER software (Heinz et al., 2010). Statistical cut-offs used for significance were p-value < 0.001 and log2 enrichment ratio > 2. Significant findings from Motif enrichment analysis of HNF1B ChIP-seq in pancreatic progenitor cells are shown in Figure 5.7. and as a table in Appendix 8. As expected, HNF1B binding was most significantly enriched at the HNF1B and HNF1A loci, since HNF1B is known to form homo and hetero-dimers (Mendel and Crabtree, 1991, Mendel et al., 1991). HNF1B binding was also significantly enriched at the loci of several genes known to be involved in pancreas development including HNF4A, ONECUT1, SOX9, RFX6, RREB1 and TEAD1. Both RT-qPCR and RNA-seq show dose-dependent effects of HNF1B on HNF4A and SOX9 expression levels in $HNF1B^{+/-}$ and $HNF1B^{-/-}$ mutant cells.

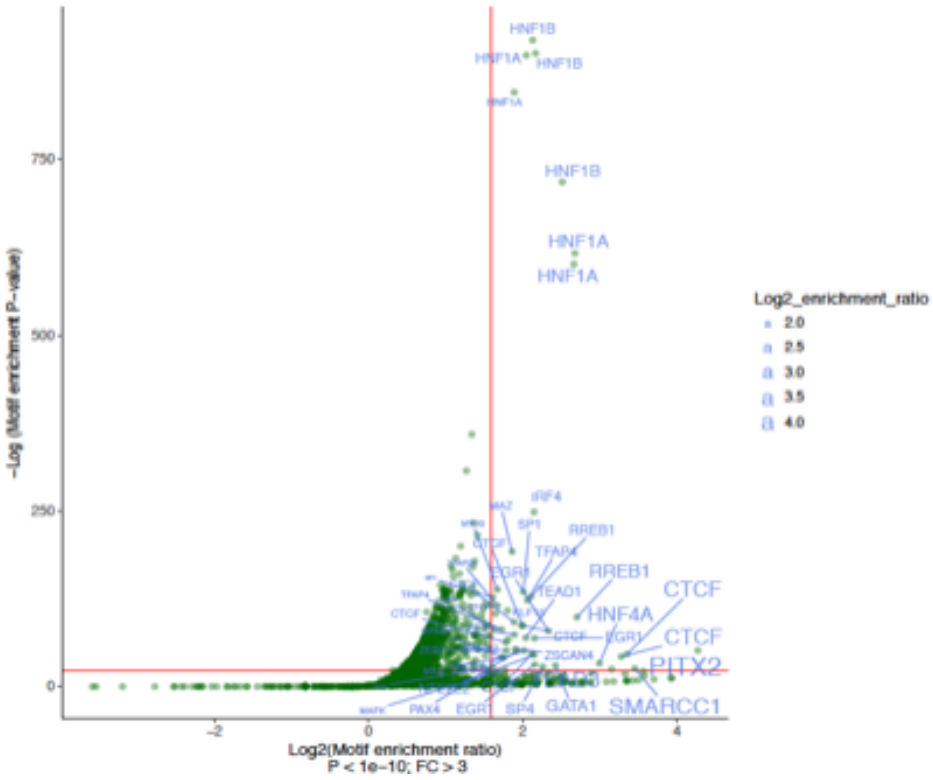


Figure 5.7. Half volcano plot showing significant findings from Motif enrichment analysis of HNF1B ChIP-seq at day 12 using Homer. Gene loci with significantly enriched HNF1B binding sites are shown. Statistical cut-offs used for significance were p-value < 0.001 and log2 enrichment ratio > 2. A table showing this data is included in Appendix 8.

We next compared our RNA-seq and ChIP-seq data sets in order to identify genes that are bound and regulated by HNF1B. This analysis revealed that the only genes which were bound by HNF1B in wild-type pancreatic progenitors, and which were significantly down-regulated in HNF1B$^{+/-}$ mutant cells were; HNF1B, SOX9 and RFX6. In contrast 33 significantly expressed genes (fpkm>1) were bound by HNF1B and down-regulated in HNF1B$^{-/-}$ iPSC-derived pancreatic progenitor cells and 20 genes were up-regulated in HNF1B$^{-/-}$ cells. Down-regulated genes included several genes known to be involved in pancreatic development including; RFX6, HNF1A, HNF4A, SOX9, ONECUT1 and HNF4G. Several of these genes are also important in liver development. Up-regulated genes did not include any combination of genes known to be important for differentiation of cells along another organ lineage.

The RNA-seq and ChIP-seq data therefore suggest that HNF1B acts in a transcriptional network with other genes to regulate development of pancreatic progenitor cells. HNF1B can be recruited to consensus binding sites in the promotor and/or enhancer regions of several genes in pancreatic progenitor cells and can directly regulate expression of these genes.

5.2.3 Genome wide chromatin accessibility effects of HNF1B

ATAC-seq was used to determine chromatin accessibility in HNF1B$^{+/+}$ compared with HNF1B$^{+/-}$ and HNF1B$^{-/-}$ cells at day 6 and day 13 of the pancreatic differentiation process. ATAC-seq sample information together with the total number of aligned fragments and mapped reads with quality score >10 are shown in Table 5.3. Biological duplicates were sequenced for each clone at day 6 and day 13 of differentiation. Peak-calling was performed using MACS2 with "--nomodel –shift -100 --extsize 200 –q 0.01" to identify open chromatin regions (peaks) that were enriched for transposase integration sites at 1% FDR. IGV and Biodalliance (25) (26) were used for visualization of raw intensities.

With these parameters 163,065 to 345,243 peaks per sample were identified. Differentially accessible chromatin regions were identified using diffReps (24). To determine how chromatin accessibility changes, we assessed quantitative differences in ATAC-seq signal intensities between HNF1B wild-type and mutant cells. Regions were defined as differentially accessible if ATAC-seq signal intensity showed fold change > 1.5 with an FDR <0.001.

Sample name	Day	HNF1B Level	Library size	Mapped reads Quality score >10
13B Cl.45 D6_1	06	WT	178050871	134560375
13B Cl.6 D6_1	06	Het	161621261	121847813
13B Cl.48 D6_1	06	Hom	144531444	94317036
13B Cl.45 D6_2	06	WT	135153739	108090749
13B Cl.6 D6_2	06	Het	118280500	80318505
13B Cl.48 D6_2	06	Hom	107841494	66533633
13B Cl.45 D13_1	13	WT	154422271	107397519
13B Cl.6 D13_1	13	Het	117339898	93620509
13B Cl.48 D13_1	13	Hom	88548774	83245808
13B Cl.45 D13_2	13	WT	191555535	95173495
13B Cl.6 D13_2	13	Het	207894080	131708525
13B Cl.48 D13_2	13	Hom	176659724	110742570

Table 5.3. General information about ATAC-seq samples showing total number of aligned fragments and the number of mapped reads with quality score >10

At day 6, 164 chromatin regions showed decreased accessibility i.e. “gained-closed” and 332 regions showed increased accessibility i.e. “gained-open” in HNF1B$^{+/-}$ cells compared with HNF1B$^{+/+}$ cells and 2914 regions showed decreased chromatin accessibility and 1614 regions showed increased accessibility in HNF1B$^{-/-}$ cells compared with HNF1B$^{+/+}$ cells. At day 13, 50 regions showed decreased accessibility and 39 regions showed increased accessibility in HNF1B$^{+/-}$ cells compared with HNF1B$^{+/+}$ cells and 1809 regions were showed decreased accessibility and 1090 regions were showed increased accessibly in HNF1B$^{-/-}$ cells compared with HNF1B$^{+/+}$ cells. This data suggests that the overall effect of HNF1B is to bind and open chromatin to allow recruitment of other transcription factors, which is well known function for genes with a HNF/FOXA domain (Kaestner, 2010).

Major differences in chromatin accessibility were detected in the HNF1B$^{-/-}$ cells compared with HNF1B$^{+/+}$ cells at both the foregut (days 6) and pancreatic progenitor (day 13) stage of the differentiation process. A preference for down-regulation of chromatin accessibility was observed with about double the number of regions compacting. However, in HNF1B$^{+/-}$ cells the effect on chromatin accessibility was smaller, with more changes at day 6 compared with day 13. This suggests that HNF1B haploinsufficiency has a greater impact on chromatin organisation at an earlier stage of development and the effects on pancreatic progenitor differentiation may be independent of chromatin change.

Genes located within a 10kb window of an observed chromatin change were identified and our RNA-seq data was used to look at the expression levels for these genes and to determine if these genes were up or down regulated. At day 6, for regions of gained-closed chromatin, 13 genes located within a 10kb window were down-regulated and none were up-regulated in HNF1B$^{+/-}$ compared with HNF1B$^{+/+}$ cells. These genes included genes known to be important in foregut and pancreas development including HNF1A, HNF4A, ONECUT1, GATA6 and PDX1. No genes located within a 10kb window were up-regulated in HNF1B$^{+/-}$ compared with HNF1B$^{+/+}$ cells. At day 6, for regions of gained-open chromatin, no genes located within a 10kb window were differentially expressed in HNF1B$^{+/-}$ compared with HNF1B$^{+/+}$ cells. At day 13, for regions of gained-closed chromatin, only 2 genes located within a 10kb window were down-regulated (DDC and PLEKHA6) and none were up-regulated in HNF1B$^{+/-}$ compared with HNF1B$^{+/+}$ cells. At day 13, for regions of gained-opened chromatin, no genes located within a 10kb window were up or down-regulated.

At day 6, for regions of gained-closed chromatin, 180 genes located within a 10kb window were differentially expressed in HNF1B$^{-/-}$ compared with HNF1B$^{+/+}$ cells. 140 of these genes were down-regulated and 40 genes were up-regulated. At day 6, for regions of gained-open chromatin, 64 genes located within a 10kb window were differentially expressed in HNF1B$^{-/-}$ compared with HNF1B$^{+/+}$ cells. 20 of these genes were down-regulated and 44 genes were up-regulated. At day 13, for regions of gained-open chromatin, 179 genes located within a 10kb window were differentially expressed in HNF1B$^{-/-}$ compared with HNF1B$^{+/+}$ cells. 67 genes were up-regulated and 112 were down-regulated. At day 13, for regions of gained-closed chromatin, 79 genes located within a 10kb window were differentially expressed in HNF1B$^{-/-}$ compared with HNF1B$^{+/+}$ cells. 53 genes were up-regulated and 26 were down-regulated.

Motif analysis in differential open chromatin accessibility regions was performed using HOMER software (27). Significantly enriched motifs are shown in Figures 5.8 and 5.9 and all enriched motifs are shown in table form in Appendix 9. Expression levels of genes with enriched motifs in open or closed chromatin regions were determined using our RNA-seq data to determine if these genes were up or down regulated.

At the primitive gut tube or foregut progenitor stage (day 6), comparison between HNF1B$^{-/-}$ and HNF1B$^{+/+}$ cells (2914 regions with reduced chromatin accessibility), revealed significant motif enrichment for several genes known to play an important role in foregut and pancreatic endoderm development such as HNF1B, HNF1A, HNF4G and HNF4A motifs. HNF1B forms homodimers and heterodimers with HNF1A (Mendel and Crabtree, 1991, Mendel et al., 1991). HNF4G is a member of the HNF4 member of TFs and has nearly identical DNA-binding and ligand-binding domains to HNF4A. HNF4G is known to be expressed in the pancreas, kidney, small intestine and testes in humans; and in the liver, although at lower levels than HNF4A (Drewes et al., 1996). Motifs for other endoderm lineage related factors such as RFX3, which forms heterodimers with RFX6 and ONECUT1 (HNF6) were also significantly enriched. Genes which showed significant motif enrichment and which were also significantly down-regulated based on our RNA-seq data at day 6 included; HNF4A, HNF1B and ONECUT1. Significantly up-regulated genes included; SIM1 and MEOX1 which are both transcription factors which may be involved in several embryonic developmental processes such as kidney development, neurogenesis, somite and blood cell development.

Decreased chromatin accessibility regions (164 regions) in HNF1B$^{+/-}$ compared with HNF1B$^{+/+}$ cells were also significantly enriched for HNF1B, HNF1A, HNF4G motifs. RNA-seq also

showed reduced expression of these genes in HNF1B$^{+/-}$ cells compared with HNF1B$^{+/+}$ cells at day 6, although the adjusted p-value did not reach significance. There were no genes which showed significant motif enrichment and which were significantly differentially expressed based on our RNA-seq data.

In regions of increased chromatin accessibility in HNF1B$^{-/-}$ cells (1614 regions) compared with HNF1B$^{+/+}$ cells at day 6, genes which showed significant motif enrichment and which were significantly down-regulated based on our RNA-seq data included; CDX2, RFX6, RFX2, HHEX and MECOM. These genes are mainly involved in transcriptional processes and in posterior axis specification. Genes which were significantly up-regulated based on our RNA-seq data included; NR2F1 and HOXB1. These transcription factors are known to mainly be involved in embryonic developmental processes such as nervous system development.

Similarly, regions with increased chromatin accessibility in HNF1B$^{+/-}$ cells (334 regions) compared with HNF1B$^{+/+}$ cells at day 6, showed enrichment of motifs for RRBE1 and STAT3. RREB1 is a transcription factor that binds specifically to the RAS-responsive elements (RRE) of gene promoters. A study by Lee et al. (Lee et al., 2012) showed increased expression of RREB1 in primitive gut tube cells differentiated from hESCs. Activating germline mutations in STAT3 were recently identified as a cause of neonatal diabetes mellitus associated with beta-cell autoimmunity (Flanagan et al., 2014b). More recently, a study by Saarimaki-Vire et al. has shown that an activating mutation in STAT3 can cause diabetes by causing premature endocrine differentiation through direct induction of NEUROG3 expression (Saarimaki-Vire et al., 2017). RNA-seq however did not show significant differential expression of these genes between HNF1B$^{+/-}$ and HNF1B$^{+/+}$ cells. Only one gene; EPAS1 showed significant motif enrichment and was significantly down-regulated between HNF1B$^{+/-}$ and HNF1B$^{+/+}$ cells. EPAS1 is a transcription factor involved in the induction of genes regulated by oxygen (hypoxia inducible factors). These genes mediate a metabolic rewiring of the hypoxic cell and play an important role in embryonic development (Dengler et al., 2014).

At day 13, regions of decreased chromatin accessibility in HNF1B$^{-/-}$ cells (1809 regions) compared with HNF1B$^{+/+}$ cells showed significant motif enrichment for several genes including some which are known to be important in pancreatic progenitor specification. Genes which showed significant motif enrichment and which were significantly down-regulated at day 13 based on our RNA-seq data included; PTF1A, HNF1B, HNF1A, HNF4A, PRDM16, SOX9,

SP8, MYC, EPAS1, RFX2, KLF4 and NR1H3 and genes which were significantly up-regulated included; SREBF1, RARG and HOXB7.

There were 1090 regions of increased chromatin accessibility in HNF1B$^{-/-}$ cells compared with HNF1B$^{+/+}$ cells at day 13. Genes which showed significant motif enrichment and which were significantly down-regulated at day 13 based on our RNA-seq data included; PTF1A, RFX6, SOX9 and RFX2. Genes which showed significant motif enrichment and which were significantly up-regulated at day 13 based on our RNA-seq data included; FOXO1, NR3C1, JUN and TFAP2C.

The effect of HNF1B haploinsufficiency on chromatin was milder at day 13 compared with day 6. There were 50 regions of decreased chromatin accessibility in HNF1B$^{+/-}$ cells compared with HNF1B$^{+/+}$. Only one gene; SOX9; showed significant motif enrichment and was significantly down-regulated at day 13. Genes which showed significant motif enrichment, and which were significantly up-regulated at day 13 based on our RNA-seq data included; MIXL1 and AR. There were 39 regions with increased chromatin accessibility in HNF1B$^{+/-}$ cells compared with HNF1B$^{+/+}$ cells at day 6. Genes which showed significant motif enrichment and which were significantly down-regulated at day 13 based on our RNA-seq data included; PTF1A, SOX9, STAT4 and NR6A1. Only one gene, JUN, showed significant motif enrichment and was significantly up-regulated at day 13 based on our RNA-seq data.

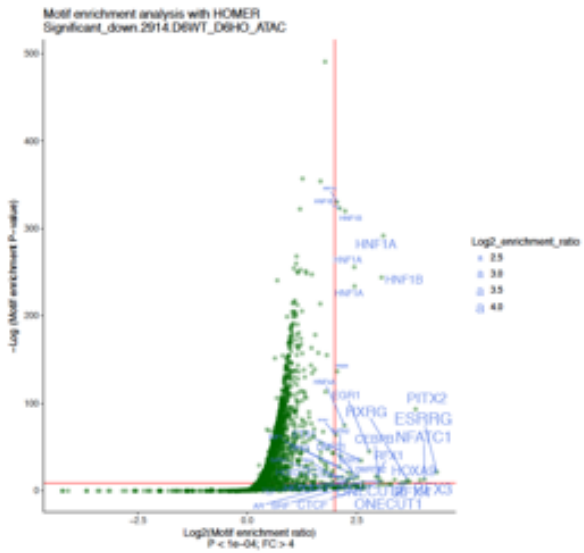
Data from the ATAC-seq experiments was integrated with the ChIP-seq dataset to further investigate the effects of HNF1B binding on chromatin accessibility. Genes which showed significant motif enrichment in regions of decreased and increased chromatin accessibility from the ATAC-seq experiments at day 13 were compared with gene loci with significantly enriched HNF1B binding from the HNF1B ChIP-seq data at day 13. For regions of decreased chromatin accessibility in HNF1B$^{-/-}$ compared with HNF1B$^{+/+}$ day 13 cells; genes which showed significant motif enrichment and enriched HNF1B binding included; HNF1A, HNF1B, HNF4A, SOX9, IRF4, SP1, CTCF, PITX2, E2F1, E2F4, RFX2, RFX1, TBR1, GATA1 and POU2F1. For regions of increased chromatin accessibility in HNF1B$^{-/-}$ compared with HNF1B$^{+/+}$ day 13 cells; genes which showed significant motif enrichment and enriched HNF1B binding included; SOX9, SP1, RFX6, RFX2, HSF1, CTCF, TFAP4, GATA2, ESRRA, E2F1 and POU2F1.

For regions of decreased chromatin accessibility in HNF1B$^{+/-}$ compared with HNF1B$^{+/+}$ day 13 cells; genes which showed significant motif enrichment and enriched HNF1B binding included; SOX9, HNF4A, HOXB9, TBR1, TFAP4, AR, CTCF, NFYA, NFYB, EGR1, TEAD1, SP1, SOX21, RREB1, SOX8, CEBPZ, TBX20, RFX2, ZBTB4, POU2F1 and SOX18. For regions of increased chromatin accessibility in HNF1B$^{-/-}$ compared with HNF1B$^{+/+}$ day 13 cells; genes which showed significant motif enrichment and enriched HNF1B binding included; RREB1, POU2F1, SOX8, TBP, ESRRA, RXRA, SOX9, HOXA13, EGR1, SP4, ALX3, SP1, HOXB13, TEAD1, GATA1, ZNF524, TBX20, IRF4, MAX and NFATC1.

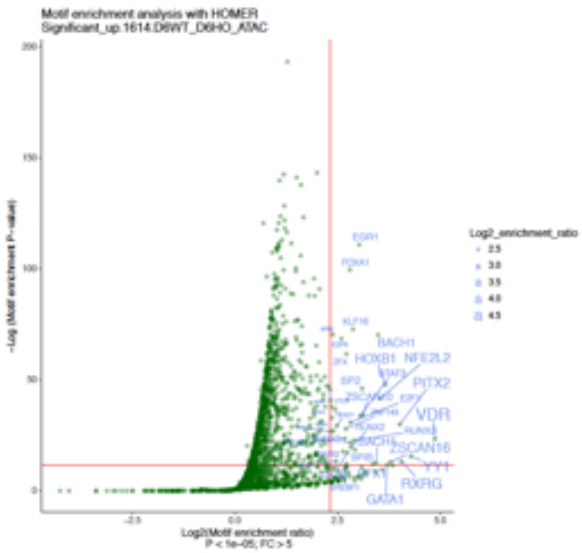
Several of these genes show very low expression (fpkm<1) in day 13 cells including; HOXA13, GATA1, TBR1, HOXB9, IRF4, GATA2, TBX20, ALX3, NFTAC1 and HOXB13. Of the remaining genes; those showing significant differential expression between HNF1B$^{-/-}$ and HNF1B$^{+/+}$ cells included; HNF4A, HNF1B, SOX9, RFX6, HNF1A, SOX8, RXRA and RFX2 and those showing significant differential expression between HNF1B$^{+/-}$ and HNF1B$^{+/+}$ cells included; HNF4A, HNF1B, SOX9, RFX6 and AR.

Day 6 – Foregut Progenitor cells

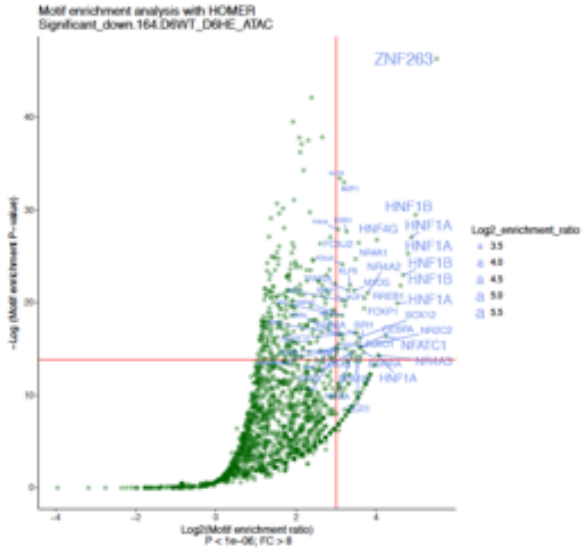
HNF1B$^{+/+}$ vs. HNF1B$^{-/-}$
Decreased chromatin accessibility



HNF1B$^{+/+}$ vs. HNF1B$^{-/-}$
Increased chromatin accessibility



HNF1B$^{+/+}$ vs. HNF1B$^{+/-}$
Decreased chromatin accessibility



HNF1B$^{+/+}$ vs. HNF1B$^{+/-}$
Increased chromatin accessibility

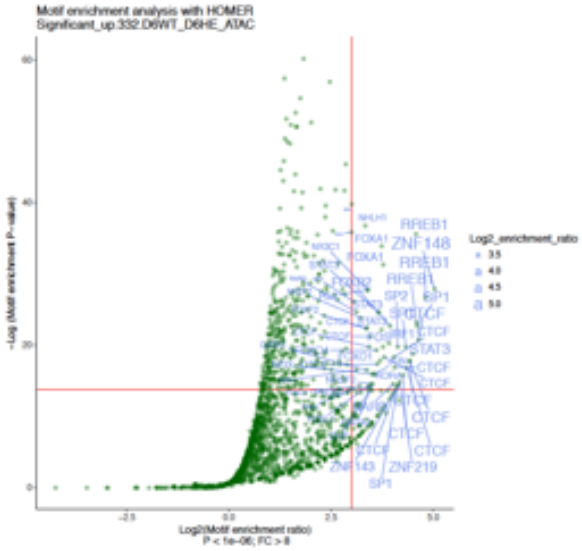
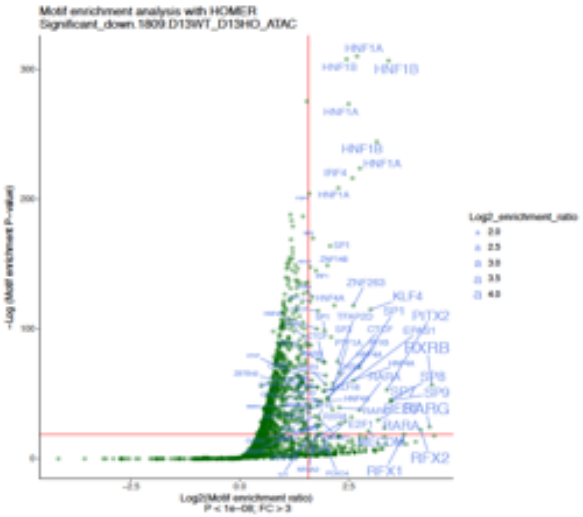


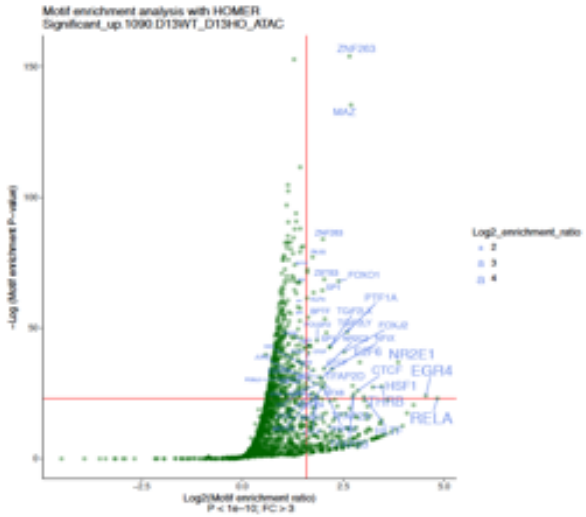
Figure 5.8. Half volcano plots showing significantly enriched motifs in regions of increased and decreased chromatin accessibility at day 6 of differentiation in HNF1B$^{-/-}$ and HNF1B$^{+/-}$ cells compared with HNF1B$^{+/+}$ cells. Data is shown in table form in Appendix 9.

Day 13 – Pancreatic Progenitor Cells

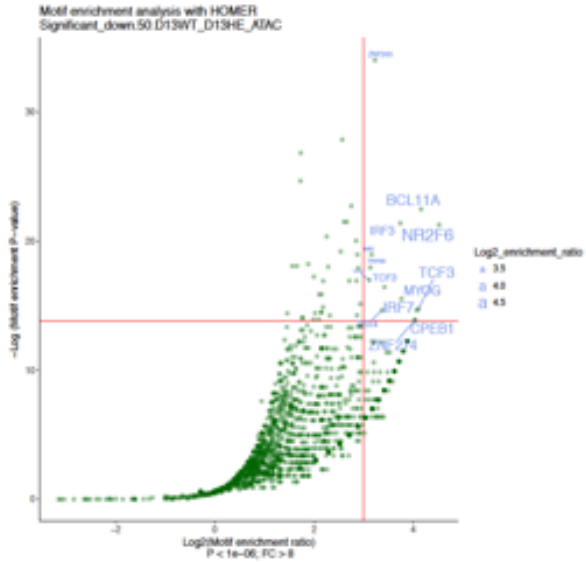
HNF1B$^{+/+}$ vs. HNF1B$^{-/-}$
Decreased chromatin accessibility



HNF1B$^{+/+}$ vs. HNF1B$^{-/-}$
Increased chromatin accessibility



HNF1B$^{+/+}$ vs. HNF1B$^{+/-}$
Decreased chromatin accessibility



HNF1B$^{+/+}$ vs. HNF1B$^{+/-}$
Increased chromatin accessibility

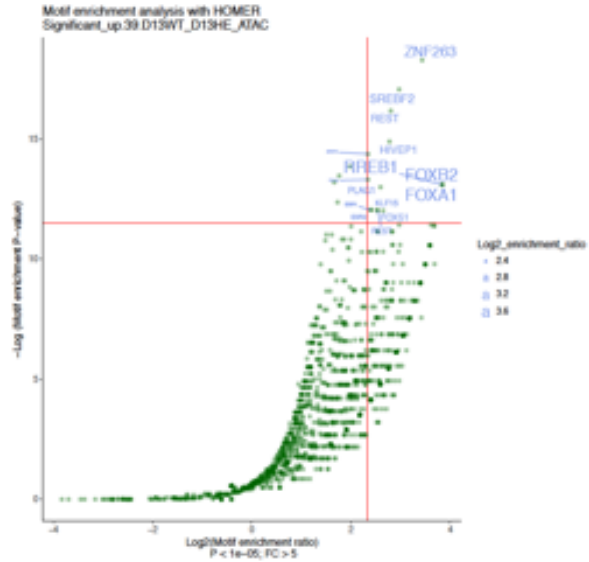


Figure 5.9. Half volcano plots showing significantly enriched motifs in regions of increased and decreased chromatin accessibility at day 13 of differentiation in HNF1B$^{-/-}$ and HNF1B$^{+/-}$ cells compared with HNF1B$^{+/+}$ cells. Data is shown in table form in Appendix 9.

5.2.4 HNF1B transcription factor binding dynamics

Binding of transcription factors (TFs) within regulatory DNA forms the basis for gene regulation. DNA sequences directly occupied by DNA-binding proteins are protected from transposition and the resulting sequence 'footprint' can therefore reveal the presence of a DNA-binding protein at a particular site. Presence of footprint infers but does not prove TF occupancy. A footprint appears as a deep notch of ATAC-seq signal / Tn5 frequencies, which can be obtained using NucleoATAC (36). The average profile of differential footprints (score > 40), showing the different distribution of insertions around the footprints are shown in Figures 5.10. and 5.11. Wellington Bootstrap (37), was used to identify unique footprints for $HNF1B^{+/+}$, $HNF1B^{+/-}$, and $HNF1B^{-/-}$ at days 6 and day 13 between regions of gained-open and gained-closed chromatin accessibility.

Motif analyses was performed for differential footprints (region extended 15bp on each flank, for footprints of very high quality (score > 40). Footprints at HNF1B, HNF1A were present at day 6 in $HNF1B^{+/+}$ but not $HNF1B^{-/-}$ cells. Footprints for HNF1B, HNF1A, HNF4A and HNF4G were present at day 13 in $HNF1B^{+/+}$ but not $HNF1B^{-/-}$ cells. This is consistent with the findings from the HNF1B ChIP-seq, indicating that foot-printing can be sensitively used to infer transcription factor binding sites.

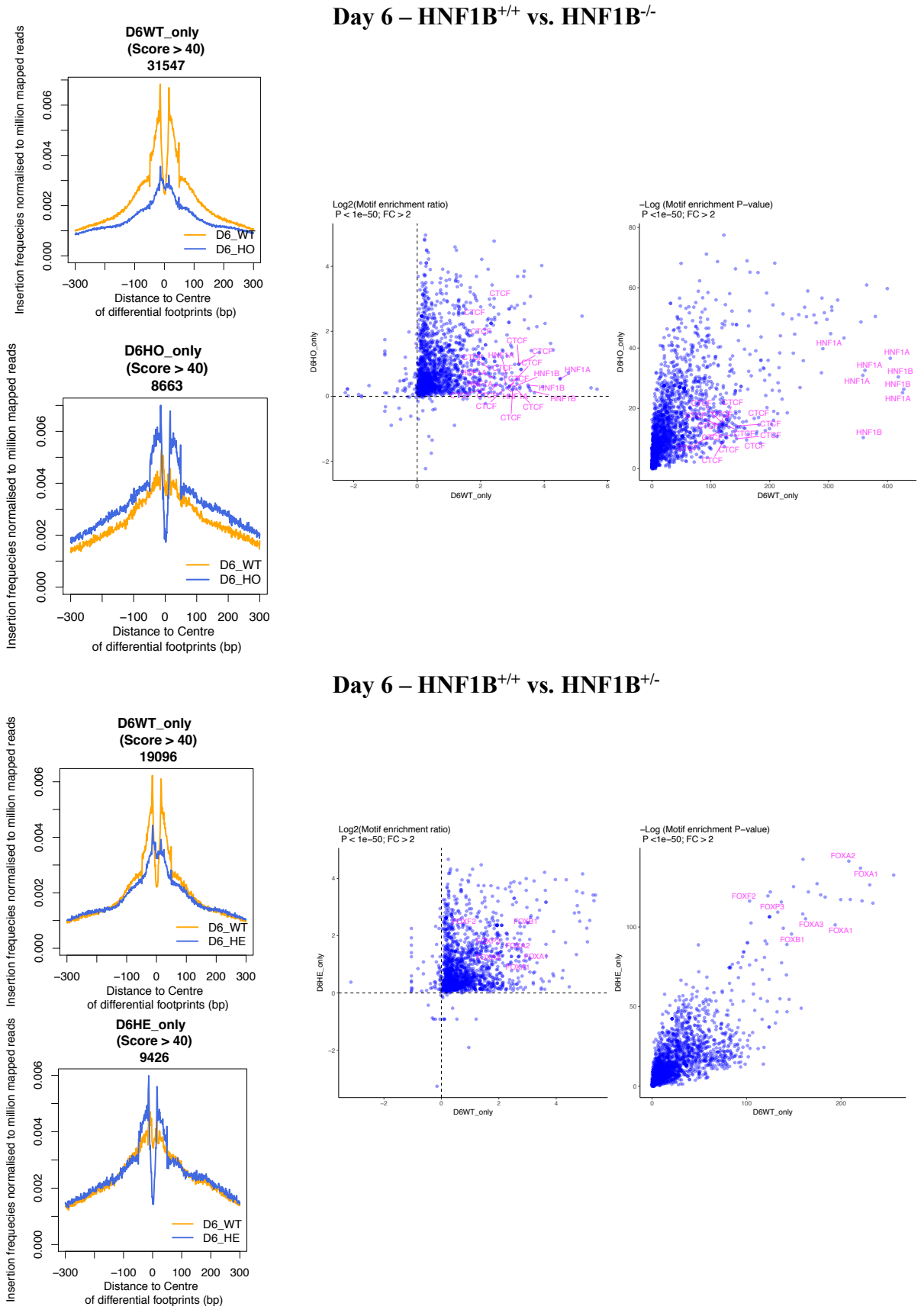


Figure 5.10. Foot-printing and motif analysis at day 6 of differentiation in HNF1B$^{-/-}$ and HNF1B$^{+/-}$ cells compared with HNF1B$^{+/+}$ cells. The number of unique footprints at Day 6 for HNF1B$^{+/+}$ (D6WT) vs. HNF1B$^{-/-}$ (D6HO) and for HNF1B$^{+/+}$ (D6WT) vs. HNF1B$^{+/-}$ (D6HE) is shown as well as enriched Motifs for differential footprints (region extended 15bp on each flank, for footprints of very high quality (score > 40).

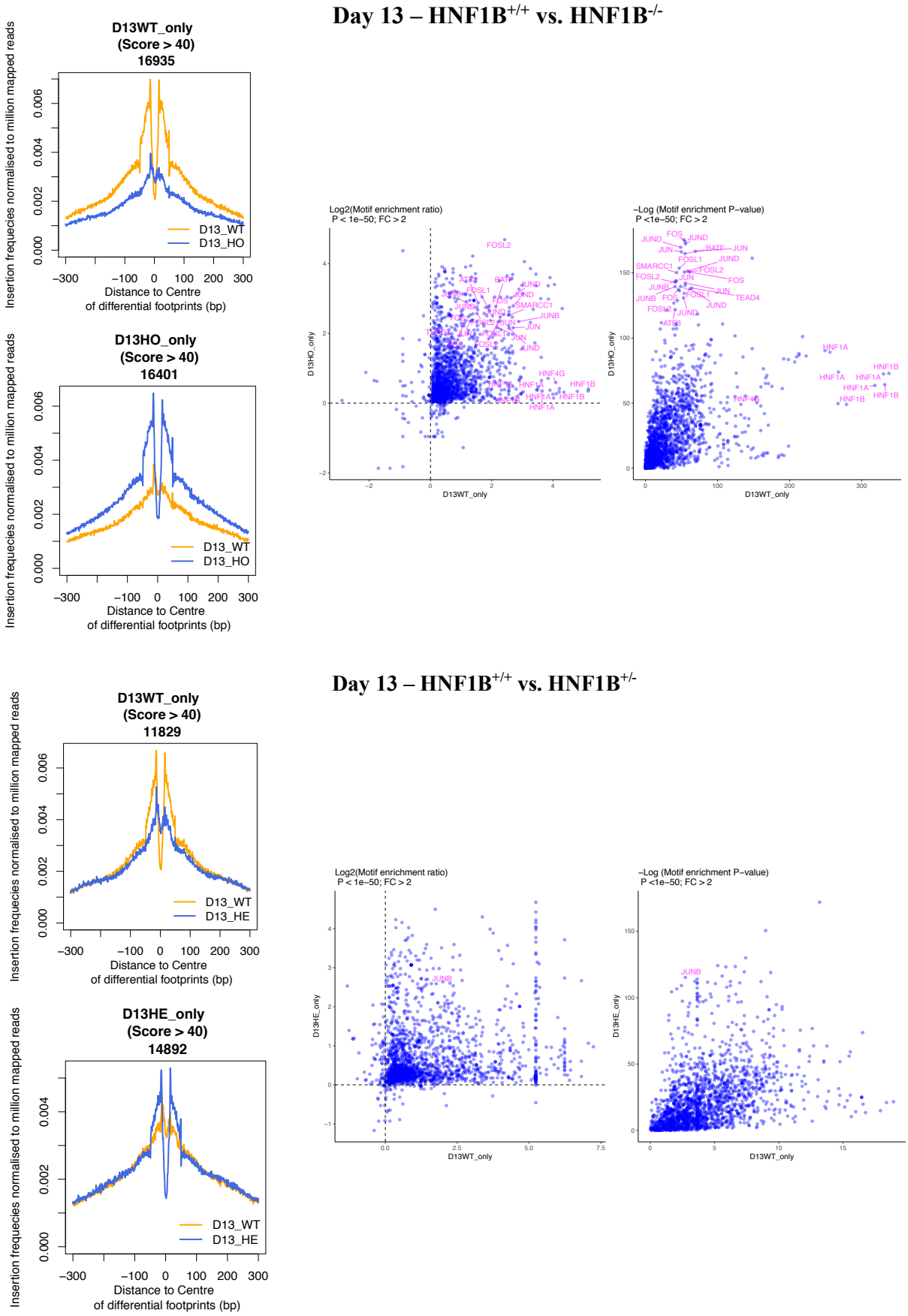


Figure 5.11. Foot-printing and motif analysis at day 13 of differentiation in HNF1B$^{-/-}$ and HNF1B$^{+/-}$ cells compared with HNF1B$^{+/+}$ cells. The number of unique footprints at Day 13 for HNF1B$^{+/+}$ (D13WT) vs. HNF1B$^{-/-}$ (D13HO) and for HNF1B$^{+/+}$ (D13WT) vs. HNF1B$^{+/-}$ (D13HE) is shown as well as enriched Motifs for differential footprints (region extended 15bp on each flank, for footprints of very high quality (score > 40).

5.3 Discussion

In this chapter, RNA-seq and ATAC-seq, were used to further define the molecular mechanisms controlled by HNF1B and the effect HNF1B on modulation of chromatin accessibility during pancreas development.

RNA-sequencing was performed on $HNF1B^{+/+}$, $HNF1B^{+/-}$ and $HNF1B^{-/-}$ cells at the foregut progenitor stage (day 6), when HNF1B starts to be significantly expressed, posterior foregut stage (day 8), pancreatic progenitor stage (day 13), endocrine progenitor stage (day 16) and hormonal cell or β-like cell stage (day 27). Consistent with the cellular phenotyping data outlined in chapter 4, RNA-seq showed that homozygous and heterozygous loss of HNF1B resulted in significant down-regulation of genes involved in foregut, pancreatic progenitor and endocrine cell differentiation. As expected; the effect on gene expression was dose sensitive and HNF1B haploinsufficiency had a milder effect on the pancreatic transcriptome compared with homozygous loss of HNF1B. In addition, genes involved in closely related developmental processes such as hepatobiliary development were also significantly down-regulated.

GO and KEGG pathway analysis showed that several biological processes and pathways were differentially expressed in $HNF1B^{-/-}$ and $HNF1B^{+/-}$ cells compared with $HNF1B^{+/+}$ cells. These included expected GO(BP) and KEGG pathways, such as tube and digestive tract development at day 6, pancreas development, endocrine pancreatic development, pancreatic secretion, and maturity onset diabetes of the young at days 8 and 13 and maturity onset diabetes of the young, endocrine pancreas development, hormone secretion, regulation of secretion, pancreatic secretion and insulin secretion at days 16 and 27. It was interesting to note that although there was significant up-regulation of pathways involved in neuronal cell development and axon guidance in the $HNF1B^{-/-}$ cells, there was no significant and consistent up-regulation of pathways involved in any specific organ developmental pathway in the $HNF1B^{+/-}$ cells.

Other potentially interesting processes and pathways which were significantly differentially expressed in $HNF1B^{-/-}$ and $HNF1B^{+/-}$ cells compared with $HNF1B^{+/+}$ cells included those involved in cell structure, motility and adhesion e.g. axon guidance, cellular metabolic processes, signalling pathways e.g. PPAR signalling, cell proliferation and cell cycle. All of these biological processes and pathways play an important role in organ development and further investigation of the genes involved and their interaction with HNF1B in vivo is required.

One of the main limitations of the RNA-seq experiments is that at each stage of the differentiation protocols the cell population is not entirely homogenous and may contain partially differentiated cells or cells which are differentiating along a different lineage. In particular, at day 16 and day 27 it was more difficult to identify significantly up or down regulated genes or pathways involved in pancreatic endocrine cell development, likely due to the heterogeneous mixture of cells obtained at these stages using our culture conditions. Use of single-cell RNA-sequencing would allow us to study this heterogeneity in more detail.

One of the aims of this project was to perform ChIP-sequencing in HNF1B$^{+/+}$, HNF1B$^{+/-}$ and HNF1B$^{-/-}$ cells at day 6 and day 13 in order to identify genes which are bound and regulated by HNF1B. Unfortunately, however these experiments were not successful during the timeframe of this project, due to the manufacturer discontinuing production of the HNF1B antibody which has been used successfully previously, and inability to find another suitable antibody. HNF1B ChIP-seq data, obtained through our collaboration with the Ferrer group in day 12 pancreatic progenitor cells derived from hESCs, showed that HNF1B binding was most significantly enriched at the HNF1B and HNF1A loci but was also enriched at the loci of several other genes known to be involved in foregut derivative and/or pancreas development. Genes which were also significantly down-regulated in HNF1B$^{-/-}$ cells included HNF1B, HNF1A, HNF4A, ONECUT1, SOX9 and RFX6. HNF1B, SOX9 and RFX6 were also significantly down-regulated in HNF1B$^{+/-}$ cells at day 13.

The ATAC-seq data showed that heterozygous and homozygous loss of HNF1B has a direct effect on modulation of chromatin accessibility during foregut and pancreas development. This suggests a role for HNF1B in maintaining the open chromatin landscape at enhancer regions in foregut and pancreatic progenitor cells. HNF1B, HNF1A, HNF4A, SOX9 and ONECUT1 motifs were enriched in regions of reduced chromatin accessibility, consistent with data from ChIP-seq experiments. The effect on chromatin was dose-sensitive and mild for HNF1B$^{+/-}$ cells compared with HNF1B$^{-/-}$ cells, especially at day 13. This suggests that the impairment of differentiation from foregut to pancreatic progenitor cells observed at day 13 for HNF1B$^{+/-}$ cells is not due to major changes in chromatin configuration.

Footprinting analysis showed that footprints at HNF1B, HNF1A were present at day 6 in HNF1B$^{+/+}$ but not HNF1B$^{-/-}$ cells. Footprints for HNF1B, HNF1A, HNF4A and HNF4G were present at day 13 in HNF1B$^{+/+}$ but not HNF1B$^{-/-}$ cells. This is consistent with the findings from the HNF1B ChIP-seq, indicating that foot-printing can be sensitively used to infer transcription

factor binding sites. The quality scores and significance thresholds used for the ATAC-seq chromatin accessibility and footprints analysis was very stringent and as further experience of this type of analysis is gained in the field, these thresholds will be further optimised.

Chapter 6 Final Discussion

6.1 Summary of main results

A summary of the main achievements and findings from this project include;

1. Generation of isogenic HNF1B wild-type, HNF1B heterozygous knockout and HNF1B homozygous knockout hiPSCs. These cell lines were used in subsequent experiments to study the function of HNF1B in human pancreas development. hiPSCs were also derived from patients with HNF1B-associated disease and could be used in future experiments.
2. A robust protocol which recapitulates human pancreas development was devised based on the previous protocol developed by our group (Cho et al., 2012). The modified protocol showed improved efficiency in the generation of $PDX1^+/NKX6.1^+$ pancreatic progenitor cells, $PDX1^+/NEUROD1^+$ endocrine progenitor cells and glucose responsive insulin producing monohormonal $CPEP^+/PDX1^+$ β-like cells.
3. The normal pattern of HNF1B expression was established in wild-type cells, which showed that HNF1B was up-regulated at the foregut progenitor stage and remained high at the pancreatic progenitor stage and throughout the rest of the differentiation up to the hormonal cell stage.
4. HNF1B homozygous knockout hiPSCs were able to differentiate into definitive endoderm cells with the same efficiency as HNF1B wild-type and HNF1B heterozygous knockout cells; however, there was a complete defect of foregut and pancreatic development *in vitro*. A dose-sensitive effect of HNF1B loss was observed since heterozygous knockout of HNF1B in hiPSCs resulted in impairment, but not complete absence, of pancreatic progenitor and endocrine cell development.
5. There was a significant reduction in the number of C-peptide positive β-like cells in HNF1B heterozygous knockout compared with HNF1B wild-type cells, although the β-like cells that were generated secreted a similar amount of C-peptide in response to glucose per β-like cell.
6. A significant reduction in cell proliferation was seen in HNF1B heterozygous knockout and homozygous knockout cells at the foregut progenitor stage and in heterozygous knockout cells at the pancreatic progenitor stage compared with HNF1B wild-type cells. No significant differences in cell apoptosis was seen between the genotypes.
7. RNA-sequencing identified significantly down- and up-regulated genes and biological pathways in HNF1B homozygous knockout and HNF1B heterozygous knockout cells

compared with HNF1B wild-type cells. As expected genes and pathways involved in foregut, pancreas and hepatobiliary system development and, at the later differentiation stages in pancreatic endocrine cell development were down-regulated in HNF1B homozygous knockout and HNF1B heterozygous knockout cells. Neuronal cell development was significantly up-regulated in HNF1B homozygous knockout cells.

8. ATAC-seq experiments detected major differences in chromatin accessibility in the HNF1B homozygous knockout cells compared with HNF1B wild-type cells at both the foregut and pancreatic progenitor stage of the differentiation process. HNF1B haploinsufficiency had smaller effects on chromatin accessibility especially at the pancreatic progenitor stage. At the foregut progenitor and pancreatic progenitor stages, comparison between HNF1B homozygous knockout and HNF1B heterozygous knockout cells and HNF1B wild-type cells revealed significant motif enrichment for several genes known to play an important role in foregut and pancreatic endoderm development.

These findings and their significance will be discussed in more detail in the next sections of this chapter.

6.2 Generation of HNF1B heterozygous and homozygous knockout hiPSC lines

Human iPSCs in combination with genome editing represent a useful model system to study the roles of particular genes during human development. In this study, CRISPR/Cas9 genetic engineering was used to generate heterozygous and homozygous hiPSC lines, which were otherwise isogenic to the wild-type hiPSC line. These lines were then differentiated along the pancreatic lineage to determine the effect of heterozygous and homozygous knockout of HNF1B on human pancreas development.

The CRISPR/Cas9 system was chosen due to its ease of use and improved efficiency compared with other targeting systems such as TALENS or ZFNs. Multiple strategies are available for genome editing of cell lines using the CRISPR/Cas9 system including use of an expression plasmid, Cas9 mRNA or recombinant Cas9 protein. Use of an expression vector is the most efficient and can contain a reporter gene (e.g. GFP), or a selection marker. However, the Cas9 DNA is degraded more slowly potentially resulting in more off-target effects. Cas9 protein is degraded more quickly, resulting in fewer off-target effects. However, it cannot be combined with a reporter gene or a selection marker (Ran et al., 2013, Kim et al., 2014).

The targeting strategy used in this project, involving use of a Cas9 plasmid, a gRNA plasmid and a plasmid containing the donor repair template and a puromycin selection cassette, was successful in generating homozygous and heterozygous HNF1B knockout iPSCs with high efficiency (up to 40%) in two hiPSC lines with different genetic backgrounds. Insertion of a puromycin resistance cassette allowed for reliable selection of targeted clones and resulted in a large disruption in the open reading frame of the gene in one or both alleles.

Another targeting strategy that could have been used involves cloning of gRNA sequences into a pSpCas9(BB)-2A-Puro vector, resulting in generation of indels in one or both alleles due to NHEJ. This method is more efficient compared with HDR. However, based on initial experience in our laboratory, despite the shorter time required for creating the gRNA and donor template repair plasmids, selection of targeted clones with functional mutations was less efficient. The "strength" of the knock-out phenotype for a given mutant cell is ultimately determined by the amount of residual gene function and was unpredictable. Hence, further validation of loss of the protein by Western Blot is always required.

It is important to note that in HNF1B-associated disease, no genotype-phenotype correlation is seen in patients. Therefore, it was not necessary to create hiPSC lines with specific mutations found in patients, and for the patient-derived hiPSC lines the specific clinical phenotype was not essential as long as they had a pathogenic mutation.

Fibroblasts were derived from 8 patients with HNF1B-related disease and hIPSC lines were successfully generated from 5 of these patients. The genotypes of these patients included whole gene deletions and frameshift, splice site and nonsense mutations. The clinical characteristics of the patients was variable and reflect the phenotypic heterogeneity seen in patients with HNF1B-asociated disease. Only 3 patients had diabetes and unfortunately hiPSC cells could not be successfully derived from these patients following fibroblast reprogramming. Out of the 5 remaining patients, hiPSCs from 2 patients were selected for gene editing based on ease of maintaining the hiPSC cells in culture, limited unwanted background differentiation and ability to efficiently differentiate the hiPSCs to the endoderm stage, when there should be no effect from the HNF1B mutation.

Unfortunately, correction of the mutations in the two patient iPSC lines (R2 and R4) was unsuccessful during the time-course of this project. This was likely due to the limited choice of efficient gRNAs in the region to be targeted. The gRNAs with the highest scores and low numbers of off-target effects in exon regions, from the CRISPR Design Tool, were located more than 100 bp from the patient mutations, resulting in high efficiency for inducing indels but significantly reduced efficiency for correction of the 1bp duplication in patient R2 and nonsense mutation in patient R4. It is estimated that single-base pair correction rates drop approximately fourfold at 100 bp away from the DSB site (Ran et al., 2013). For the 2nd targeting method using recombinant Cas9 protein, no correction of the mutation or introduction of new mutations occurred suggesting either that transfection of the recombinant Cas9 protein or one of the other components was unsuccessful, or rapid degradation of the Cas9 protein occurred prior to DNA cleavage.

Overall these experiments reflect the importance of selecting the correct targeting strategy due to variability in targeting efficiencies depending on cell line, targeted locus, type of repair donor template and location of modification relative to the DSB site.

Genetic engineering of two hiPSC lines with different genetic backgrounds was successful in generating HNF1B heterozygous and homozygous iPSCs with high efficiency. These cell lines

were used in subsequent experiments to study the function of HNF1B in human pancreas development.

6.3 Pancreatic differentiation protocol

A robust protocol which recapitulates human pancreas development was devised based on a previous protocol developed by our group (Cho et al., 2012). The main improvements to the previous protocol included; 1) introduction of Vitamin C at the foregut induction stage of the protocol to improve cell confluency and survival and prevention of premature induction of NEUROG3, 2) addition of the Protein Kinase C activator at the posterior foregut stage of the protocol to induce increased expression of PDX1, 3) use of EGF in combination with NA to increase expression of NKX6-1^{+} in PDX1^{+} pancreatic progenitor cells and to direct cells towards the endocrine/ductal lineage rather than the acinar (PTF1A^{+}) lineage (Nostro et al., 2015), 4) addition of AlkV inhibitor and use of high glucose media (25mM) to induce increased expression of NEUROG3 (Pagliuca et al., 2014, Rezania et al., 2014, Russ et al., 2015), 5) maturation of endocrine progenitors to hormonal cells (mainly β-like cells) by maintaining them in high glucose media with RA and T3 (in B27 supplement) to induce specification of monohormonal CPEP^{+}/PDX1^{+}/NKX6.1^{+} cells which respond to GSIS (Rezania et al., 2014, Russ et al., 2015).

These changes improved the efficiency of the protocol to generate PDX1^{+}/NKX6.1^{+} pancreatic progenitor cells, PDX1^{+}/NEUROD1^{+} endocrine progenitor cells and monohormonal CPEP^{+}/PDX1^{+} β-like cells. A 2-dimensional culture system was used in this study due to ease of culture and ability to phenotype cells efficiently at different stage of the pancreatic differentiation protocol. The hiPSC-derived β-like cells were consistently responsive to GSIS, which is a significant improvement for a 2D culture system. Although we were able to produce monohormonal, functional β-like cells using our protocol, these cells mostly showed low expression of NKX6-1 and MAFA which are important markers for mature β-like cells. The efficiency of the protocol was still cell line dependent; however, this difference between cell lines was less compared to previous protocols used.

Further improvements which could be made to the protocol include use of a 3D culture system to improve the efficiency of generation of endocrine progenitor and more mature β-like cells (Pagliuca et al., 2014, Rezania et al., 2014, Russ et al., 2015). Human organoid systems could

also be used to the test the biophysical and biochemical properties of the niche that sustains pancreatic progenitor cells and β-like cells (Broutier et al., 2016, Greggio et al., 2013). A recent study by Bader et al. (Bader et al., 2016), highlighted the importance of a 3D environment to induce β-cell maturation. Further improvements in GMP culture conditions and understanding of the immunogenicity of hiPSC-derived β-like cells will be required prior to their use for cell therapy to treat diabetes as well as improvement in encapsulation devices to deliver the cells (Ellis et al., 2017, Vegas et al., 2016).

6.4 Effect of HNF1B mutations on pancreatic development

HNF1B plays a crucial role in the organogenesis of several tissues, such as the gut, pancreas, liver, genitourinary tract and kidneys, and in the function of adult tissues. HNF1B is highly conserved in structure and function within vertebrates. However, mice with heterozygous mutations in HNF1B show no phenotype, while humans develop multi-system disease, including diabetes, mediated by haploinsufficiency as the disease mechanism.

Importantly, diseases induced by haploinsufficiency often show phenotypic variability between humans and mice. In addition to HNF1B, heterozygous mutations in several other genes, such as HNF1A, HNF4A and GATA6 which cause monogenic diabetes in humans, have no effect when haploinsufficient in the mouse. This variation in phenotype is likely to be because of species-specific differences in the dosage of HNF1B that is required for full function. HNF1B exists in a tightly regulated feedback circuit in tissues where it is expressed (Arda et al., 2013, Servitja and Ferrer, 2004). Therefore, even subtle differences in the relative activity and temporal expression of the gene may have profound consequences on the overall network activity. Further molecular studies are necessary to fully understand salient differences between pancreas development in mice and humans.

Consistent with previous studies, we found that HNF1B expression is significantly up-regulated during differentiation from definitive endoderm to the foregut progenitor stage, reaching a maximum on days 5 to 6 of our differentiation protocol (Cho et al., 2012, D'Amour et al., 2006, Kroon et al., 2008). HNF1B expression persisted throughout the rest of the differentiation up to the hormonal cell stage (day 27). These findings are consistent with the known expression pattern and important role of HNF1B in foregut specification and pancreas development observed in mice and humans.

Homozygous knockout of HNF1B in hiPSCs resulted in a complete defect of foregut and pancreatic development *in vitro*. HNF1B is essential for the efficient formation of posterior foregut, therefore, homozygous loss of HNF1B protein expression in the human embryo is likely to be lethal due to a primary defect in gut tube / foregut formation. A dose-sensitive effect of HNF1B loss was observed since heterozygous knockout of HNF1B in humans resulted in a milder phenotype including impairment of foregut, pancreatic progenitor and endocrine cell development, although some functionality of the produced β-like cells was maintained. This suggests that full expression of HNF1B may only be required at an early stage of differentiation and that its function may be dispensable during endocrine lineage specification.

Of note, cell proliferation was reduced in HNF1B heterozygous knockout cells between the foregut progenitor and pancreatic progenitor stage of differentiation and in homozygous knockout cells at foregut progenitor stage but not at later stages. During embryogenesis, development of pancreatic tissue requires the proliferation of progenitor cells, followed by their differentiation into endocrine or acinar cells. A reduced number of pancreatic progenitor cells during embryonic development would predict a smaller pancreas (Stanger et al., 2007), a phenotype observed in some HNF1B heterozygous patients even in the absence of overt diabetes or pancreatic exocrine insufficiency. About 80% of the islet cell mass present at birth is generated by the proliferation and differentiation of endocrine progenitors, with the other 20% coming from islet cell proliferation (Bouwens and Rooman, 2005). After birth, the predominant mechanism for islet cell mass maintenance is self-duplication.

The findings from this study are consistent with HNF1B-associated diabetes being due to defects in cell expansion and maintenance of pancreatic progenitor cells during development, which leads to reduced beta cell mass at birth and diabetes later in life. The age at onset of diabetes will depend on number and functionality of beta cells at birth and how these are affected by interaction with environmental factors throughout life.

A recent study by DeVas et al. (De Vas et al., 2015), used constitutive and inducible conditional inactivation of Hnf1b to study its effect on pancreas development in mice. The study showed that Hnf1b is required for the proliferation and survival of multipotent progenitor cells through modulation of FGF and Notch pathways. In vivo ChIP on E12.5 pancreata showed that deletion of Hnf1b resulted in down-regulation of Fgfr4 and qRT-PCR of Notch pathway genes showed up-regulation of Hey factors, which are inhibitors of the Ptf1a transcriptional complex, and

down-regulation of the Notch ligand Dll1, which is activated by Ptf1a. Hnf1b was also found to be essential for the generation of endocrine precursors, likely by directly regulating Neurog3 expression as well as by acting directly on Glis3 and Onecut1 in the control of Neurog3.

Consistent with the DeVas study, our RNA-seq data also showed significant down-regulation of FGFR4 at day 13 (pancreatic progenitor stage) in HNF1B homozygous knockout cells and HNF1B heterozygous knockout cells compared with HNF1B wild-type cells. FGFR3 and FGFR2 were also significantly down-regulated in HNF1B homozygous knockout cells. PTF1A, HES1 and DLL1 were also significantly down-regulated in HNF1B homozygous knockout cells compared with HNF1B wild-type cells but only PTF1A was significantly down-regulated in HNF1B heterozygous knockout cells. There was no up-regulation of HEY factors at day 13, but HEY1 and HEY2 were significantly up-regulated at the endocrine progenitor stage (day 16). At the endocrine progenitor stage there was significant down-regulation of NEUROG3 and ONECUT1 in HNF1B homozygous knockout cells compared with HNF1B wild-type cells; however GLIS3 was not differentially expressed.

The effect of HNF1B on exocrine and ductal cell development was not explored in our study and there are only very few in vitro protocols which have been developed to differentiate pancreatic progenitor cells to exocrine or ductal cells (Tulpule et al., 2013). 3-dimensional in vitro pancreatic organoid or in vivo systems however could be used to study the role of specific genes on all 3 pancreatic lineages (Broutier et al., 2016, Bruin et al., 2013, Greggio et al., 2013, Kroon et al., 2008).

Epidermal growth factor (EGF) stimulates proliferation of pancreatic progenitor cells resulting in increased $PDX1^{+}/NKX6.1^{+}$ cells. A study by Verdeguer at al. (Verdeguer et al., 2010) showed that Hnf1b only caused polycystic kidney disease in mice if it was inactivated in rapidly proliferating tubular cells due to its role as a bookmarking factor. In order to investigate if a similar effect is seen in human pancreas differentiation, we increased the concentration of EGF to induce cells to proliferate more rapidly to see if this exacerbates the phenotype. Higher (3x) concentrations of EGF resulted in an increase in the number of proliferating or EdU positive cells in both HNF1B wild-type and HNF1B heterozygous knockout cell lines. At each concentration, the percentage of proliferating cells was greater in HNF1B wild-type compared with HNF1B heterozygous knockout cell lines; however, there was no exacerbation of the phenotype in HNF1B heterozygous knockout or HNF1B homozygous knockout cells. This suggests that HNF1B may not act as a bookmarking factor in human pancreatic cells; however

further studies regarding its role as a bookmarking factor in the pancreas as in the kidney are required.

There has been one previous study which used human iPSC lines from patients with HNF1B-associated diabetes due to an S148L/+ mutation to study the effect of this mutation on endoderm and pancreas development (Teo et al., 2016). The study showed that a decrease in HNF1B expression did not affect endoderm differentiation into foregut or the subsequent induction of pancreatic progenitors and HNF1B haploinsufficiency appeared to increase the expression of endoderm makers such as SOX17 and pancreatic bud markers such as PDX1. However, the study revealed a decrease in PAX6 expression during the differentiation of HNF1B mutant iPSCs into endocrine cells. The results from our study were not consistent with these findings and in contrast we found that HNF1B haploinsufficiency resulted in significant impairment of foregut and pancreas development. A significant decrease in PAX6 expression was also seen in our study in $HNF1B^{+/-}$ an $HNF1B^{-/-}$ cells compared with $HNF1B^{+/+}$ cells in day 27 β-like cells. RNA-seq, ChIP-seq and ATAC-seq experiments did not however suggest and direct interaction between HNF1B and PAX6.

The main limitations of the study by Teo et al. were that their protocol did not allow the production of glucose responsive β-like cells and isogenic cell lines were not used, and therefore genetic background could account for any differences in gene expression seen.

While simple constitutive knockouts are useful and informative, it would have also been desirable to engineer inducible or conditional loss-of-function models e.g. using the optimized inducible gene knockdown or knockout (sOPTiKD or sOPTiKO) platform (Bertero et al., 2016). This would allow us to investigate the dose-related effect of loss of HNF1B by inducing different percentages of knockdown of HNF1B and also to directly study the role of HNF1B at particular stages of the pancreas development in more detail.

Our current in vitro protocol does not allow the generation of functionally mature adult-like islet cells or other pancreatic cell lineages. In vivo experiments were therefore performed to study the function of HNF1B in acinar, exocrine and mature adult-like islet cells. We were able to show that HNF1B wild-type and HNF1B heterozygous knockout pancreatic progenitor cells can differentiate to all three pancreatic lineages: ductal, endocrine and exocrine cells following transplantation under the kidney capsule in mice, while transplantation of HNF1B homozygous knockout cells in vivo did not result in the development of any pancreatic structures. However,

due to multiple technical challenges (described in chapter 4) we were unable to determine if there are any significant differences in the development of endocrine, exocrine and ductal pancreatic cells and insulin secretion in vivo between HNF1B wild-type and HNF1B heterozygous knockout cells. Further optimisation of the in vivo experiments is required.

6.5 Exploring the transcriptional network controlled by HNF1B

We performed an in-depth comparison of the global transcriptional profile of HNF1B wild-type, HNF1B heterozygous knockout and HNF1B homozygous knockout hiPSC lines at each stage of the differentiation process using RNA-seq. In addition, ATAC-seq was used to determine chromatin accessibility in HNF1B wild-type, compared with HNF1B heterozygous knockout and HNF1B homozygous knockout cells at the foregut progenitor (day 6) and pancreatic progenitor (day 13) of the pancreatic differentiation process. Data from the RNA-seq and ATAC-seq experiments was also integrated with HNF1B ChIP-seq data from pancreatic progenitor cells derived from hESCs, obtained through our collaboration with the Ferrer group (Cebola et al., 2015).

RNA-seq data showed significant and dose-dependent effects of loss of HNF1B on the transcriptional network involved in foregut, pancreatic progenitor and pancreatic endocrine cell differentiation. These results were consistent with our qRT-PCR results looking at the expression of specific markers at each stage of the pancreatic differentiation protocol. At day 16 and day 27 it was more difficult to identify significant up or down regulated genes or pathways in HNF1B heterozygous knockout cells compared with HNF1B wild-type cells, likely due to the heterogeneous mixture of cells obtained at these stages using our culture conditions. Use of single-cell RNA-seq would allow us to study this heterogeneity in more detail and to classify subpopulations within known cell types (Muraro et al., 2016).

GO and KEGG pathway analysis showed that several biological processes and pathways were differentially expressed in HNF1B heterozygous knockout and HNF1B homozygous knockout cells compared with HNF1B wild-type cells. Potentially interesting processes and pathways which were significantly differentially expressed in HNF1B homozygous knockout and HNF1B heterozygous knockout cells compared with HNF1B wild-type cells included those involved in cell structure, motility and adhesion e.g. axon guidance, cellular metabolic processes, signalling pathways e.g. PPAR signalling, cell proliferation and cell cycle. Further

exploration of the genes involved in these biological processes and pathways and their interaction with HNF1B in vivo is required.

KEGG pathway analysis showed that genes associated with the Peroxisome proliferator-activated receptor (PPAR) signalling pathway were down-regulated at all stages of the differentiation process, particularly in HNF1B homozygous knockout cells. Peroxisome proliferator-activated receptors (PPARs) compose a family of three nuclear receptors (PPARalpha, PPARbeta/delta, PPARgamma) which act as lipid sensors to modulate gene expression. PPAR pathways are known to play an important role in lipid metabolism and cell proliferation (Feige et al., 2006). In the pancreas, PPAR-γ has been shown to directly regulate key β cell genes involved in glucose sensing, insulin secretion and insulin gene transcription (Gupta et al., 2010). This finding is of particular interest since HNF1B gene mutations are known to have effects on lipid metabolism since patients may develop hepatic insulin resistance and dyslipidaemia (Brackenridge et al., 2006, Pearson et al., 2004). Hepatic insulin resistance without any impairment of peripheral glucose uptake has not previously been described in any other type of diabetes. Studying patients with HNF1B mutations may therefore provide insights into the mechanism of hepatic insulin resistance in complex polygenic T2D.

HNF1B is known to play an important role in hepatobiliary development and liver function. Using the HNF1B mutant hiPSCs generated in this study, it would be interesting to differentiate these lines along the hepatic lineage in order to determine the effects of loss of HNF1B on metabolic pathways such as lipogenesis and gluconeogenesis. Previous experiments performed by our group and other studies however have shown that while genes involved in gluconeogenesis were up-regulated in hiPSC-derived hepatocytes following insulin stimulation, an effect on lipogeneis could not be reliably determined (Ding et al., 2013). This is likely due to immature nature of hiPSC derived hepatocytes and sub-optimal culture conditions for studying metabolic pathways in these cells.

Major differences in chromatin accessibility were detected in the HNF1B homozygous knockout cells compared with HNF1B wild-type cells at both the foregut progenitor (day 6) and pancreatic progenitor (day 13) stages of the differentiation process. A preference for down-regulation of chromatin accessibility was observed with about double the number of regions compacting. However, in HNF1B heterozygous knockout cells the effect on chromatin accessibility was smaller, with more changes at day 6 compared with day 13.

Motif enrichment analysis of regions of differential chromatin accessibility combined with RNA-seq expression data at day 6 and day 13 revealed significant motif enrichment and down-regulation of several genes known to play an important role in foregut and pancreatic endoderm development such as HNF1B, HNF1A, HNF4G, HNF4A and ONECUT1, in HNF1B homozygous knockout and HNF1B heterozygous knockout cells compared with HNF1B wild-type cells.

The ATAC-seq data showed that homozygous loss of HNF1B had major effects on chromatin accessibility at day 6 and day 13, suggesting that the overall effect of HNF1B is to bind and open chromatin to allow recruitment of other transcription factors, which is well known function for genes with a HNF/FOXA domain (Kaestner, 2010). However, HNF1B haploinsufficiency has a greater impact on chromatin organisation at an earlier stage of development and the effects on pancreatic progenitor differentiation may be independent of chromatin change. The relationship between the effect of HNF1B on chromatin accessibility and the expression of downstream target genes requires further investigation.

6.6 Concluding remarks

The results from this study provide further insights into the molecular mechanisms by which HNF1B regulates human pancreas development and function, revealing that HNF1B haploinsufficiency impairs the expansion and maintenance of pancreatic progenitor cells in vitro. In vivo, this would likely result in reduced beta cell numbers at birth and diabetes later in life in patients with HNF1B-associated disease. These mechanisms suggest that the capacity to produce pancreatic progenitor cells during embryonic life could determine individual susceptibility to diabetes.

Large-scale production of cells from human iPSC lines have enabled quantitative genome-wide analyses, such as ChIP-Seq, ATAC-seq and RNA-Seq, and thus can help reveal the transcriptional and epigenetic networks involving HNF1B in humans. In this project, such analyses have been used to explore the consequences of loss of one or both alleles of HNF1B on the transcriptional networks directing pancreatic development and β-cell function. We have revealed potential biological processes and pathways involving HNF1B, which could be of interest for drug development and improvement of protocols to generate β-cells for cell replacement therapy in diabetes. Single-cell RNA-sequencing can also be used to explore the

heterogeneity of the cell populations obtained during the pancreatic differentiation process and can be used to interrogate in more detail the mechanisms by which HNF1B haploinsufficiency affects pancreatic development.

Thus far, hPSCs have been mainly used to study the effect of a single genetic locus on disease phenotypes by comparing mutant to control wild-type hPSCs. hPSCs can also be used interrogate complex genetic interactions, which has broad implications for investigation of multigenic human traits. Diseases caused by heterozygous mutations are likely to be susceptible to even subtle decreases in the expression or activity of the remaining allele that could be caused by the environment, background genetic modifiers, or other stochastic events. These factors can account for the wide range of clinical phenotypes observed in individuals carrying HNF1B heterozygous mutations and the age at which patients develop diabetes. Therefore, potential treatment options for HNF1B-associated diabetes may involve augmenting the expression or activity of HNF1B or its downstream targets, a strategy that may be extended to type 2 diabetic patients with SNPs in the HNF1B gene which may result in dysregulated gene expression or activity.

Chapter 7 References

1. Ahlgren, U., Pfaff, S. L., Jessell, T. M., Edlund, T. & Edlund, H. 1997. Independent requirement for ISL1 in formation of pancreatic mesenchyme and islet cells. *Nature,* 385**,** 257-60.

2. Ahmad, Z., Rafeeq, M., Collombat, P. & Mansouri, A. 2015. Pax6 Inactivation in the Adult Pancreas Reveals Ghrelin as Endocrine Cell Maturation Marker. *PLoS One,* 10**,** e0144597.

3. Al-Khawaga, S., Memon, B., Butler, A. E., Taheri, S., Abou-Samra, A. B. & Abdelalim, E. M. 2017. Pathways governing development of stem cell-derived pancreatic beta cells: lessons from embryogenesis. *Biol Rev Camb Philos Soc*.

4. Andersson, R. 2015. Promoter or enhancer, what's the difference? Deconstruction of established distinctions and presentation of a unifying model. *Bioessays,* 37**,** 314-23.

5. Apelqvist, A., Li, H., Sommer, L., Beatus, P., Anderson, D. J., Honjo, T., Hrabe De Angelis, M., Lendahl, U. & Edlund, H. 1999. Notch signalling controls pancreatic cell differentiation. *Nature,* 400**,** 877-81.

6. Arda, H. E., Benitez, C. M. & Kim, S. K. 2013. Gene regulatory networks governing pancreas development. *Dev Cell,* 25**,** 5-13.

7. Bader, E., Migliorini, A., Gegg, M., Moruzzi, N., Gerdes, J., Roscioni, S. S., Bakhti, M., Brandl, E., Irmler, M., Beckers, J., Aichler, M., Feuchtinger, A., Leitzinger, C., Zischka, H., Wang-Sattler, R., Jastroch, M., Tschop, M., Machicao, F., Staiger, H., Haring, H. U., Chmelova, H., Chouinard, J. A., Oskolkov, N., Korsgren, O., Speier, S. & Lickert, H. 2016. Identification of proliferative and mature beta-cells in the islets of Langerhans. *Nature,* 535**,** 430-4.

8. Bamgbola, O. 2016. Metabolic consequences of modern immunosuppressive agents in solid organ transplantation. *Ther Adv Endocrinol Metab,* 7**,** 110-27.

9. Bannister, A. J. & Kouzarides, T. 2011. Regulation of chromatin by histone modifications. *Cell Res,* 21**,** 381-95.

10. Barbacci, E., Chalkiadaki, A., Masdeu, C., Haumaitre, C., Lokmane, L., Loirat, C., Cloarec, S., Talianidis, I., Bellanne-Chantelot, C. & Cereghini, S. 2004. HNF1beta/TCF2 mutations impair transactivation potential through altered co-regulator recruitment. *Hum Mol Genet,* 13**,** 3139-49.

11. Barbacci, E., Reber, M., Ott, M. O., Breillat, C., Huetz, F. & Cereghini, S. 1999. Variant hepatocyte nuclear factor 1 is required for visceral endoderm specification. *Development,* 126**,** 4795-805.

12. Beards, F., Frayling, T., Bulman, M., Horikawa, Y., Allen, L., Appleton, M., Bell, G. I., Ellard, S. & Hattersley, A. T. 1998. Mutations in hepatocyte nuclear factor 1beta are not a common cause of maturity-onset diabetes of the young in the U.K. *Diabetes,* 47**,** 1152-4.

13. Beckers, D., Bellanne-Chantelot, C. & Maes, M. 2007. Neonatal cholestatic jaundice as the first symptom of a mutation in the hepatocyte nuclear factor-1beta gene (HNF-1beta). *J Pediatr,* 150**,** 313-4.

14. Bellanne-Chantelot, C., Clauin, S., Chauveau, D., Collin, P., Daumont, M., Douillard, C., Dubois-Laforgue, D., Dusselier, L., Gautier, J. F., Jadoul, M., Laloi-Michelin, M., Jacquesson, L., Larger, E., Louis, J., Nicolino, M., Subra, J. F., Wilhem, J. M., Young, J., Velho, G. & Timsit, J. 2005. Large genomic rearrangements in the hepatocyte nuclear factor-1beta (TCF2) gene are the most frequent cause of maturity-onset diabetes of the young type 5. *Diabetes,* 54**,** 3126-32.

15. Bernstein, B. E., Stamatoyannopoulos, J. A., Costello, J. F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M. A., Beaudet, A. L., Ecker, J. R., Farnham, P. J., Hirst, M., Lander, E. S., Mikkelsen, T. S. & Thomson, J. A. 2010. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol,* 28**,** 1045-8.

16. 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.

17. Bhushan, A., Itoh, N., Kato, S., Thiery, J. P., Czernichow, P., Bellusci, S. & Scharfmann, R. 2001. Fgf10 is essential for maintaining the proliferative capacity of epithelial progenitor cells during early pancreatic organogenesis. *Development,* 128**,** 5109-17.

18. Bingham, C. & Hattersley, A. T. 2004. Renal cysts and diabetes syndrome resulting from mutations in hepatocyte nuclear factor-1beta. *Nephrol Dial Transplant,* 19**,** 2703-8.

19. Bockenhauer, D. & Jaureguiberry, G. 2016. HNF1B-associated clinical phenotypes: the kidney and beyond. *Pediatr Nephrol,* 31**,** 707-14.

20. Body-Bechou, D., Loget, P., D'herve, D., Le Fiblec, B., Grebille, A. G., Le Guern, H., Labarthe, C., Redpath, M., Cabaret-Dufour, A. S., Sylvie, O., Fievet, A., Antignac, C., Heidet, L., Taque, S. & Patrice, P. 2014. TCF2/HNF-1beta mutations: 3 cases of fetal severe pancreatic agenesis or hypoplasia and multicystic renal dysplasia. *Prenat Diagn,* 34**,** 90-3.

21. Boland, M. J., Nazor, K. L. & Loring, J. F. 2014. Epigenetic regulation of pluripotency and differentiation. *Circ Res,* 115**,** 311-24.

22. Bouwens, L. & Rooman, I. 2005. Regulation of pancreatic beta-cell mass. *Physiol Rev,* 85**,** 1255-70.

23. Boyadjiev, S. A. & Jabs, E. W. 2000. Online Mendelian Inheritance in Man (OMIM) as a knowledgebase for human developmental disorders. *Clin Genet,* 57**,** 253-66.

24. Boyle, A. P., Davis, S., Shulha, H. P., Meltzer, P., Margulies, E. H., Weng, Z., Furey, T. S. & Crawford, G. E. 2008. High-resolution mapping and characterization of open chromatin across the genome. *Cell,* 132**,** 311-22.

25. Brackenridge, A., Pearson, E. R., Shojaee-Moradie, F., Hattersley, A. T., Russell-Jones, D. & Umpleby, A. M. 2006. Contrasting insulin sensitivity of endogenous glucose production rate in subjects with hepatocyte nuclear factor-1beta and -1alpha mutations. *Diabetes,* 55**,** 405-11.

26. Brissova, M., Blaha, M., Spear, C., Nicholson, W., Radhika, A., Shiota, M., Charron, M. J., Wright, C. V. & Powers, A. C. 2005. Reduced PDX-1 expression impairs islet response to insulin resistance and worsens glucose homeostasis. *Am J Physiol Endocrinol Metab,* 288**,** E707-14.

27. Brissova, M., Shiota, M., Nicholson, W. E., Gannon, M., Knobel, S. M., Piston, D. W., Wright, C. V. & Powers, A. C. 2002. Reduction in pancreatic transcription factor PDX-1 impairs glucose-stimulated insulin secretion. *J Biol Chem,* 277**,** 11225-32.

28. Broutier, L., Andersson-Rolf, A., Hindley, C. J., Boj, S. F., Clevers, H., Koo, B. K. & Huch, M. 2016. Culture and establishment of self-renewing human and mouse adult liver and pancreas 3D organoids and their genetic manipulation. *Nat Protoc,* 11**,** 1724-43.

29. Bruin, J. E., Rezania, A., Xu, J., Narayan, K., Fox, J. K., O'neil, J. J. & Kieffer, T. J. 2013. Maturation and function of human embryonic stem cell-derived pancreatic progenitors in macroencapsulation devices following transplant into mice. *Diabetologia,* 56**,** 1987-98.

30. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods,* 10**,** 1213-8.

31. Butler, A. E., Janson, J., Bonner-Weir, S., Ritzel, R., Rizza, R. A. & Butler, P. C. 2003. Beta-cell deficit and increased beta-cell apoptosis in humans with type 2 diabetes. *Diabetes,* 52**,** 102-10.

32. 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.

33. 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.

34. Cha, J. Y., Kim, H., Kim, K. S., Hur, M. W. & Ahn, Y. 2000. Identification of transacting factors responsible for the tissue-specific expression of human glucose transporter type 2 isoform gene. Cooperative role of hepatocyte nuclear factors 1alpha and 3beta. *J Biol Chem,* 275**,** 18358-65.

35. Chandra, V., Albagli-Curiel, O., Hastoy, B., Piccand, J., Randriamampita, C., Vaillant, E., Cave, H., Busiah, K., Froguel, P., Vaxillaire, M., Rorsman, P., Polak, M. & Scharfmann, R. 2014. RFX6 regulates insulin secretion by modulating Ca2+ homeostasis in human beta cells. *Cell Rep,* 9**,** 2206-18.

36. Chawla, K., Tripathi, S., Thommesen, L., Laegreid, A. & Kuiper, M. 2013. TFcheckpoint: a curated compendium of specific DNA-binding RNA polymerase II transcription factors. *Bioinformatics,* 29**,** 2519-20.

37. Chen, S., Borowiak, M., Fox, J. L., Maehr, R., Osafune, K., Davidow, L., Lam, K., Peng, L. F., Schreiber, S. L., Rubin, L. L. & Melton, D. 2009. A small molecule that directs differentiation of human ESCs into the pancreatic lineage. *Nat Chem Biol,* 5**,** 258-65.

38. Chen, Y., Pan, F. C., Brandes, N., Afelik, S., Solter, M. & Pieler, T. 2004. Retinoic acid signaling is essential for pancreas development and promotes endocrine at the expense of exocrine cell differentiation in Xenopus. *Dev Biol,* 271**,** 144-60.

39. Chen, Y. Z., Gao, Q., Zhao, X. Z., Chen, Y. Z., Bennett, C. L., Xiong, X. S., Mei, C. L., Shi, Y. Q. & Chen, X. M. 2010. Systematic review of TCF2 anomalies in renal cysts and

diabetes syndrome/maturity onset diabetes of the young type 5. *Chin Med J (Engl),* 123**,** 3326-33.

40. Chi, K. R. 2016. Reveling in the Revealed: A growing toolbox for surveying the activity of entire genomes. *The Scientist.*

41. 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.

42. Clissold, R. L., Hamilton, A. J., Hattersley, A. T., Ellard, S. & Bingham, C. 2014. HNF1B-associated renal and extra-renal disease-an expanding clinical spectrum. *Nat Rev Nephrol*.

43. Coffinier, C., Barra, J., Babinet, C. & Yaniv, M. 1999a. Expression of the vHNF1/HNF1beta homeoprotein gene during mouse organogenesis. *Mech Dev,* 89**,** 211-3.

44. Coffinier, C., Gresh, L., Fiette, L., Tronche, F., Schutz, G., Babinet, C., Pontoglio, M., Yaniv, M. & Barra, J. 2002. Bile system morphogenesis defects and liver dysfunction upon targeted deletion of HNF1beta. *Development,* 129**,** 1829-38.

45. Coffinier, C., Thepot, D., Babinet, C., Yaniv, M. & Barra, J. 1999b. Essential role for the homeoprotein vHNF1/HNF1beta in visceral endoderm differentiation. *Development,* 126**,** 4785-94.

46. Conrad, E., Stein, R. & Hunter, C. S. 2014. Revealing transcription factors during human pancreatic beta cell development. *Trends Endocrinol Metab,* 25**,** 407-14.

47. The Encode Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature,* 489**,** 57-74.

48. 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.

49. Danaei, G., Finucane, M. M., Lu, Y., Singh, G. M., Cowan, M. J., Paciorek, C. J., Lin, J. K., Farzadfar, F., Khang, Y. H., Stevens, G. A., Rao, M., Ali, M. K., Riley, L. M., Robinson, C. A. & Ezzati, M. 2011. National, regional, and global trends in fasting plasma glucose and diabetes prevalence since 1980: systematic analysis of health examination surveys and epidemiological studies with 370 country-years and 2.7 million participants. *Lancet,* 378**,** 31-40.

50. 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.

51. De Vas, M. G., Kopp, J. L., Heliot, C., Sander, M., Cereghini, S. & Haumaitre, C. 2015. Hnf1b controls pancreas morphogenesis and the generation of Ngn3+ endocrine progenitors. *Development,* 142**,** 871-82.

52. Dengler, V. L., Galbraith, M. & Espinosa, J. M. 2014. Transcriptional regulation by hypoxia inducible factors. *Crit Rev Biochem Mol Biol,* 49**,** 1-15.

53. Ding, Q., Lee, Y. K., Schaefer, E. A., Peters, D. T., Veres, A., Kim, K., Kuperwasser, N., Motola, D. L., Meissner, T. B., Hendriks, W. T., Trevisan, M., Gupta, R. M., Moisan, A., Banks, E., Friesen, M., Schinzel, R. T., Xia, F., Tang, A., Xia, Y., Figueroa, E., Wann, A., Ahfeldt, T., Daheron, L., Zhang, F., Rubin, L. L., Peng, L. F., Chung, R. T., Musunuru, K. & Cowan, C. A. 2013. A TALEN genome-editing system for generating human stem cell-based disease models. *Cell Stem Cell,* 12**,** 238-51.

54. Dong, P. D., Munson, C. A., Norton, W., Crosnier, C., Pan, X., Gong, Z., Neumann, C. J. & Stainier, D. Y. 2007. Fgf10 regulates hepatopancreatic ductal system patterning and differentiation. *Nat Genet,* 39**,** 397-402.

55. Dorrell, C., Schug, J., Canaday, P. S., Russ, H. A., Tarlow, B. D., Grompe, M. T., Horton, T., Hebrok, M., Streeter, P. R., Kaestner, K. H. & Grompe, M. 2016. Human islets contain four distinct subtypes of beta cells. *Nat Commun,* 7**,** 11756.

56. Drewes, T., Senkel, S., Holewa, B. & Ryffel, G. U. 1996. Human hepatocyte nuclear factor 4 isoforms are encoded by distinct and differentially expressed genes. *Mol Cell Biol,* 16**,** 925-31.

57. Edghill, E. L., Bingham, C., Ellard, S. & Hattersley, A. T. 2006a. Mutations in hepatocyte nuclear factor-1beta and their related phenotypes. *J Med Genet,* 43**,** 84-90.

58. Edghill, E. L., Bingham, C., Slingerland, A. S., Minton, J. A., Noordam, C., Ellard, S. & Hattersley, A. T. 2006b. Hepatocyte nuclear factor-1 beta mutations cause neonatal diabetes and intrauterine growth retardation: support for a critical role of HNF-1beta in human pancreatic development. *Diabet Med,* 23**,** 1301-6.

59. El-Khairi, R. & Vallier, L. 2016. The role of hepatocyte nuclear factor 1beta in disease and development. *Diabetes Obes Metab,* 18 Suppl 1**,** 23-32.

60. Ellard, S., Lango Allen, H., De Franco, E., Flanagan, S. E., Hysenaj, G., Colclough, K., Houghton, J. A., Shepherd, M., Hattersley, A. T., Weedon, M. N. & Caswell, R. 2013. Improved genetic testing for monogenic diabetes using targeted next-generation sequencing. *Diabetologia,* 56**,** 1958-63.

61. Ellis, C., Ramzy, A. & Kieffer, T. J. 2017. Regenerative medicine and cell-based approaches to restore pancreatic function. *Nat Rev Gastroenterol Hepatol,* 14**,** 612-628.

62. Elsea, S. H. & Lucas, R. E. 2002. The mousetrap: what we can learn when the mouse model does not mimic the human disease. *ILAR J,* 43**,** 66-79.

63. Faguer, S., Esposito, L., Casemayou, A., Pirson, Y., Decramer, S., Cartery, C., Hazzan, M., Garrigue, V., Roussey, G., Cointault, O., Ho, T., Merville, P., Devuyst, O., Gourdy, P., Chassaing, N., Bascands, J. L., Kamar, N., Schanstra, J. P., Rostaing, L. & Chauveau, D. 2016. Calcineurin Inhibitors Downregulate HNF-1beta and May Affect the Outcome of HNF1B Patients After Renal Transplantation. *Transplantation,* 100**,** 1970-8.

64. Feige, J. N., Gelman, L., Michalik, L., Desvergne, B. & Wahli, W. 2006. From molecular action to physiological outputs: peroxisome proliferator-activated receptors are nuclear receptors at the crossroads of key cellular functions. *Prog Lipid Res,* 45**,** 120-59.

65. Ferre, S., Bongers, E. M., Sonneveld, R., Cornelissen, E. A., Van Der Vlag, J., Van Boekel, G. A., Wetzels, J. F., Hoenderop, J. G., Bindels, R. J. & Nijenhuis, T. 2013. Early development

of hyperparathyroidism due to loss of PTH transcriptional repression in patients with HNF1beta mutations? *J Clin Endocrinol Metab,* 98**,** 4089-96.

66. Flanagan, S. E., De Franco, E., Lango Allen, H., Zerah, M., Abdul-Rasoul, M. M., Edge, J. A., Stewart, H., Alamiri, E., Hussain, K., Wallis, S., De Vries, L., Rubio-Cabezas, O., Houghton, J. A., Edghill, E. L., Patch, A. M., Ellard, S. & Hattersley, A. T. 2014a. Analysis of transcription factors key for mouse pancreatic development establishes NKX2-2 and MNX1 mutations as causes of neonatal diabetes in man. *Cell Metab,* 19**,** 146-54.

67. Flanagan, S. E., Haapaniemi, E., Russell, M. A., Caswell, R., Allen, H. L., De Franco, E., Mcdonald, T. J., Rajala, H., Ramelius, A., Barton, J., Heiskanen, K., Heiskanen-Kosma, T., Kajosaari, M., Murphy, N. P., Milenkovic, T., Seppanen, M., Lernmark, A., Mustjoki, S., Otonkoski, T., Kere, J., Morgan, N. G., Ellard, S. & Hattersley, A. T. 2014b. Activating germline mutations in STAT3 cause early-onset multi-organ autoimmune disease. *Nat Genet,* 46**,** 812-814.

68. Fujikura, J., Hosoda, K., Iwakura, H., Tomita, T., Noguchi, M., Masuzaki, H., Tanigaki, K., Yabe, D., Honjo, T. & Nakao, K. 2006. Notch/Rbp-j signaling prevents premature endocrine and ductal cell differentiation in the pancreas. *Cell Metab,* 3**,** 59-65.

69. Fukuda, A., Kawaguchi, Y., Furuyama, K., Kodama, S., Horiguchi, M., Kuhara, T., Kawaguchi, M., Terao, M., Doi, R., Wright, C. V., Hoshino, M., Chiba, T. & Uemoto, S. 2008. Reduction of Ptf1a gene dosage causes pancreatic hypoplasia and diabetes in mice. *Diabetes,* 57**,** 2421-31.

70. Furuta, H., Furuta, M., Sanke, T., Ekawa, K., Hanabusa, T., Nishi, M., Sasaki, H. & Nanjo, K. 2002. Nonsense and missense mutations in the human hepatocyte nuclear factor-1 beta gene (TCF2) and their relation to type 2 diabetes in Japanese. *J Clin Endocrinol Metab,* 87**,** 3859-63.

71. Gaffney, D. J., Mcvicker, G., Pai, A. A., Fondufe-Mittendorf, Y. N., Lewellen, N., Michelini, K., Widom, J., Gilad, Y. & Pritchard, J. K. 2012. Controls of nucleosome positioning in the human genome. *PLoS Genet,* 8**,** e1003036.

72. Gierl, M. S., Karoulias, N., Wende, H., Strehle, M. & Birchmeier, C. 2006. The zinc-finger factor Insm1 (IA-1) is essential for the development of pancreatic beta cells and intestinal endocrine cells. *Genes Dev,* 20**,** 2465-78.

73. Giresi, P. G., Kim, J., Mcdaniell, R. M., Iyer, V. R. & Lieb, J. D. 2007. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res,* 17**,** 877-85.

74. Gradwohl, G., Dierich, A., Lemeur, M. & Guillemot, F. 2000. neurogenin3 is required for the development of the four endocrine cell lineages of the pancreas. *Proc Natl Acad Sci U S A,* 97**,** 1607-11.

75. Grapin-Botton, A. 2005. Antero-posterior patterning of the vertebrate digestive tract: 40 years after Nicole Le Douarin's PhD thesis. *Int J Dev Biol,* 49**,** 335-47.

76. Greggio, C., De Franceschi, F., Figueiredo-Larsen, M., Gobaa, S., Ranga, A., Semb, H., Lutolf, M. & Grapin-Botton, A. 2013. Artificial three-dimensional niches deconstruct pancreas development in vitro. *Development,* 140**,** 4452-62.

77. 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.

78. Guariguata, L., Whiting, D. R., Hambleton, I., Beagley, J., Linnenkamp, U. & Shaw, J. E. 2014. Global estimates of diabetes prevalence for 2013 and projections for 2035. *Diabetes Res Clin Pract,* 103**,** 137-49.

79. Gudmundsson, J., Sulem, P., Steinthorsdottir, V., Bergthorsson, J. T., Thorleifsson, G., Manolescu, A., Rafnar, T., Gudbjartsson, D., Agnarsson, B. A., Baker, A., Sigurdsson, A., Benediktsdottir, K. R., Jakobsdottir, M., Blondal, T., Stacey, S. N., Helgason, A., Gunnarsdottir, S., Olafsdottir, A., Kristinsson, K. T., Birgisdottir, B., Ghosh, S., Thorlacius, S., Magnusdottir, D., Stefansdottir, G., Kristjansson, K., Bagger, Y., Wilensky, R. L., Reilly, M. P., Morris, A. D., Kimber, C. H., Adeyemo, A., Chen, Y., Zhou, J., So, W. Y., Tong, P. C., Ng, M. C., Hansen, T., Andersen, G., Borch-Johnsen, K., Jorgensen, T., Tres, A., Fuertes, F., Ruiz-Echarri, M., Asin, L., Saez, B., Van Boven, E., Klaver, S., Swinkels, D. W., Aben, K. K., Graif, T., Cashy, J., Suarez, B. K., Van Vierssen Trip, O., Frigge, M. L., Ober, C., Hofker, M. H., Wijmenga, C., Christiansen, C., Rader, D. J., Palmer, C. N., Rotimi, C., Chan, J. C., Pedersen, O., Sigurdsson, G., Benediktsson, R., Jonsson, E., Einarsson, G. V., Mayordomo, J. I., Catalona, W. J., Kiemeney, L. A., Barkardottir, R. B., Gulcher, J. R., Thorsteinsdottir, U., Kong, A. & Stefansson, K. 2007. Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet,* 39**,** 977-83.

80. Gupta, D., Kono, T. & Evans-Molina, C. 2010. The role of peroxisome proliferator-activated receptor gamma in pancreatic beta cell function and survival: therapeutic implications for the treatment of type 2 diabetes mellitus. *Diabetes Obes Metab,* 12**,** 1036-47.

81. Haldorsen, I. S., Vesterhus, M., Raeder, H., Jensen, D. K., Sovik, O., Molven, A. & Njolstad, P. R. 2008. Lack of pancreatic body and tail in HNF1B mutation carriers. *Diabet Med,* 25**,** 782-7.

82. Harries, L. W., Bingham, C., Bellanne-Chantelot, C., Hattersley, A. T. & Ellard, S. 2005. The position of premature termination codons in the hepatocyte nuclear factor -1 beta gene determines susceptibility to nonsense-mediated decay. *Hum Genet,* 118**,** 214-24.

83. Harries, L. W., Ellard, S., Jones, R. W., Hattersley, A. T. & Bingham, C. 2004. Abnormal splicing of hepatocyte nuclear factor-1 beta in the renal cysts and diabetes syndrome. *Diabetologia,* 47**,** 937-42.

84. 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.

85. Haumaitre, C., Fabre, M., Cormier, S., Baumann, C., Delezoide, A. L. & Cereghini, S. 2006. Severe pancreas hypoplasia and multicystic renal dysplasia in two human fetuses carrying novel HNF1beta/MODY5 mutations. *Hum Mol Genet,* 15**,** 2363-75.

86. Hebrok, M., Kim, S. K. & Melton, D. A. 1998. Notochord repression of endodermal Sonic hedgehog permits pancreas development. *Genes Dev,* 12**,** 1705-13.

87. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H. & Glass, C. K. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell,* 38**,** 576-89.

88. Hex, N., Bartlett, C., Wright, D., Taylor, M. & Varley, D. 2012. Estimating the current and future costs of Type 1 and Type 2 diabetes in the UK, including direct health costs and indirect societal and productivity costs. *Diabet Med,* 29**,** 855-62.

89. 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.

90. Hua, H., Shang, L., Martinez, H., Freeby, M., Gallagher, M. P., Ludwig, T., Deng, L., Greenberg, E., Leduc, C., Chung, W. K., Goland, R., Leibel, R. L. & Egli, D. 2013. iPSC-derived beta cells model diabetes due to glucokinase deficiency. *J Clin Invest,* 123**,** 3146-53.

91. 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**,** 3514-22.

92. Jennings, R. E., Berry, A. A., Strutt, J. P., Gerrard, D. T. & Hanley, N. A. 2015. Human pancreas development. *Development,* 142**,** 3126-37.

93. Jensen, J., Heller, R. S., Funder-Nielsen, T., Pedersen, E. E., Lindsell, C., Weinmaster, G., Madsen, O. D. & Serup, P. 2000. Independent development of pancreatic alpha- and beta-cells from neurogenin3-expressing precursors: a role for the notch pathway in repression of premature differentiation. *Diabetes,* 49**,** 163-76.

94. Jensen, J. N., Rosenberg, L. C., Hecksher-Sorensen, J. & Serup, P. 2007. Mutant neurogenin-3 in congenital malabsorptive diarrhea. *N Engl J Med,* 356**,** 1781-2; author reply 1782.

95. Johnson, J. D., Ahmed, N. T., Luciani, D. S., Han, Z., Tran, H., Fujita, J., Misler, S., Edlund, H. & Polonsky, K. S. 2003. Increased islet apoptosis in Pdx1+/- mice. *J Clin Invest,* 111**,** 1147-60.

96. Jonsson, J., Carlsson, L., Edlund, T. & Edlund, H. 1994. Insulin-promoter-factor 1 is required for pancreas development in mice. *Nature,* 371**,** 606-9.

97. Kaestner, K. H. 2010. The FoxA factors in organogenesis and differentiation. *Curr Opin Genet Dev,* 20**,** 527-32.

98. Kalinichenko, V. V., Zhou, Y., Bhattacharyya, D., Kim, W., Shin, B., Bambal, K. & Costa, R. H. 2002. Haploinsufficiency of the mouse Forkhead Box f1 gene causes defects in gall bladder development. *J Biol Chem,* 277**,** 12369-74.

99. 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.

100. Kettunen, J. L. T., Parviainen, H., Miettinen, P. J., Farkkila, M., Tamminen, M., Salonen, P., Lantto, E. & Tuomi, T. 2017. Biliary Anomalies in Patients With HNF1B Diabetes. *J Clin Endocrinol Metab,* 102**,** 2075-2082.

101. Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J. S. 2014. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res,* 24**,** 1012-9.

102. Kim, S. K., Hebrok, M. & Melton, D. A. 1997. Notochord to endoderm signaling is required for pancreas development. *Development,* 124**,** 4243-52.

103. Kitanaka, S., Miki, Y., Hayashi, Y. & Igarashi, T. 2004. Promoter-specific repression of hepatocyte nuclear factor (HNF)-1 beta and HNF-1 alpha transcriptional activity by an HNF-1 beta missense mutant associated with Type 5 maturity-onset diabetes of the young with hepatic and biliary manifestations. *J Clin Endocrinol Metab,* 89**,** 1369-78.

104. Kornberg, R. D. 1974. Chromatin structure: a repeating unit of histones and DNA. *Science,* 184**,** 868-71.

105. Kornfeld, J. W., Baitzel, C., Konner, A. C., Nicholls, H. T., Vogt, M. C., Herrmanns, K., Scheja, L., Haumaitre, C., Wolf, A. M., Knippschild, U., Seibler, J., Cereghini, S., Heeren, J., Stoffel, M. & Bruning, J. C. 2013. Obesity-induced overexpression of miR-802 impairs glucose metabolism through silencing of Hnf1b. *Nature,* 494**,** 111-5.

106. Kotalova, R., Dusatkova, P., Cinek, O., Dusatkova, L., Dedic, T., Seeman, T., Lebl, J. & Pruhova, S. 2015. Hepatic phenotypes of HNF1B gene mutations: a case of neonatal cholestasis requiring portoenterostomy and literature review. *World J Gastroenterol,* 21**,** 2550-7.

107. 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.

108. Larsen, H. L. & Grapin-Botton, A. 2017. The molecular and morphogenetic basis of pancreas organogenesis. *Semin Cell Dev Biol,* 66**,** 51-68.

109. Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M. T. & Carey, V. J. 2013. Software for computing and annotating genomic ranges. *PLoS Comput Biol,* 9**,** e1003118.

110. Ledig, S., Schippert, C., Strick, R., Beckmann, M. W., Oppelt, P. G. & Wieacker, P. 2011. Recurrent aberrations identified by array-CGH in patients with Mayer-Rokitansky-Kuster-Hauser syndrome. *Fertil Steril,* 95**,** 1589-94.

111. Lee, D. H., Ko, J. J., Ji, Y. G., Chung, H. M. & Hwang, T. 2012. Proteomic identification of RREB1, PDE6B, and CD209 up-regulated in primitive gut tube differentiated from human embryonic stem cells. *Pancreas,* 41**,** 65-73.

112. Lee, T. I. & Young, R. A. 2013. Transcriptional regulation and its misregulation in disease. *Cell,* 152**,** 1237-51.

113. Lerner, J., Bagattin, A., Verdeguer, F., Makinistoglu, M. P., Garbay, S., Felix, T., Heidet, L. & Pontoglio, M. 2016. Human mutations affect the epigenetic/bookmarking function of HNF1B. *Nucleic Acids Res,* 44**,** 8097-111.

114. Lewis, S. L. & Tam, P. P. 2006. Definitive endoderm of the mouse embryo: formation, cell fates, and morphogenetic function. *Dev Dyn,* 235**,** 2315-29.

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

116. Loirat, C., Bellanne-Chantelot, C., Husson, I., Deschenes, G., Guigonis, V. & Chabane, N. 2010. Autism in three patients with cystic or hyperechogenic kidneys and chromosome 17q12 deletion. *Nephrol Dial Transplant,* 25**,** 3430-3.

117. Lokmane, L., Haumaitre, C., Garcia-Villalba, P., Anselme, I., Schneider-Maunoury, S. & Cereghini, S. 2008. Crucial role of vHNF1 in vertebrate hepatic specification. *Development,* 135**,** 2777-86.

118. 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.

119. 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.

120. Malecki, M. T. 2010. The search for undiagnosed MODY patients: what is the next step? *Diabetologia,* 53**,** 2465-7.

121. Mccracken, K. W. & Wells, J. M. 2012. Molecular pathways controlling pancreas induction. *Semin Cell Dev Biol,* 23**,** 656-62.

122. Mefford, H. C., Clauin, S., Sharp, A. J., Moller, R. S., Ullmann, R., Kapur, R., Pinkel, D., Cooper, G. M., Ventura, M., Ropers, H. H., Tommerup, N., Eichler, E. E. & Bellanne-Chantelot, C. 2007. Recurrent reciprocal genomic rearrangements of 17q12 are associated with renal disease, diabetes, and epilepsy. *Am J Hum Genet,* 81**,** 1057-69.

123. Mendel, D. B. & Crabtree, G. R. 1991. HNF-1, a member of a novel class of dimerizing homeodomain proteins. *J Biol Chem,* 266**,** 677-80.

124. Mendel, D. B., Hansen, L. P., Graves, M. K., Conley, P. B. & Crabtree, G. R. 1991. HNF-1 alpha and HNF-1 beta (vHNF-1) share dimerization and homeo domains, but not activation domains, and form heterodimers in vitro. *Genes Dev,* 5**,** 1042-56.

125. Molotkov, A., Molotkova, N. & Duester, G. 2005. Retinoic acid generated by Raldh2 in mesoderm is required for mouse dorsal endodermal pancreas development. *Dev Dyn,* 232**,** 950-7.

126. Mootha, V. K., Lindgren, C. M., Eriksson, K. F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstrale, M., Laurila, E., Houstis, N., Daly, M. J., Patterson, N., Mesirov, J. P., Golub, T. R., Tamayo, P., Spiegelman, B., Lander, E. S., Hirschhorn, J. N., Altshuler, D. & Groop, L. C. 2003. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet,* 34**,** 267-73.

127. Moreno-De-Luca, D., Mulle, J. G., Kaminsky, E. B., Sanders, S. J., Genestar, Myers, S. M., Adam, M. P., Pakula, A. T., Eisenhauer, N. J., Uhas, K., Weik, L., Guy, L., Care, M. E., Morel, C. F., Boni, C., Salbert, B. A., Chandrareddy, A., Demmer, L. A., Chow, E. W., Surti, U., Aradhya, S., Pickering, D. L., Golden, D. M., Sanger, W. G., Aston, E., Brothman, A. R., Gliem, T. J., Thorland, E. C., Ackley, T., Iyer, R., Huang, S., Barber, J. C., Crolla, J. A., Warren, S. T., Martin, C. L. & Ledbetter, D. H. 2010. Deletion 17q12 is a recurrent copy number variant that confers high risk of autism and schizophrenia. *Am J Hum Genet,* 87**,** 618-30.

128. Morris, A. P., Voight, B. F., Teslovich, T. M., Ferreira, T., Segre, A. V., Steinthorsdottir, V., Strawbridge, R. J., Khan, H., Grallert, H., Mahajan, A., Prokopenko, I., Kang, H. M., Dina,

C., Esko, T., Fraser, R. M., Kanoni, S., Kumar, A., Lagou, V., Langenberg, C., Luan, J., Lindgren, C. M., Muller-Nurasyid, M., Pechlivanis, S., Rayner, N. W., Scott, L. J., Wiltshire, S., Yengo, L., Kinnunen, L., Rossin, E. J., Raychaudhuri, S., Johnson, A. D., Dimas, A. S., Loos, R. J., Vedantam, S., Chen, H., Florez, J. C., Fox, C., Liu, C. T., Rybin, D., Couper, D. J., Kao, W. H., Li, M., Cornelis, M. C., Kraft, P., Sun, Q., Van Dam, R. M., Stringham, H. M., Chines, P. S., Fischer, K., Fontanillas, P., Holmen, O. L., Hunt, S. E., Jackson, A. U., Kong, A., Lawrence, R., Meyer, J., Perry, J. R., Platou, C. G., Potter, S., Rehnberg, E., Robertson, N., Sivapalaratnam, S., Stancakova, A., Stirrups, K., Thorleifsson, G., Tikkanen, E., Wood, A. R., Almgren, P., Atalay, M., Benediktsson, R., Bonnycastle, L. L., Burtt, N., Carey, J., Charpentier, G., Crenshaw, A. T., Doney, A. S., Dorkhan, M., Edkins, S., Emilsson, V., Eury, E., Forsen, T., Gertow, K., Gigante, B., Grant, G. B., Groves, C. J., Guiducci, C., Herder, C., Hreidarsson, A. B., Hui, J., James, A., Jonsson, A., Rathmann, W., Klopp, N., Kravic, J., Krjutskov, K., Langford, C., Leander, K., Lindholm, E., Lobbens, S., Mannisto, S., et al. 2012. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet,* 44**,** 981-90.

129. Muraro, M. J., Dharmadhikari, G., Grun, D., Groen, N., Dielen, T., Jansen, E., Van Gurp, L., Engelse, M. A., Carlotti, F., De Koning, E. J. & Van Oudenaarden, A. 2016. A Single-Cell Transcriptome Atlas of the Human Pancreas. *Cell Syst,* 3**,** 385-394 e3.

130. Murphy, R., Ellard, S. & Hattersley, A. T. 2008. Clinical implications of a molecular genetic classification of monogenic beta-cell diabetes. *Nat Clin Pract Endocrinol Metab,* 4**,** 200-13.

131. Murtaugh, L. C., Law, A. C., Dor, Y. & Melton, D. A. 2005. Beta-catenin is essential for pancreatic acinar but not islet development. *Development,* 132**,** 4663-74.

132. Naya, F. J., Huang, H. P., Qiu, Y., Mutoh, H., Demayo, F. J., Leiter, A. B. & Tsai, M. J. 1997. Diabetes, defective pancreatic morphogenesis, and abnormal enteroendocrine differentiation in BETA2/neuroD-deficient mice. *Genes Dev,* 11**,** 2323-34.

133. Nostro, M. C., Sarangi, F., Yang, C., Holland, A., Elefanty, A. G., Stanley, E. G., Greiner, D. L. & Keller, G. 2015. Efficient generation of NKX6-1+ pancreatic progenitors from multiple human pluripotent stem cell lines. *Stem Cell Reports,* 4**,** 591-604.

134. 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.

135. Ohlsson, H., Karlsson, K. & Edlund, T. 1993. IPF1, a homeodomain-containing transactivator of the insulin gene. *EMBO J,* 12**,** 4251-9.

136. Oliver-Krasinski, J. M. & Stoffers, D. A. 2008. On the origin of the beta cell. *Genes Dev,* 22**,** 1998-2021.

137. Oram, R. A., Edghill, E. L., Blackman, J., Taylor, M. J., Kay, T., Flanagan, S. E., Ismail-Pratt, I., Creighton, S. M., Ellard, S., Hattersley, A. T. & Bingham, C. 2010. Mutations in the hepatocyte nuclear factor-1beta (HNF1B) gene are common with combined uterine and renal malformations but are not found with isolated uterine malformations. *Am J Obstet Gynecol,* 203**,** 364 e1-5.

138. 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.

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

140. Pearson, E. R., Badman, M. K., Lockwood, C. R., Clark, P. M., Ellard, S., Bingham, C. & Hattersley, A. T. 2004. Contrasting diabetes phenotypes associated with hepatocyte nuclear factor-1alpha and -1beta mutations. *Diabetes Care,* 27**,** 1102-7.

141. Piccand, J., Strasser, P., Hodson, D. J., Meunier, A., Ye, T., Keime, C., Birling, M. C., Rutter, G. A. & Gradwohl, G. 2014. Rfx6 maintains the functional identity of adult pancreatic beta cells. *Cell Rep,* 9**,** 2219-32.

142. Piper, J., Assi, S. A., Cauchy, P., Ladroue, C., Cockerill, P. N., Bonifer, C. & Ott, S. 2015. Wellington-bootstrap: differential DNase-seq footprinting identifies cell-type determining transcription factors. *BMC Genomics,* 16**,** 1000.

143. 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. *Mech Dev,* 116**,** 223-6.

144. 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.

145. Poll, A. V., Pierreux, C. E., Lokmane, L., Haumaitre, C., Achouri, Y., Jacquemin, P., Rousseau, G. G., Cereghini, S. & Lemaigre, F. P. 2006. A vHNF1/TCF2-HNF6 cascade regulates the transcription factor network that controls generation of pancreatic precursor cells. *Diabetes,* 55**,** 61-9.

146. Ponts, N., Harris, E. Y., Prudhomme, J., Wick, I., Eckhardt-Ludka, C., Hicks, G. R., Hardiman, G., Lonardi, S. & Le Roch, K. G. 2010. Nucleosome landscape and control of transcription in the human malaria parasite. *Genome Res,* 20**,** 228-38.

147. Radman-Livaja, M. & Rando, O. J. 2010. Nucleosome positioning: how is it established, and why does it matter? *Dev Biol,* 339**,** 258-66.

148. Rahier, J., Guiot, Y., Goebbels, R. M., Sempoux, C. & Henquin, J. C. 2008. Pancreatic beta-cell mass in European subjects with type 2 diabetes. *Diabetes Obes Metab,* 10 Suppl 4**,** 32-42.

149. Raile, K., Klopocki, E., Holder, M., Wessel, T., Galler, A., Deiss, D., Muller, D., Riebel, T., Horn, D., Maringa, M., Weber, J., Ullmann, R. & Gruters, A. 2009. Expanded clinical spectrum in hepatocyte nuclear factor 1b-maturity-onset diabetes of the young. *J Clin Endocrinol Metab,* 94**,** 2658-64.

150. Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A. & Zhang, F. 2013. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc,* 8**,** 2281-308.

151. 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. *J Clin Invest,* 120**,** 3127-36.

152. Rebouissou, S., Vasiliu, V., Thomas, C., Bellanne-Chantelot, C., Bui, H., Chretien, Y., Timsit, J., Rosty, C., Laurent-Puig, P., Chauveau, D. & Zucman-Rossi, J. 2005. Germline

hepatocyte nuclear factor 1alpha and 1beta mutations in renal cell carcinomas. *Hum Mol Genet,* 14**,** 603-14.

153. 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., Johnson, J. D. & Kieffer, T. J. 2014. Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. *Nat Biotechnol,* 32**,** 1121-33.

154. Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W. & Smyth, G. K. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res,* 43**,** e47.

155. Roelandt, P., Antoniou, A., Libbrecht, L., Van Steenbergen, W., Laleman, W., Verslype, C., Van Der Merwe, S., Nevens, F., De Vos, R., Fischer, E., Pontoglio, M., Lemaigre, F. & Cassiman, D. 2012. HNF1B deficiency causes ciliary defects in human cholangiocytes. *Hepatology,* 56**,** 1178-81.

156. Rossi, J. M., Dunn, N. R., Hogan, B. L. & Zaret, K. S. 2001. Distinct mesodermal signals, including BMPs from the septum transversum mesenchyme, are required in combination for hepatogenesis from the endoderm. *Genes Dev,* 15**,** 1998-2009.

157. Rubio-Cabezas, O., Hattersley, A. T., Njolstad, P. R., Mlynarski, W., Ellard, S., White, N., Chi, D. V. & Craig, M. E. 2014. The diagnosis and management of monogenic diabetes in children and adolescents. *Pediatr Diabetes,* 15 Suppl 20**,** 47-64.

158. 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.

159. Saarimaki-Vire, J., Balboa, D., Russell, M. A., Saarikettu, J., Kinnunen, M., Keskitalo, S., Malhi, A., Valensisi, C., Andrus, C., Eurola, S., Grym, H., Ustinov, J., Wartiovaara, K., Hawkins, R. D., Silvennoinen, O., Varjosalo, M., Morgan, N. G. & Otonkoski, T. 2017. An Activating STAT3 Mutation Causes Neonatal Diabetes through Premature Induction of Pancreatic Differentiation. *Cell Rep,* 19**,** 281-294.

160. Santos, A., Wernersson, R. & Jensen, L. J. 2015. Cyclebase 3.0: a multi-organism database on cell-cycle regulation and phenotypes. *Nucleic Acids Res,* 43**,** D1140-4.

161. Schaffer, A. E., Freude, K. K., Nelson, S. B. & Sander, M. 2010. Nkx6 transcription factors and Ptf1a function as antagonistic lineage determinants in multipotent pancreatic progenitors. *Dev Cell,* 18**,** 1022-9.

162. Semple, R. K., Savage, D. B., Cochran, E. K., Gorden, P. & O'rahilly, S. 2011. Genetic syndromes of severe insulin resistance. *Endocr Rev,* 32**,** 498-514.

163. Serls, A. E., Doherty, S., Parvatiyar, P., Wells, J. M. & Deutsch, G. H. 2005. Different thresholds of fibroblast growth factors pattern the ventral foregut into liver and lung. *Development,* 132**,** 35-47.

164. Servitja, J. M. & Ferrer, J. 2004. Transcriptional networks controlling pancreatic development and beta cell function. *Diabetologia,* 47**,** 597-613.

165. 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.

166. Shen, L., Shao, N. Y., Liu, X., Maze, I., Feng, J. & Nestler, E. J. 2013. diffReps: detecting differential chromatin modification sites from ChIP-seq data with biological replicates. *PLoS One,* 8**,** e65598.

167. Sherwood, R. I., Chen, T. Y. & Melton, D. A. 2009. Transcriptional dynamics of endodermal organ formation. *Dev Dyn,* 238**,** 29-42.

168. 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,* 20**,** 675-688 e6.

169. Shields, B. M., Hicks, S., Shepherd, M. H., Colclough, K., Hattersley, A. T. & Ellard, S. 2010. Maturity-onset diabetes of the young (MODY): how many cases are we missing? *Diabetologia,* 53**,** 2504-8.

170. Shlyueva, D., Stampfel, G. & Stark, A. 2014. Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet,* 15**,** 272-86.

171. Smith, S. B., Qu, H. Q., Taleb, N., Kishimoto, N. Y., Scheel, D. W., Lu, Y., Patch, A. M., Grabs, R., Wang, J., Lynn, F. C., Miyatsuka, T., Mitchell, J., Seerke, R., Desir, J., Vanden Eijnden, S., Abramowicz, M., Kacet, N., Weill, J., Renard, M. E., Gentile, M., Hansen, I., Dewar, K., Hattersley, A. T., Wang, R., Wilson, M. E., Johnson, J. D., Polychronakos, C. & German, M. S. 2010. Rfx6 directs islet formation and insulin production in mice and humans. *Nature,* 463**,** 775-80.

172. 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.

173. Soyer, J., Flasse, L., Raffelsberger, W., Beucher, A., Orvain, C., Peers, B., Ravassard, P., Vermot, J., Voz, M. L., Mellitzer, G. & Gradwohl, G. 2010. Rfx6 is an Ngn3-dependent winged helix transcription factor required for pancreatic islet cell development. *Development,* 137**,** 203-12.

174. Spagnoli, F. M. & Brivanlou, A. H. 2008. The Gata5 target, TGIF2, defines the pancreatic region by modulating BMP signals within the endoderm. *Development,* 135**,** 451-61.

175. Spence, J. R., Lange, A. W., Lin, S. C., Kaestner, K. H., Lowy, A. M., Kim, I., Whitsett, J. A. & Wells, J. M. 2009. Sox17 regulates organ lineage segregation of ventral foregut progenitor cells. *Dev Cell,* 17**,** 62-74.

176. Spence, J. R. & Wells, J. M. 2007. Translational embryology: using embryonic principles to generate pancreatic endocrine cells from embryonic stem cells. *Dev Dyn,* 236**,** 3218-27.

177. Spurdle, A. B., Thompson, D. J., Ahmed, S., Ferguson, K., Healey, C. S., O'mara, T., Walker, L. C., Montgomery, S. B., Dermitzakis, E. T., Australian National Endometrial Cancer

Study, G., Fahey, P., Montgomery, G. W., Webb, P. M., Fasching, P. A., Beckmann, M. W., Ekici, A. B., Hein, A., Lambrechts, D., Coenegrachts, L., Vergote, I., Amant, F., Salvesen, H. B., Trovik, J., Njolstad, T. S., Helland, H., Scott, R. J., Ashton, K., Proietto, T., Otton, G., National Study of Endometrial Cancer Genetics, G., Tomlinson, I., Gorman, M., Howarth, K., Hodgson, S., Garcia-Closas, M., Wentzensen, N., Yang, H., Chanock, S., Hall, P., Czene, K., Liu, J., Li, J., Shu, X. O., Zheng, W., Long, J., Xiang, Y. B., Shah, M., Morrison, J., Michailidou, K., Pharoah, P. D., Dunning, A. M. & Easton, D. F. 2011. Genome-wide association study identifies a common variant associated with risk of endometrial cancer. *Nat Genet,* 43**,** 451-4.

178. Stafford, D. & Prince, V. E. 2002. Retinoic acid signaling is required for a critical early step in zebrafish pancreatic development. *Curr Biol,* 12**,** 1215-20.

179. Stanger, B. Z., Tanaka, A. J. & Melton, D. A. 2007. Organ size is limited by the number of embryonic progenitor cells in the pancreas but not the liver. *Nature,* 445**,** 886-91.

180. Stekelenburg, C. M. & Schwitzgebel, V. M. 2016. Genetic Defects of the beta-Cell That Cause Diabetes. *Endocr Dev,* 31**,** 179-202.

181. Stepniewski, J., Kachamakova-Trojanowska, N., Ogrocki, D., Szopa, M., Matlok, M., Beilharz, M., Dyduch, G., Malecki, M. T., Jozkowicz, A. & Dulak, J. 2015. Induced pluripotent stem cells as a model for diabetes investigation. *Sci Rep,* 5**,** 8597.

182. Sterneckert, J. L., Reinhardt, P. & Scholer, H. R. 2014. Investigating human disease using stem cell models. *Nat Rev Genet,* 15**,** 625-39.

183. 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.

184. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S. & Mesirov, J. P. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A,* 102**,** 15545-50.

185. Sun, J., Zheng, S. L., Wiklund, F., Isaacs, S. D., Purcell, L. D., Gao, Z., Hsu, F. C., Kim, S. T., Liu, W., Zhu, Y., Stattin, P., Adami, H. O., Wiley, K. E., Dimitrov, L., Sun, J., Li, T., Turner, A. R., Adams, T. S., Adolfsson, J., Johansson, J. E., Lowey, J., Trock, B. J., Partin, A. W., Walsh, P. C., Trent, J. M., Duggan, D., Carpten, J., Chang, B. L., Gronberg, H., Isaacs, W. B. & Xu, J. 2008. Evidence for two independent prostate cancer risk-associated loci in the HNF1B gene at 17q12. *Nat Genet,* 40**,** 1153-5.

186. Taylor, B. L., Liu, F. F. & Sander, M. 2013. Nkx6.1 is essential for maintaining the functional state of pancreatic beta cells. *Cell Rep,* 4**,** 1262-75.

187. Teo, A. K., Lau, H. H., Valdez, I. A., Dirice, E., Tjora, E., Raeder, H. & Kulkarni, R. N. 2016. Early Developmental Perturbations in a Human Stem Cell Model of MODY5/HNF1B Pancreatic Hypoplasia. *Stem Cell Reports,* 6**,** 357-67.

188. Teo, A. K., Wagers, A. J. & Kulkarni, R. N. 2013. New opportunities: harnessing induced pluripotency for discovery in diabetes and metabolism. *Cell Metab,* 18**,** 775-91.

189. Thomas, G., Jacobs, K. B., Yeager, M., Kraft, P., Wacholder, S., Orr, N., Yu, K., Chatterjee, N., Welch, R., Hutchinson, A., Crenshaw, A., Cancel-Tassin, G., Staats, B. J.,

Wang, Z., Gonzalez-Bosquet, J., Fang, J., Deng, X., Berndt, S. I., Calle, E. E., Feigelson, H. S., Thun, M. J., Rodriguez, C., Albanes, D., Virtamo, J., Weinstein, S., Schumacher, F. R., Giovannucci, E., Willett, W. C., Cussenot, O., Valeri, A., Andriole, G. L., Crawford, E. D., Tucker, M., Gerhard, D. S., Fraumeni, J. F., Jr., Hoover, R., Hayes, R. B., Hunter, D. J. & Chanock, S. J. 2008. Multiple loci identified in a genome-wide association study of prostate cancer. *Nat Genet,* 40**,** 310-5.

190. 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.

191. Toyoda, T., Mae, S., Tanaka, H., Kondo, Y., Funato, M., Hosokawa, Y., Sudo, T., Kawaguchi, Y. & Osafune, K. 2015. Cell aggregation optimizes the differentiation of human ESCs and iPSCs into pancreatic bud-like progenitor cells. *Stem Cell Res,* 14**,** 185-97.

192. Tsompana, M. & Buck, M. J. 2014. Chromatin accessibility: a window into the genome. *Epigenetics Chromatin,* 7**,** 33.

193. Tsuchiya, A., Sakamoto, M., Yasuda, J., Chuma, M., Ohta, T., Ohki, M., Yasugi, T., Taketani, Y. & Hirohashi, S. 2003. Expression profiling in ovarian clear cell carcinoma: identification of hepatocyte nuclear factor-1 beta as a molecular marker and a possible molecular target for therapy of ovarian clear cell carcinoma. *Am J Pathol,* 163**,** 2503-12.

194. Tulpule, A., Kelley, J. M., Lensch, M. W., Mcpherson, J., Park, I. H., Hartung, O., Nakamura, T., Schlaeger, T. M., Shimamura, A. & Daley, G. Q. 2013. Pluripotent stem cell models of Shwachman-Diamond syndrome reveal a common mechanism for pancreatic and hematopoietic dysfunction. *Cell Stem Cell,* 12**,** 727-36.

195. Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. & Luscombe, N. M. 2009. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet,* 10**,** 252-63.

196. Vegas, A. J., Veiseh, O., Gurtler, M., Millman, J. R., Pagliuca, F. W., Bader, A. R., Doloff, J. C., Li, J., Chen, M., Olejnik, K., Tam, H. H., Jhunjhunwala, S., Langan, E., Aresta-Dasilva, S., Gandham, S., Mcgarrigle, J. J., Bochenek, M. A., Hollister-Lock, J., Oberholzer, J., Greiner, D. L., Weir, G. C., Melton, D. A., Langer, R. & Anderson, D. G. 2016. Long-term glycemic control using polymer-encapsulated human stem cell-derived beta cells in immune-competent mice. *Nat Med,* 22**,** 306-11.

197. Verdeguer, F., Le Corre, S., Fischer, E., Callens, C., Garbay, S., Doyen, A., Igarashi, P., Terzi, F. & Pontoglio, M. 2010. A mitotic transcriptional switch in polycystic kidney disease. *Nat Med,* 16**,** 106-10.

198. Vethe, H., Bjorlykke, Y., Ghila, L. M., Paulo, J. A., Scholz, H., Gygi, S. P., Chera, S. & Raeder, H. 2017. Probing the missing mature beta-cell proteomic landscape in differentiating patient iPSC-derived cells. *Sci Rep,* 7**,** 4780.

199. Vivante, A., Kohl, S., Hwang, D. Y., Dworschak, G. C. & Hildebrandt, F. 2014. Single-gene causes of congenital anomalies of the kidney and urinary tract (CAKUT) in humans. *Pediatr Nephrol,* 29**,** 695-704.

200. Voss, T. C. & Hager, G. L. 2014. Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nat Rev Genet,* 15**,** 69-81.

201. Wang, L., Coffinier, C., Thomas, M. K., Gresh, L., Eddu, G., Manor, T., Levitsky, L. L., Yaniv, M. & Rhoads, D. B. 2004. Selective deletion of the Hnf1beta (MODY5) gene in beta-cells leads to altered gene expression and defective insulin release. *Endocrinology,* 145**,** 3941-9.

202. Wells, J. M., Esni, F., Boivin, G. P., Aronow, B. J., Stuart, W., Combs, C., Sklenka, A., Leach, S. D. & Lowy, A. M. 2007. Wnt/beta-catenin signaling is required for development of the exocrine pancreas. *BMC Dev Biol,* 7**,** 4.

203. Wells, J. M. & Melton, D. A. 1999. Vertebrate endoderm development. *Annu Rev Cell Dev Biol,* 15**,** 393-410.

204. Winckler, W., Weedon, M. N., Graham, R. R., Mccarroll, S. A., Purcell, S., Almgren, P., Tuomi, T., Gaudet, D., Bostrom, K. B., Walker, M., Hitman, G., Hattersley, A. T., Mccarthy, M. I., Ardlie, K. G., Hirschhorn, J. N., Daly, M. J., Frayling, T. M., Groop, L. & Altshuler, D. 2007. Evaluation of common variants in the six known maturity-onset diabetes of the young (MODY) genes for association with type 2 diabetes. *Diabetes,* 56**,** 685-93.

205. Wu, S. M. & Hochedlinger, K. 2011. Harnessing the potential of induced pluripotent stem cells for regenerative medicine. *Nat Cell Biol,* 13**,** 497-505.

206. Xi, H., Shulha, H. P., Lin, J. M., Vales, T. R., Fu, Y., Bodine, D. M., Mckay, R. D., Chenoweth, J. G., Tesar, P. J., Furey, T. S., Ren, B., Weng, Z. & Crawford, G. E. 2007. Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome. *PLoS Genet,* 3**,** e136.

207. Xie, R., Everett, L. J., Lim, H. W., Patel, N. A., Schug, J., Kroon, E., Kelly, O. G., Wang, A., D'amour, K. A., Robins, A. J., Won, K. J., Kaestner, K. H. & Sander, M. 2013. Dynamic chromatin remodeling mediated by polycomb proteins orchestrates pancreatic differentiation of human embryonic stem cells. *Cell Stem Cell,* 12**,** 224-37.

208. Yusa, K., Rashid, S. T., Strick-Marchand, H., Varela, I., Liu, P. Q., Paschon, D. E., Miranda, E., Ordonez, A., Hannan, N. R., 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 alpha1-antitrypsin deficiency in induced pluripotent stem cells. *Nature,* 478**,** 391-4.

209. Zaret, K. S. 2008. Genetic programming of liver and pancreas progenitors: lessons for stem-cell differentiation. *Nat Rev Genet,* 9**,** 329-40.

210. 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.

211. Zhou, Q., Law, A. C., Rajagopal, J., Anderson, W. J., Gray, P. A. & Melton, D. A. 2007. A multipotent progenitor domain guides pancreatic organogenesis. *Dev Cell,* 13**,** 103-14.

212. Zhu, Z. & Huangfu, D. 2013. Human pluripotent stem cells: an emerging model in developmental biology. *Development,* 140**,** 705-17.

213. Zhu, Z., Li, Q. V., Lee, K., Rosen, B. P., Gonzalez, F., Soh, C. L. & Huangfu, D. 2016. Genome Editing of Lineage Determinants in Human Pluripotent Stem Cells Reveals Mechanisms of Pancreatic Development and Diabetes. *Cell Stem Cell,* 18**,** 755-68.

214. Zorn, A. M. & Wells, J. M. 2009. Vertebrate endoderm development and organ formation. *Annu Rev Cell Dev Biol,* 25**,** 221-51.