Modelling Monogenic Diabetes Mutations Using Authentic Human Beta Cell Models



Claire Elizabeth Duff

Harris Manchester College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas Term 2020

Declaration

The data and interpretation detailed in this thesis represent my own work, with contributions from others specified below. Experiments were performed at the Oxford Centre for Diabetes, Endocrinology and Metabolism (OCDEM) and the Wellcome Centre for Human Genetics (WCHG) under the supervision of Professor Anna Gloyn and Dr Benoît Hastoy, with input from Dr Christian Honoré and Dr Mattias Hansson at Novo Nordisk. The project was funded by the Medical Research Council and Novo Nordisk.

Karyotyping of genome-engineered hiPSC lines was done by Dr Daniela Moralli and Mary-Anne Durin at the Chromosome Dynamics Core (WCHG) (Chapters 3 and 5). Sequencing data (Chapter 4 and Chapter 5) was generated by the Oxford Genomics Centre at the WCHG and data analysis was supported by Dr Agata Wesolowska-Andersen (WCHG), Dr Marta Pérez Alcántara (WCHG) and Dr Han Sun (Stanford University, USA).

Generation of the *INS*-KO (knockout) EndoC-βH1 line (Chapter 4) was performed by Dr Antje Grotz and EndoC-βH1 culture was supported by Sameena Nawaz. Acquirement of immunofluorescence images was assisted by Dr Benoît Hastoy (Chapter 4). Generation of the *HNF1A* KO (Chapter 5) was performed by Dr Shailesh Gupta (AstraZeneca)

This thesis is approximately 47,000 words and no part has been submitted for any other degree at this or any other university.

Acknowledgements

The last four years have given me many happy memories but also challenges, not least completing during a global pandemic! The experience would have been so much lesser if not for the encouragement I have received from friends, family and colleagues. Firstly, I would like to thank my supervisors Anna Gloyn, Benoît Hastoy, Christian Honoré and Mattias Hansson. Anna, for your support and guidance even from over 5,000 miles away. Benoît, for taking on supervising me in my final year, yet still investing so much in developing me as a scientist. I am grateful to Christian and Mattias for hosting me at Novo Nordisk and to Christian for all his constructive feedback. I would also like to thank Nicola Beer for supervising me for the first six months of my project and who made me feel so welcome when I joined the lab. Also, to Matt Simmonds, whose encouragement led to my application to this DPhil.

To the Gloyn team and beyond who provided support, laughs and shared in all things mojo when it was most needed: Agata, Amy, Anastasia, Antje, Benoite, Carla, Dani, Emily, Grace, Mandy, Jo, Joely, Katia, Phalguni, Pippa, Rikke and Shahana. I particularly want to thank Nicole, for sharing in the singing and dancing whilst in tissue culture or at the FACS machine. To Fernando, for your enthusiasm for science and calm reassurance. Marta, for being so generous with your time in helping me with R analysis. I would also like to thank Sameena who patiently listened to many a tale of experimental woe and provided a constant source of encouragement and support.

I am very grateful to those outside the lab with whom I have enjoyed spending time with, be it in person or virtually: Alice, Annie, Becky, all at the East Oxford Connect Group, Helen and Jess. Lastly to Jenni (and Oscar the cat), for providing a house which has become my home – thank you for sharing in sofa slug times, biscuit o'clock and all my other quirks.

Finally, I would like to thank my family – my mum, dad and brother, Tim. For listening to me talk about my project, even when it probably made little sense to you. More than that, for all the love and support you have provided not only during my DPhil but in every aspect of life.

Abstract

Genome-wide association studies have identified over 400 loci associated with risk for type 2 diabetes (T2D), a large proportion are associated with pancreatic islet dysfunction. The parallel emergence of genome editing technology such as CRISPR/Cas9, alongside the ability to derive endocrine pancreatic cells from human stem cells, provides unprecedented opportunity to investigate and examine the mechanisms that underpin beta cell dysfunction. Given that several monogenic genes are also associated with T2D risk, investigating the mechanisms by which beta cell dysfunction occurs in monogenic diabetes also presents the opportunity to gain additional insight into T2D pathogenesis. The aim of my thesis was to model monogenic diabetes variants in *INS* and *HNF1A* in human beta cell models using both genome-edited human induced pluripotent stem cells (hiPSCs) and EndoC-βH1 cells.

To investigate the role of *INS* mutations on beta cell function, I generated the first p.C43G *INS* hiPSC line, including a reporter line, using CRISPR/Cas9 editing and differentiated these cells along the pancreatic endocrine lineage. Gene expression analysis revealed no influence of the p.C43G mutation on either differentiation efficiency or ER stress, reflecting that differentiated hiPSCs may require further functional maturation in order to model *INS*-associated monogenic diabetes.

Expression of the p.C43G *INS* mutant in the recently generated *INS*-KO EndoC-βH1 line was appropriately packaged into secretory vesicles. Subsequent gene expression and RNA-Seq analysis showed no significant differences between cells transfected with either the wildtype or p.C43G construct, echoing the findings from hiPSC differentiations. This would suggest that the negative effect of the mutation might translate through different cellular mechanisms. However, this study revealed various technical challenges, such as with transfection efficiency that considerably reduced the ability to fully characterise p.C43G expression in beta cells.

Finally, I sought to investigate another gene involved in monogenic diabetes, *HNF1A*. As mouse models of HNF1A-diabetes do not fully recapitulate the human phenotype, I utilised a hiPSC line to characterise *HNF1A* knockout in parallel with a knockdown (KD) model of *HNF1A* within the EndoC-βH1 cell line. Loss of *HNF1A* led to decreased expression of insulin and glucagon at the final stage of differentiation. The *HNF1A* KO line, in comparison to the wildtype line, had an enrichment of differentially expressed genes involved in regulation of beta cell function, beta cell development and insulin secretion, recapitulating several features of the patient phenotype. Further comparisons with differentiated cells from a patient with an HNF1A-MODY causing variant and with my *HNF1A* KD EndoC-βH1 data, showed a significant overlap between the datasets demonstrating that many of the same gene signatures can be captured between different cellular models of *HNF1A*.

My thesis illustrates the complementarity of utilising both iPSCs and EndoC-βH1 for disease modelling with each addressing different aspects of diabetes pathogenesis. This approach, although still being optimised, provides a better understanding of the cellular physiology that underlies genetic predisposition for diabetes.

Abbreviations

ANOVA	Analysis of variance
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
BLC	Beta-like cells
BMI	Body mass index
bp	Base pair
BSA	Bovine serum albumin
cDNA	Complementary DNA
CRISPR	Clustered regularly interspaced short palindromic repeats
CT	Cycle threshold
DE	Definitive endoderm
DEG	Differentially expressed gene
DNA	Deoxyribonucleic acid
DSB	Double-strand break
EM	Electron microscopy
EN	Endocrine-like cells
ER	Endoplasmic reticulum
EP	Endocrine progenitors
EV	Empty vector
FACS	Fluorescence-activated cell sorting
FDR	False discovery rate
gDNA	Genomic DNA
GLUT	Glucose transporter
GO	Gene ontology
GSEA	Gene set enrichment analysis
GSIS	Glucose-stimulated insulin secretion
GWAS	Genome-wide association study
HDR	Homology directed repair
hESC	Human embryonic stem cells
hiPSC	Human induced pluripotent stem cells
HNF1A	Hepatocyte nuclear factor-1 alpha homeobox A
indel	Insertion/deletion
INS	Insulin
K_{ATP} channel	ATP-sensitive potassium channel

KD	Knockdown
kDa	Kilodalton
KO	Knockout
LFC	Log_2 fold change
MODY	Maturity onset diabetes of the young
NDM	Neonatal diabetes mellitus
NHEJ	Non-homologous end joining
NMD	Nonsense-mediated mRNA decay
NT	Non-targeting
PAM	Protospacer adjacent motif
PBS(T)	Phosphate buffered saline (tween)
PCA	Principal component analysis
PCR	Polymerase chain reaction
PNDM	Permanent neonatal diabetes mellitus
PFG	Posterior foregut
PGT	primitive gut tube
P/S	Penicillin-Streptomycin
QC	Quality Control
RNA	Ribonucleic Acid
RNAi	RNA interference
RT	Room temperature
RT-qPCR	Reverse transcription quantitative real-time PCR
SDM	Site directed mutagenesis
SEM	Standard error of the mean
Seq	Sequencing
sgRNA	Single guide RNA
siRNA	Small interfering RNA
ssODN	Single-stranded oligodeoxynucleotide
TIDE	Tracking of Indels by DEcomposition
T1D	Type 1 diabetes
T2D	Type 2 diabetes
UPR	Unfolded protein response
WT	Wildtype

Table of Contents

1. Introduction....1
1.1 Diabetes Mellitus....2
1.1.1 Diabetes is a worldwide health concern....2
1.1.2 Monogenic diabetes....2
1.1.3 Type 1 Diabetes (T1D)....2
1.1.4 Type 2 Diabetes (T2D) – a worldwide health concern....3
1.1.5 There are currently no disease modifying therapies for Type 2 Diabetes....4
1.1.6 Advantages of tackling diabetes using a genetic approach....5
1.2 The Pancreatic Beta Cell....8
1.2.1 Insulin and its role in glycaemic control....8
1.2.2 Glucose stimulated insulin secretion....10
1.2.3 Insulin and its role as a growth hormone....11
1.2.4 The biosynthesis of insulin....12
1.3 Monogenic diabetes....17
1.3.1 Neonatal diabetes mellitus (NDM)....17
1.3.2 Maturity onset of diabetes in the young (MODY)....19
1.4 Islet and beta cell models....21
1.4.1 Differences between rodent and human islets....21
1.4.2 Animal cell models....22
1.4.3 Human islets....23
1.4.4 EndoC-βH1....24
1.4.5 Human pluripotent stem cells....25
1.5 *INS* mutations as a cause of monogenic diabetes....30
1.5.1 Previous investigations of diabetes-associated *INS* mutations....32
1.5.2 Overexpression models of *INS* mutations....33
1.5.3 Modelling of *INS* mutations in human stem cell models....33
1.6 *HNF1A*....34
1.6.1 *HNF1A* – a transcription factor regulating genes involved in glucose transport and lipid metabolism....34
1.6.2 *HNF1A* and its role in beta cell development....35
1.6.3 *HNF1A* as a cause of MODY....36
1.6.4 *HNF1A* and its role in Type 2 Diabetes....37
1.6.5 *HNF1A* in animal models....37
1.6.6 *HNF1A* in human models....38
1.7 Aims of thesis....40
2 Materials and Methods....42

2.1 Cell Culture 43
2.1.1 General maintenance of cell lines 43
2.2 Genomic DNA extraction 45
2.2.1 DNA quantification 45
2.3 Gene expression analysis 46
2.3.1 RNA extraction 46
2.3.2 cDNA synthesis 46
2.3.3 Real-time TaqMan® quantitative PCR 46
2.4 Transcriptome profiling 47
2.4.1 RNA extraction, sequencing and quantification 47
2.4.2 Differential gene expression analysis 48
2.5 Protein expression studies 49
2.5.1 Cell Lysis 49
2.5.2 Protein Quantification 50
2.5.3 Western Blotting 50
2.5.4 Transient Transfections 51
2.5.5 Flow cytometry 52
2.6 CRISPR/Cas9 genome editing in hiPSCs 54
2.7 *In vitro* differentiation of hiPSC towards beta-like cells 56
3 Generation and characterisation of a diabetes causing *INS* mutation in an hiPSC model 61
3.1 Introduction 62
3.2 Experimental aims 64
3.3 Methods 66
3.3.1 Genome Editing 66
3.3.2 Transfection of hiPSCs 67
3.3.3 Design of ssODN repair templates 69
3.3.4 Differentiation experiments 73
3.3.5 Cell sorting using FACS 73
3.3.6 Quantitative real-time PCR (RT-qPCR) 74
3.3.7 Western Blotting analysis 74
3.4 Results 76
3.4.1 Technical challenges in generating an hiPSC p.C96Y model 76
3.4.2 Generation of hiPSC *INS* models 77
3.4.3 Genotyping of edited *INS* hiPSC lines 78
3.4.4 Quality control of *INS* hiPSC models 79

3.5 Differentiation of SB *INS* p.C43G and NKX6.1-GFP *INS* p.C43G hiPSC models and associated control lines into pancreatic beta-like cells and cellular phenotyping at selected stages of development 82
3.5.1 Evaluation of developmental markers of beta-cell differentiation in *INS* p.C43G and WT clones 82
3.6 Analysing transcript levels of *INS* p.C43G hiPSC lines differentiated along the endocrine lineage 91
3.6.1 Separation of NKX6.1-GFP+ population using FACS 91
3.6.2 No differences in expression of insulin at stage 6 or 7 between groups 92
3.6.3 No difference in levels of ER stress markers between groups 94
3.7 Discussion 96
4 Characterisation of the p.C43G *INS* mutation in the EndoC-βH1 cell line 100
4.1 Introduction 101
4.2 Methods 104
4.2.1 Cloning of *INS* expression plasmids 104
4.2.1 Transient transfections 107
4.2.2 Quantitative real-time PCR (RT-qPCR) 107
4.2.3 Immunofluorescence staining 107
4.2.4 Confocal microscopy 108
4.2.5 Western Blotting 109
4.2.6 Generation of an *INS* KO EndoC-βH1 cell line 109
4.2.7 Transcriptome profiling 110
4.3 Results 112
4.3.1 Characterisation of the *INS*-KO EndoC-βH1 cell line 112
4.3.2 Establishing a re-expression system in the *INS*-KO cell line 115
4.3.3 Establishing a transfection system of *INS*-KO cells with a previously published plasmid 117
4.3.4 Transfection of hProCpepGFP shows incomplete processing of insulin 117
4.3.5 New cloning strategy to re-introduce the stop codon to the *INS* sequence and prevent expression of the Myc-FLAG tag in the pCVM6_*INS*_WT_Myc plasmid 118
4.3.6 Transfected cells demonstrated increased *INS* expression as measured by RT-qPCR but insulin protein was unable to be detected via Western Blot 118
4.3.7 Immunofluorescence reveals successful transfection but a low level of co-transfection 121
4.3.8 Wildtype and mutant *INS* are both localised to the secretory vesicles 122
4.3.9 Transfection of the wildtype or mutant construct does not induce ER stress when evaluated by RT-qPCR. 124
4.3.10 Principal component analysis revealed a separation of samples treated with tunicamycin 125

4.4 Discussion 130
5 Characterisation of HNF1A knockout beta cell models 133
5.1 Introduction 134
5.2 Methods 136
5.2.1 Genome editing 136
5.2.2 Differentiation experiments 137
5.2.3 Quantitative real-time PCR (RT-qPCR) 137
5.2.4 Western Blotting Analysis 138
5.2.5 Transcriptome profiling 139
5.2.6 RNA interference 140
5.3 Results 141
5.3.1 Generation of the *HNF1A* KO cell line 141
5.3.2 Characterisation of *HNF1A* KO hiPSC model 141
5.3.3 Differentiation of *HNF1A* KO hiPSC models into pancreatic beta-like cells and cellular phenotyping at three stages of development 143
5.3.4 Transcriptome profiling of *HNF1A* KO and WT hiPSC lines differentiated along the endocrine lineage 153
5.3.5 Analysis demonstrates a significant overlap between *HNF1A* KO DEGs and *HNF1A* targets identified through an independent *HNF1A* knock-down experiment in the EndoC-βH1 cell line 162
5.4 Discussion 165
6 Discussion 168
Bibliography 177

1

Introduction

1.1 Diabetes Mellitus

1.1.1 Diabetes is a worldwide health concern

Diabetes mellitus is a group of lifelong metabolic diseases with the common feature of hyperglycaemia due to failure in insulin production, insulin resistance or a decrease in insulin function (Alberti and Zimmet, 1998). Diabetes is recognised as a substantial economic and health burden, due to the high mortality and morbidity rate, and cases are expected to increase rapidly over the coming years (WHO 2016). Diabetes is classified into subtypes of monogenic, Type 1 Diabetes (T1D) and Type 2 Diabetes (T2D). The heterogeneous nature of the diseases means that misdiagnosis can be common and therefore it is important to correctly identify the type of diabetes to optimise treatment and minimise potential complications.

1.1.2 Monogenic diabetes

Monogenic diabetes is caused by single gene mutations in genes that control either the production or release of insulin (Hattersley, et al., 2006). There are two major classifications of monogenic diabetes: maturity-onset diabetes of the young (MODY) and neonatal diabetes mellitus (NDM) with both invariably presenting at a younger age than T2D, with NDM usually diagnosed within the first six months of life (Edghill, et al., 2006, Juszczak, et al., 2016). Monogenic diabetes is discussed in more detail later in the Chapter.

1.1.3 Type 1 Diabetes (T1D)

Type 1 Diabetes (T1D), comprising 10% of diabetes cases, is a chronic autoimmune disease whereby the immune cells of the body attack the pancreatic beta cells leading to insulin deficiency (Gepts, 1965, Nerup, et al., 1971). Specific T1D-associated autoantibodies, such as those that target insulinoma-associated protein 2 (IA-2), *SLC30A8*

and the 65-kD isoform of glutamic acid decarboxylase (GAD65) can act as biomarkers of T1D and be present before symptoms arise (Baekkeskov, et al., 1990, Ziegler, et al., 1999, Wenzlau, et al., 2007, Ilonen, et al., 2013). Although a polygenic disease influenced by environmental factors, the main genetic risk is associated with the human leukocyte antigen (HLA) region (Rich, et al., 2006, Rich, et al., 2009, Pociot, et al., 2010). Another strong genetic association is with *INS*, with two common variants in the non-transcribed variable number of tandem repeat (VNTR), approximately 0.5kb upstream of *INS*, having an association (Bell, et al., 1984). A smaller number of tandem repeats is associated with increased risk of T1D whereas longer repeats are protective (Bennett, et al., 1997, Pugliese, et al., 1997, Motzo, et al., 2004, Concannon, et al., 2009). Genetic studies have thus far revealed more than 150 non-HLA genetic factors that influence T1D risk (Cudworth and Woodrow, 1975, Todd, et al., 2007, Cooper, et al., 2012, Evangelou, et al., 2014, Robertson, et al., 2020). The use of a T1D genetic risk score (T1D-GRS) for distinguishing T1D from other forms of diabetes is discussed later in the Chapter.

1.1.4 Type 2 Diabetes (T2D) – a worldwide health concern

Currently, there are over 450 million individuals diagnosed with Type 2 Diabetes (T2D) worldwide, with the number set to continue increasing over the coming years (International Diabetes Federation, 2019). In the UK alone, it is expected that 5 million people will be diagnosed with T2D by 2025 (Diabetes-UK 2019). T2D is characterised by prolonged hyperglycaemia as a result of defective glucose-stimulated insulin secretion from the pancreatic beta cells and insulin resistance in the peripheral tissues (Defronzo, 1979, Kahn, et al., 1993, Weyer, et al., 1999). Those with T2D are often at an increased risk of secondary complications such as cardiovascular disease, nephropathy and lower limb amputation (American Diabetes Association, 2018; International Diabetes

Federation, 2019). Treatment of such secondary complications places an immense burden on the healthcare system, with 80% of the NHS's diabetes' budget being spent on treating complications (Diabetes-UK, 2019). As such, effective therapeutic targets need to be identified to reduce the chance of secondary complications and to improve quality of life for individuals with T2D.

1.1.5 There are currently no disease modifying therapies for Type 2 Diabetes

In the first instance, patients with T2D are placed on diet and exercise programmes, with the aim of both managing their T2D and reducing the risk of secondary complications (Davies, et al., 2018). These lifestyle changes have been shown to have considerable success, with a third of participants in one study still in remission for diabetes two years following treatment, with remission linked to sustained and substantial weight loss (Lean, et al., 2019). However, most patients will still require pharmacological intervention, albeit with the challenge that the available therapies for T2D are not curative and are often non-specific (Kahn, et al., 2006, Marin-Penalver, et al., 2016). In addition, patients may be on several medications alongside lifestyle alterations to try to reduce the chance of secondary complications. Due to the heterogeneous nature of the disease, treatment often has to be specifically tailored to the individual.

Treatment can take one of many approaches. Metformin is the first line treatment with a complex mechanism of action, which includes increasing insulin sensitivity, activating AMP-activated protein kinase (AMPK) inhibiting hepatic glucose production, and altering the gut microbiota (Zhou, et al., 2001, Madiraju, et al., 2014, Wu, et al., 2017, Gormsen, et al., 2019). Sulphonylureas act by closing ATP-sensitive potassium channels, causing voltage-gated Ca^{2+} channels to open and thus increasing insulin release from pancreatic beta cells (Ashcroft and Rorsman, 1989, Gribble and Ashcroft, 1999).

SGLT2 inhibitors, often known as gliflozins, and regularly used when metformin is ineffective, work by inhibiting sodium-glucose transport protein 2 (SGLT2), preventing glucose reabsorption from the proximal tubule of the kidney. Alongside the decrease in blood glucose levels, SGLT2 inhibitors have been found to have cardiovascular benefits (Zinman, et al., 2016, Hodson and Rorsman, 2020, Saponaro, et al., 2020). Another approach to treating diabetes has been to utilise dipeptidyl peptidase-4 (DPP4) inhibitors and glucagon-like peptide-1 receptor agonists (GLP1-RA) to enhance the incretin effect. These therapies work by inhibiting glucagon secretion, increasing glucose uptake in peripheral tissues and delaying gastric emptying (Drucker, et al., 1987, Wettergren, et al., 1993, Hvidberg, et al., 1994). Liraglutide, an example of a GLP1-RA drug, has been shown to diminish appetite and aid in weight loss in obese patients (Pi-Sunyer, et al., 2015). Another option is to use thiazolidinediones (TZDs), which lower blood glucose and alter lipid metabolism via interaction with the peroxisome proliferator-activated receptor gamma (PPARɣ) causing increased insulin sensitivity in peripheral adipocytes (Tontonoz, et al., 1995, Mayerson, et al., 2002). The number of differing treatments is indicative of the heterogeneous nature of the disease, with therapies serving to treat complications or conditions arising from the disease rather than combat the underlying pathophysiology. Utilising genetic approaches would enable the increased implementation of personalised therapies, alongside the ability to identify those who may be at increased risk of complications, improving patient outcomes and lowering healthcare costs.

1.1.6 Advantages of tackling diabetes using a genetic approach

Monogenic, T1D and T2D are unified in their pathogenesis of disease, centred on either dysfunction or destruction of the beta cell. Several different methods have been used in genetic approaches to different forms of diabetes. Identifying genes responsible for

monogenic forms of diabetes has traditionally been easier as the variants co-segregate in families and are of large-effect. Lessons learnt from monogenic forms of diabetes can be extrapolated to Type 2 Diabetes. Furthermore, many of the genes responsible for monogenic forms of diabetes also harbour common variants which increase the risk of developing Type 2 Diabetes (Fuchsberger, et al., 2016). This is due to the many shared loci between monogenic diabetes and T2D, particularly in regard to genes involved in pancreatic development and function, in addition to insulin secretion and ER stress (Pasquali, et al., 2014, Thurner, et al., 2018). For instance, mutations in genes such as *KCNJ11* and *ABCC8*, which have an essential function in the regulation of insulin secretion, cause NDM but mutations have also been shown to form a high-risk haplotype which can predict the progression from impaired glucose tolerance to T2D (Gloyn, et al., 2003, Gloyn, et al., 2004, Babenko, et al., 2006). Additionally monogenic diabetes loci such as *HNF1A*, *GLIS3*, *GCK*, and *PAX4* also harbour variants associated with T2D risk (Voight, et al., 2010, Cho, et al., 2011, Rees, et al., 2011, Fu, et al., 2013). Therefore, it can be seen that exploring different types of monogenic diabetes can provide insight into beta cell physiology and a better understanding of insulin secretion and its regulation.

Although environmental factors such as diet, smoking and physical activity are known to influence diabetes risk, familial studies have shown that a proportion of T2D risk is heritable (Almgren, et al., 2011, Hu, 2011, Willemsen, et al., 2015). For instance, dizygotic twins demonstrate a concordance rate of 20-30%, which increases to 70% in monozygotic twins (Kaprio, et al., 1992). An individual's risk of T2D is 40% if one parent has T2D and almost 70% if both parents have the disease (Tillil and Kobberling, 1987).

Genome-wide association studies (GWAS) allow the screening of large numbers of individuals in an unbiased approach to identify not only common variants in T2D, but

rare and low-frequency variants with large effect sizes, with meta-analyses increasing the robustness and reliability of the findings. Since the earliest genome-wide association study, there have been over 400 genetic loci identified that are implicated in T2D risk, with many variants residing in non-coding regions (Mahajan, et al., 2018, Spracklen, et al., 2020, Vujkovic, et al., 2020). GWAS have greatly improved the ability to discover candidate genes that are particularly important in beta cell physiology and pathophysiology (Billings and Florez, 2010, Voight, et al., 2010, Dimas, et al., 2014, Fuchsberger, et al., 2016, Udler, et al., 2018).

Utilising genetics also provides the opportunity to discriminate different forms of diabetes from one another, enabling an earlier diagnosis for other family members through screening, enabling optimisation of treatment and informing of potential heritability of their disease. A T1D genetic risk score (T1D-GRS), comprising T1D-associated common genetic variants, has been shown to differentiate monogenic diabetes from T1D (Patel, et al., 2016). Furthermore, the T1D-GRS can help distinguish T1D from T2D in young adults (Oram, et al., 2016). Alongside this, advancements in sequencing methods have allowed the opportunity for patients with neonatal diabetes to receive genetic testing encompassing a known panel of neonatal diabetes genes, rather than selecting a much smaller number of genes to test based on clinical information. This broad testing approach detected the causal gene in 80% of neonatal diabetes patients. Such a diagnosis allows the early intervention of correct therapeutic approaches, alongside being informed on the long-term management and progression of the particular subtype of neonatal diabetes (De Franco, et al., 2015).

These approaches demonstrate the advantage of tackling diabetes using genetics. Firstly, through translation of findings of the large effect sizes of genes involved in monogenic diabetes to more complex forms of diabetes, and additionally utilising large populations

in an unbiased way to find T2D-associated genetic loci with much smaller effect sizes. Further, advancements in sequencing methods allows the discrimination of different forms of diabetes. These methods not only show that investigating genetic associations with diabetes can uncover novel mechanisms and insights into pathogenesis of the disease but that the beta cell should be a central focus point for identifying further therapeutic targets.

1.2 The Pancreatic Beta Cell

The endocrine islets of Langerhans, found throughout the pancreas, only make up about 1-2% of the pancreatic mass yet receive up to 15% of the entire pancreatic blood flow (Ballian and Brunicardi, 2007). The pancreatic islets are comprised of insulin-producing beta cells (60%), glucagon-producing alpha cells (30%), with the remaining 10% incorporating somatostatin-producing delta cells, ghrelin-producing epsilon cells and pancreatic polypeptide-producing gamma cells, although these ratios can vary between individuals (Orci, et al., 1976, Brissova, et al., 2005, Andralojc, et al., 2009, Hauge-Evans, et al., 2009). Each cell type has a particular function and is involved in the biosynthesis and secretion of their specific hormones. Together, the cells work to maintain glucose homeostasis. The beta cell is of particular importance, as it is the only cell which can effectively reduce blood glucose levels via the release of insulin in response to hyperglycaemia, with dysregulation of insulin at the centre of diabetes (Coore and Randle, 1964, Malaisse, et al., 1983, Henquin, et al., 1994).

1.2.1 Insulin and its role in glycaemic control

Insulin plays the central role in the maintenance of normal blood glucose within the body (Bell, et al., 1979, Henquin, 1980). Exclusively expressed and secreted by the beta cells in the islets of Langerhans, insulin is vital for both stimulating glucose uptake from the

systemic circulation and suppression of gluconeogenesis within the liver (Hedeskov, 1980, Gerich, 1993, Henquin, et al., 1994). Here, and in muscle tissue, insulin binds to tyrosine kinase receptors, stimulating phosphorylation of specific tyrosine residues which leads to activation of intracellular messengers such as PI3K and Akt (Wilson and Cushman, 1994, Hirashima, et al., 2003). Activation of these intracellular messengers leads to glycogenesis and the conversion of glucose to pyruvate, eventually causing an increase in ATP production (Oakes, et al., 1997, Friedrichsen, et al., 2013). Within the muscle, insulin also stimulates uptake of amino acids through relevant channels, leading to an increase in protein synthesis (Bonadonna, et al., 1993). Insulin enables uptake of glucose in skeletal muscle and adipose tissues via trafficking of GLUT4 transporters from the intracellular compartments to the cellular surface (James, et al., 1987, Vilaro, et al., 1989). Insulin also promotes glucose uptake within adipose and liver tissue, alongside increasing lipoprotein lipase activity within adipose tissue (Sadur and Eckel, 1982, Sambandam, et al., 1999, Kersten, 2001). Excess stimulation of lipogenesis leads to decreased insulin action, eventual insulin resistance and progression to diabetes (Lefebvre, et al., 1998, Abel, et al., 2001, Yore, et al., 2014). Within this setting, increased insulin levels promote accelerated acetyl CoA carboxylase production whilst preventing the suppression of hormone sensitive lipase in adipose tissue, resulting in higher levels of circulating fatty acid levels (Witters, et al., 1988, Fex and Mulder, 2008). Excess fatty acids increases insulin resistance and can cause hypertension, increasing the risk of cardiovascular disease (Reaven, et al., 1988, Santomauro, et al., 1999, Inoguchi, et al., 2000). Insulin gene expression is increased in response to increased blood glucose levels and is regulated by several islet-enriched transcription factors such as *PDX1*, *NEUROD1* and *MAFA* (Permutt and Kipnis, 1972, Goodison, et al., 1992, Naya, et al., 1995, Ohneda, et al., 2000, Aramata, et al., 2007). The central role of insulin in metabolic

processes is evident by the complexity of diabetes-associated traits that can occur as a result of insulin dysregulation.

1.2.2 Glucose stimulated insulin secretion

Blood glucose levels are strictly monitored, with normal glucose range between 4-6mM. When blood glucose increases, glucose is taken up by the beta cells via the glucose transporters GLUT1 and GLUT3 (McCulloch, et al., 2011). Glucose is converted into ATP via glycolysis, which results in an increased intracellular ratio of ATP:ADP leading to closure of the K_{ATP} channel (Ashcroft, et al., 1984, Ashcroft and Rorsman, 1989). In turn, this causes depolarisation of the membrane and the opening of voltage-dependent Ca^{2+} channels (VDCCs) which leads to calcium entering the cell causing an increase in cytoplasmic free Ca^{2+} (Rorsman, et al., 2012). This increase in calcium triggers exocytosis of the insulin-containing vesicles, culminating with the release of the hormone into the extracellular space (Daniel, et al., 1999, Henquin, 2000, Rorsman and Ashcroft, 2018) (Figure 1.2.1).

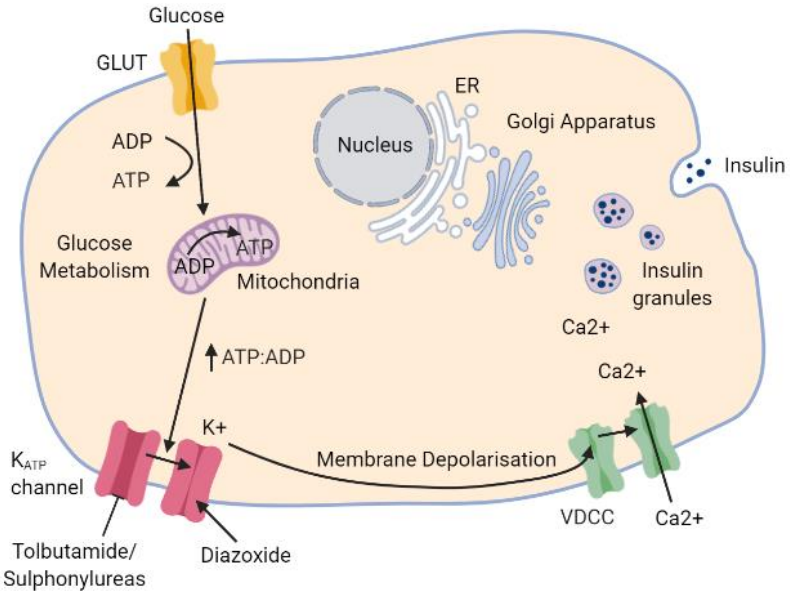


Figure 1.2.1 Glucose-stimulated insulin secretion in the pancreatic beta cell
Glucose uptake is mediate via glucose transporters (GLUT), whereby glucose is broken down via aerobic glycolysis in the cytoplasm and the Krebs cycle in the mitochondria, generating ATP. The increase in the ATP:ADP ratio inhibits K_{ATP} channel activity, inducing membrane depolarisation and opening of voltage-dependent calcium channels (VDCC). Calcium influx within the cell results in the exocytosis of secretory vesicles containing insulin and subsequent release of insulin into the extracellular space. The figure was generated using icons from Biorender.

1.2.3 Insulin and its role as a growth hormone

Beyond its role maintaining blood glucose levels, insulin has both direct and indirect functions in the control of normal body growth. Insulin is one of the major anabolic hormones in the body, increasing the synthesis of DNA, RNA and proteins in target tissues (Manchester, 1972). In the foetus, insulin is present in the pancreas as early as the 10^{th} gestational week and is required to process maternal glucose, enabling tissue growth and fat deposition (Grasso, et al., 1980, Milner and Hill, 1984). There is a direct relationship in normal pregnancies between maternal blood glucose levels in the third trimester and the birthweight of the child (Farmer, et al., 1988, Breschi, et al., 1993, Sermer, et al., 1995). The impact of insulin on growth can be seen in infants of mothers with diabetes whereby infants have an increased birth weight (He, et al., 2015). In contrast, babies with agenesis of the pancreas, can be emaciated and short in stature, and

infants with neonatal diabetes can have slower growth (Allen, et al., 2011, Naylor, et al., 2011). This evidence suggests that it is not only the maternal glycaemic level that affects foetal growth but that foetal genetics also influence birthweight, a theory known as the foetal insulin hypothesis (Hattersley and Tooke, 1999). This concept is demonstrated by heterozygous mutations in the glucokinase gene (*GCK*), where inheritance of a *GCK* mutation led to decreased birth weight, whereas maternal hyperglycaemia due to a *GCK* mutation led to increased birth weight (Hattersley, et al., 1998).

1.2.4 The biosynthesis of insulin

Preproinsulin is directed to the ER via its signal peptide

The insulin gene (*INS*) produces the precursor of insulin, preproinsulin, a polypeptide formed of an N-terminal signal sequence, with a 30 amino acid B and a 21 amino acid A chain connected by C-peptide. The N-terminal signal sequence directs preproinsulin towards the ER where the signal sequence is subsequently cleaved upon entry into the ER (Steiner, et al., 1967). The mechanism of insulin biosynthesis is summarised in the Figure 1.2.2.

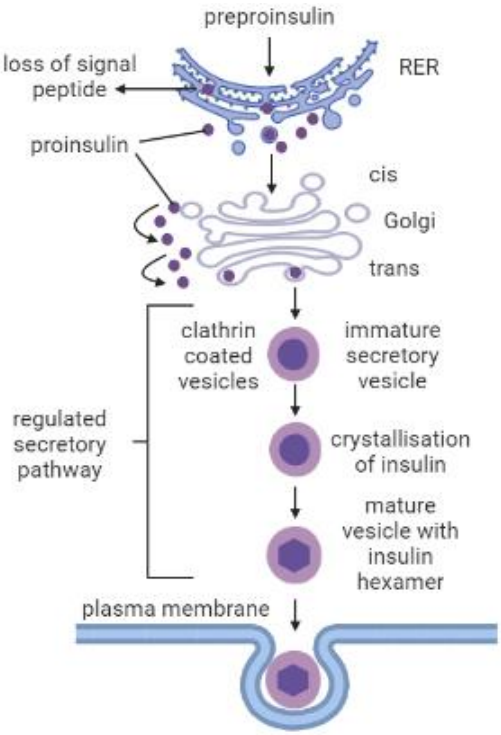


Figure 1.2.2 Mechanism of insulin biosynthesis
Preproinsulin is directed to the ER via its signal peptide which is cleaved upon entry to the rough endoplasmic reticulum (RER). The resultant proinsulin is folded in the RER into its native conformation by formation of three highly conserved disulphide bonds. Correctly folded proinsulin is transported to the Golgi apparatus where it is packaged into clathrin-coated immature secretory vesicles. A process of maturation occurs within the vesicles, where proinsulin is cleaved by prohormone convertases PC1/3 and 2 into mature insulin and C-peptide. Following the liberation of C-peptide, an insulin hexamer forms, stabilised by calcium and zinc ions. The resultant crystallisation of insulin produces an insoluble hexamer, termed large dense core vesicle due to its appearance as observed using an electron microscope. Figure created using icons from Biorender.

Folding of proinsulin in the ER

The resultant proinsulin is folded in the rough ER into its native conformation by formation of three highly conserved disulphide bonds. The first bond is formed between the B chain 19th residue and the A chain 20th residue, followed by B7-A7 and A6-A11 (Chang, et al., 2003, Haataja, et al., 2016). The oxidative environment catalyses the bond formation process alongside various chaperone proteins and folding enzymes, including protein disulphide isomerase (Munro and Pelham, 1987, Winter, et al., 2002). The resultant folded structure has hydrophobic interfaces which interact to form homodimers

(Dodson and Steiner, 1998). However, newly synthesised proinsulin may fail to form its native conformation. To prevent over accumulation of misfolded protein in the ER there is a quality control system in place, the unfolded protein response (UPR), which is activated when the load of unfolded proteins exceeds capacity (Goodge and Hutton, 2000, Credle, et al., 2005).

Successfully folded proinsulin is transported to the Golgi apparatus

The subsequent exit of correctly folded proinsulin homodimer from the ER is dependent on coat protein complex II (COPII) coated vesicles which are transported to the Golgi apparatus. The Golgi, containing zinc and calcium, has a favourable environment for the proinsulin homodimers to join with two others to form a hexamer (Emdin, et al., 1980). Proinsulin is sorted into the clathrin-coated vesicles of the regulated secretory pathway, although the mechanism by which this occurs has yet to be fully determined (Steiner and James, 1992, Fang, et al., 2015). The immature secretory vesicles are packaged with proinsulin alongside markers of the secretory vesicles such as contain chromogranin-A and -B (Hutton, et al., 1987, Kim, et al., 2001, Huh, et al., 2003).

Prohormone convertases cleave C-peptide from the proinsulin hexamer

On the vesicular membrane, there are V-type H+-ATPase pumps enabling protons to enter the lumen, activating the pH-sensitive and calcium-dependent enzymes, the prohormone convertases (Orci, et al., 1987, Sun-Wada, et al., 2006, Deriy, et al., 2009). The prohormone convertases cleave the C-peptide from the proinsulin molecule. Until recently, the general consensus was that the action of prohormone convertase 1/3 (PC1/3), encoded for by *PCSK1*, cleaves the B-chain– C-peptide junction at Arg31-Arg32, with the resultant dibasic residues Arg31/32 removed by carboxypeptidase E (Kemmler and Steiner, 1970). Following this, prohormone convertase 2 (PC2), encoded

for by *PCSK2*, cleaves between the A-chain and C-peptide at Lys64-Arg65 (Davidson and Hutton, 1987, Davidson, et al., 1987, Kaufmann, et al., 1995). Cleavage is followed by exopeptidase removal of the C-terminal Lys and Arg exposed residues, carried out by carboxypeptidase E or H (Figure 1.2.3) (Guest, et al., 1989, Lipkind and Steiner, 1999). Whilst mouse studies have shown that knockout of either *PCSK1* or *PCSK2* leads to accumulation of proinsulin intermediates, humans with *PSCK2* mutations do not exhibit hyperproinsulinaemia and those with mutant *PCSK1* have circulating hyperproinsulinaemia (Furuta, et al., 1998, Zhu, et al., 2002, Zheng, et al., 2012, Stijnen, et al., 2016). In addition, transcriptome analysis of human primary beta cells has shown that human beta cells express significantly more *PCSK1* compared to *PCSK2* than mice, indicating that PC1/3 may be the preferred major maturation pathway in human secretory vesicles (Zheng, et al., 2012, Bramswig, et al., 2013, Benner, et al., 2014, Stijnen, et al., 2016, Ramzy, et al., 2020). However, there is still debate within the field about the exact role of the two genes. Previous studies have demonstrated a role for *PCSK2* variants with increased T2D risk in certain populations, suggesting that it is required for proinsulin processing (Jonsson, et al., 2012, Chang, et al., 2015). It may be that *PCSK2* provides a compensatory mechanism in situations where there is excess pressure on proinsulin processing.

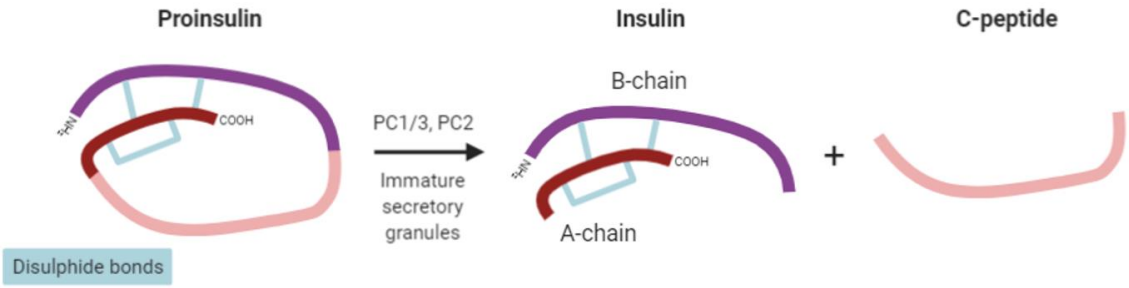


Figure 1.2.3 Maturation of proinsulin into insulin and C-peptide
The insulin precursor, proinsulin (*left*) is folded in the ER with stabilisation via three disulphide bonds. The proinsulin then undergoes further maturation with cleavage of C-peptide by consecutive action of PC1/3 and PC2 to form insulin. Figure created using icons from Biorender

Crystallisation of the insulin hexamer

Following the cleavage of C-peptide, each insulin hexamer is stabilised by zinc ions, which are imported into the lumen via pancreatic-specific ZnT-8 channels, encoded for by *SLC30A8*, which has been recently characterised for its association with T2D risk (Dodson and Steiner, 1998, Dwivedi, et al., 2019). The resultant insulin hexamer is far less soluble than the proinsulin hexamer, which leads to crystallisation of insulin (Dodson and Steiner, 1998). The appearance of insulin crystals within the vesicles using electron microscopy resulted in them being referred to as dense-core vesicles – observed as a dark structure with clear halo surround (Hill, et al., 1991, Dodson and Steiner, 1998, Chimienti, et al., 2004). Upon release into the blood stream, the change in pH and decrease in zinc concentration triggers the insulin hexamers to separate and become monomers (Carpenter and Wilcox, 2014). The insulin monomers are then able to bind to the insulin receptor. Insulin is secreted in equivalent molar amounts to C-peptide, allowing C-peptide to be used as a surrogate measurement to accurately assess insulin production from the beta cell (Jones and Hattersley, 2013). Insulin itself can be metabolised by the liver whereas C-peptide is not, therefore measuring C-peptide more accurately represents an individual's insulin secretion (Shapiro, et al., 1988).

Excess pressure on the beta cell machinery caused by proinsulin misfolding can lead to ER stress and apoptosis

A sustained stimulation of insulin biosynthesis can cause increased pressure on the UPR, which translates to an apparent increase in secretory pathway activity. Overwhelming the UPR can initiate ER stress, which may lead to additional proinsulin misfolding, compounding the problem and causing beta cell apoptosis and resultant diabetes (Butler, et al., 2003, Hodish, et al., 2010, Cnop, et al., 2012). Therefore, any defect in insulin

processing will greatly impact on the ability of the beta cell to respond to changes in blood glucose and result in a vicious cycle of increased insulin production and ER stress. Not only could the misfolded proinsulin prevent correctly folded proinsulin exiting the ER but an increased proportion of released proinsulin could worsen the quality of glycaemic control, as proinsulin is around 10x less efficient as a glycaemic agent compared with insulin (Yu and Kitbachi, 1973, Liu, et al., 2010).

1.3 Monogenic diabetes

Monogenic diabetes represents around 1-3% of all diabetes cases, resulting from mutations in a single gene (Hattersley, et al., 2006, Shields, et al., 2010, Pihoker, et al., 2013, Shepherd, et al., 2016). The majority of monogenic diabetes can be classified as neonatal diabetes mellitus (NDM) or maturity-onset diabetes of the young (MODY), with both invariably presenting at a younger age than T2D (Juszczak, et al., 2016, Misra and Owen, 2018).

1.3.1 Neonatal diabetes mellitus (NDM)

Neonatal diabetes (NDM) is a rare form of diabetes usually diagnosed within the first six months of life and comprises both transient and permanent neonatal diabetes (PNDM) (Flanagan, et al., 2007). In transient neonatal diabetes, individuals are diagnosed soon after birth and go into apparent remission by 18 months but typically develop T2D in later life (Temple, et al., 2000). Permanent neonatal diabetes generally presents within the first six months of life, sometimes within the first year, and then persists throughout the lifespan (Shield, 2000, Rubio-Cabezas, et al., 2012). The most frequently associated mutations with NDM are genes involved in beta cell development and function (Rubio-Cabezas and Ellard, 2013). There are currently 30 genetic causes of NDM identified

accounting for approximately 80% of cases (De Franco, et al., 2015, Johnson, et al., 2017, Johnson, et al., 2019, De Franco, et al., 2020).

Autosomal dominant mutations within two genes which account for up to 60% of NDM cases are heterozygous activating mutations in *ABCC8* and *KCNJ11* (Flanagan, et al., 2006, Greeley, et al., 2011). Both genes encode subunits of the pancreatic beta cell K_{ATP} channel, with mutations causing the potassium channels to permanently move to an open configuration, preventing depolarisation of the cell membrane and thus preventing insulin secretion (Gloyn, et al., 2004, Greeley, et al., 2018). Oral sulphonylureas can be administered to patients, targeting the K_{ATP} channel enabling stimulation of insulin secretion (Pearson, et al., 2006, Rafiq, et al., 2008). Oral sulphonylureas have been shown to have better long-term efficacy and safety than insulin treatment, with patients therefore having an improved quality of life (Pearson, et al., 2003, Bowman, et al., 2018). The second most common cause of neonatal diabetes are heterozygous mutations in *INS* (Stoy, et al., 2007, Edghill, et al., 2008). *INS* as a cause of neonatal diabetes is discussed in more detail later in this Chapter.

Two of the many genes that cause neonatal diabetes are mutations in transcription factors *PDX1* and *GATA6*. Mutations in either one of these genes can lead to a spectrum of phenotypes but can cause severe defects in pancreatic development, resulting in pancreatic agenesis (Allen, et al., 2011, De Franco, et al., 2013, De Franco, et al., 2013).

Autosomal recessive mutations in *EIF2AK3*, which encodes protein kinase R-like endoplasmic reticulum kinase (PERK), can cause NDM as part of Wolcott-Rallison syndrome (Wolcott and Rallison, 1972, Delepine, et al., 2000). These patients commonly have skeletal abnormalities and liver dysfunction, and is the most frequent cause of PNDM in consanguineous families, followed by mutations in *INS* (Rubio-Cabezas, et al.,

2009). Lack of PERK activity, which is involved in mediating the unfolded protein response (UPR) in beta cells, triggers beta cell apoptosis (Harding, et al., 2000, Harding, et al., 2001). Recessive loss-of-function mutations cause Wolfram syndrome, often referred to as DIDMOAD after its four most common features (Diabetes Insipidus, Diabetes Mellitus, Optic Atrophy and Deafness) (Gunn, et al., 1976, Barrett, et al., 1995). Despite early-onset of diabetes being a fundamental feature of the disease, NDM is only associated with rare autosomal dominant mutations (De Franco, et al., 2017). Two causative genes for Wolfram syndrome (WS) have been identified – *WFS1* and *WFS2* (Inoue, et al., 1998, Amr, et al., 2007). The typical form of WS is caused by *WFS1* mutations, encoding wolframin, which in the disease state causes hyperactivation of *ATF6* and therefore continual activation of genes that promote cellular apoptosis such as *DDIT3* and *BIP*. *WFS1* expression has also been demonstrated to be induced during insulin secretion, as well as being a component of IRE1 and PERK signalling, suggesting that *WFS1* protects beta cells against ER stress (Fonseca, et al., 2005, Yamada, et al., 2006, Fonseca, et al., 2010, De Franco, et al., 2017).

1.3.2 Maturity onset of diabetes in the young (MODY)

MODY accounts for 1-2% of all diabetes cases with up to 80% being initially misdiagnosed as T1D (Juszczak, et al., 2016). Mutations in *HNF1A*, *HNF1B*, *HNF4A* and *GCK* make up the vast majority of cases in the UK and usually presents in adults under 25 years of age (Kavvoura and Owen, 2012). Correct diagnosis allows many patients to alter their treatment, as well as being more informed about the progression of their disease and whether to seek testing for other family members (Thanabalasingham, et al., 2012). The current list of genetic causes of MODY are detailed in Table 1.3.1.

Table 1.3.1 Genetic mutations leading to MODY Genes are ordered in their MODY subtype classification. Incidence is the percentage of MODY cases that the gene causes.

Gene	Incidence (%)	Reference(s)
HNF4A	5-10	(Yamagata, et al., 1996)
GCK	30-50	(Froguel, et al., 1992, Hattersley, et al., 1992, Vionnet, et al., 1992, Pearson, et al., 2001)
HNF1A	30-60	(Yamagata, et al., 1996, Frayling, et al., 1997)
PDX1	<1	(Stoffers, et al., 1997, Fajans, et al., 2010)
HNF1B	<5	(Horikawa, et al., 1997, Bellanne-Chantelot, et al., 2004, Faguer, et al., 2011)
NEUROD1	<1	(Malecki, et al., 1999, Horikawa, et al., 2018)
CEL	<1	(Raeder, et al., 2006, Raeder, et al., 2014)
PAX4	<1	(Mauvais-Jarvis, et al., 2004, Plengvidhya, et al., 2007)
INS	<1	(Edghill, et al., 2008, Molven, et al., 2008, Meur, et al., 2010)
ABCC8	<1	(Bowman, et al., 2012)
KCNJ11	<1	(Bonnefond, et al., 2012, Liu, et al., 2013)

Rare forms of MODY include those caused by genes related to beta cell development such as *PDX1* and *NEUROD1* (Stoffers, et al., 1997, Malecki, et al., 1999). Heterozygous mutations in *PDX1* may cause a mild form of diabetes with diminished insulin secretion (Stoffers, et al., 1997, Clocquet, et al., 2000). However, homozygous or compound heterozygous mutations often lead to pancreatic agenesis and a more severe phenotype with a diagnosis of NDM (Stoffers, et al., 1997, Schwitzgebel, et al., 2003, Thomas, et al., 2009). Similarly, patients with heterozygous mutations in *NEUROD1* often present with mild diabetes with some patients even remaining glucose tolerant (Malecki, et al., 1999, Horikawa, et al., 2018). Interestingly, the incidence of *NEUROD1* mutations is almost double in females compared to males, with differences in clinical presentation also varying between ethnic groups (Kristinsson, et al., 2001, Gonsorcikova, et al., 2008,

Szopa, et al., 2016). In addition to being diagnosed much earlier than heterozygotes, some patients with homozygous *NEUROD1* mutations have also been diagnosed with neurological abnormalities and visual impairments (Rubio-Cabezas, et al., 2010, Demirbilek, et al., 2018). In recent years, genome-wide association studies (GWAS) have identified several genes causing monogenic diabetes, such as *HNF1A*, as harbouring common or low frequency variants which are associated with altered risk for T2D (Voight, et al., 2010, Fuchsberger, et al., 2016). Given that several monogenic genes are also associated with T2D risk, investigating the mechanisms by which beta cell dysfunction occurs in NDM and MODY also presents the opportunity to gain additional insight into T2D pathogenesis.

1.4 Islet and beta cell models

Due to the high proportion of T2D GWAS loci that are associated with insulin secretion pathways, the islet has become a primary focus for understanding the mechanisms at play in T2D (Voight, et al., 2010, Gaulton, et al., 2015, Fuchsberger, et al., 2016, Thurner, et al., 2018). However, it has been difficult to translate the findings from T2D GWAS, due to the shortage of human islet models available for study. In this regard, alternative models, such as rodent islets, have been utilised to investigate the development and function of both healthy and dysfunctional beta cells.

1.4.1 Differences between rodent and human islets

Rodents have been used extensively in the study of pancreatic development, alongside modelling of both monogenic diabetes and T2D (Sellick, et al., 2004, Watanabe, et al., 2009, Wijesekara, et al., 2010). However, there are differences in both architecture and physiology of the islet between mouse and man. Mice and humans display different sensitivity to glucose levels driven by K_{ATP} channel activity (Henquin, et al., 2006,

Henquin, et al., 2006, Skelin Klemen, et al., 2017). Mice have a higher percentage of beta cells compared to humans and these are largely found in the inner core of the islet, surrounded by alpha cells (Cabrera, et al., 2006, Dolensek, et al., 2015). Mouse islet formation occurs directly preceding birth, in contrast to the islet-like clusters seen as early as 11-15 weeks in humans (Sarkar, et al., 2008). Furthermore, there are species-specific differences in the expression and function of transcription factors during the developmental process. Mice have been shown to have biphasic *Neurog3* expression compared to a single-phase in humans, in addition to differences in the timings of when distinct endocrine subpopulations appear (Piper, et al., 2004, Johansson, et al., 2007, Jennings, et al., 2013). There is also a dissimilarity in the specificity of expression of genes between species. For instance, *UCN3* expression is limited to beta cells within mice but found in all cell types bar delta cells within human islets (Blum, et al., 2012, Baron, et al., 2016). The two species also demonstrate functional dissimilarities in glucose stimulated insulin secretion. Mice primarily express *Glut2* (*Slc2a2*) via which they facilitate glucose uptake, whereas in humans glucose enters the beta cells via both GLUT1 and GLUT3 (*SLC2A1* and *SLC2A3*) (De Vos, et al., 1995, Heimberg, et al., 1995, McCulloch, et al., 2011). Furthermore, the electrophysiological properties of the beta cells are different, with variations in the predominant type of ion channels, such as with voltage-gated calcium and sodium channels (Schulla, et al., 2003, Braun, et al., 2008, Nica, et al., 2013, Zhang, et al., 2014). Mice also express two insulin genes, *Ins1* and *Ins2*, in contrast to the one (*INS*) by humans (Wentworth, et al., 1986).

1.4.2 Animal cell models

Although there are several differences between rodents and human islets, they have still been an invaluable resource in modelling both monogenic and T2D (Wang, et al., 1999, Sellick, et al., 2004, van Burck, et al., 2010, Flannick, et al., 2014). However, there are

many instances where the animal model may not recapitulate the human phenotype, such as heterozygous mutations within *HNF1A* and *HNF4A*, which leads to MODY in humans but within mice a homozygous knockout is required to produce a diabetic phenotype (Pontoglio, et al., 1998, Miura, et al., 2006). Mouse models have also failed to recapitulate the T2D protective effect of rare loss-of-function mutations in *SLC30A8*, with *Slc30a8* KO mouse models being inconclusive with either no impact on glucose homeostasis or impaired glucose tolerance (Pound, et al., 2009, Pound, et al., 2012, Flannick, et al., 2014, Flannick, et al., 2019). Several of the dissimilarities between the species may be due to the previously discussed differences between islet architecture and function, in addition to the contrasting physiologies of the two species outside of the islet.

1.4.3 Human islets

Primary human islets, considered the gold standard for diabetes disease modelling, have been a valuable resource for advancing our understanding of the gene expression of beta cells, whilst uncovering the similarities and differences between human and rodent islets (Benner, et al., 2014, Hart and Powers, 2019). Substantial progress has been made in our understanding of the beta cell transcriptome and chromatin state (Gaulton, et al., 2010, Parker, et al., 2013, Segerstolpe, et al., 2016, Varshney, et al., 2017, Thurner, et al., 2018). However, there are several obstacles that prevent their routine use in research. Islet isolation is a costly and logistically challenging procedure with a large variability in the quality of human islet preparation. Up to a quarter of islet preparations may have impaired insulin secretion in either a static assay or in a dynamic perifusion system (Kayton, et al., 2015). Factors such as BMI, pancreas cold ischaemia time, age and HbA1c have all been shown to affect *in vitro* insulin secretion, meaning that there can be a large degree of heterogeneity between islets (Kayton, et al., 2015, Lyon, et al., 2016, Henquin, 2018). Islets cannot currently be genetically manipulated and are not routinely

accessed for biopsy, preventing their use for interrogative research on a regular basis. Due to these limitations, a more readily available resource is required in order to carry out disease modelling. The pros and cons of the commonly used human islet and beta cells models are outlined in Table 1.4.1.

Table 1.4.1 Human pancreatic islet and beta cell models

Model	Pros	Cons
Primary islets[1]	Comprise all endocrine cell types	Limited availability Heterogeneous genotypes Large variation in quality/viability Difficult to maintain cell functionality *in vitro* Genetic manipulation limited
Human beta cell lines[2]	Expandable Glucose responsive Similar transcriptional markers and functional properties to human primary beta cells	Immortalised/proliferative Aneuploid Low clonal efficiency Genetic manipulation limited Batch to batch variations Reduced functional resolution
Human pluripotent stem cell-derived beta-like cells[3]	Can differentiate into any cell type Easy to genetically manipulate	Variable differentiation efficiency Functionally immature Polyhormonal cells

1. (Kayton, et al., 2015, Lyon, et al., 2016, Henquin, 2018, Hart and Powers, 2019)
2. (Ravassard, et al., 2011, Hastoy, et al., 2018, Tsonkova, et al., 2018, Lawlor, et al., 2019) 3. (Bruin, et al., 2014, Hrvatin, et al., 2014, Pagliuca, et al., 2014, Rezania, et al., 2014, Grobarczyk, et al., 2015, Kyttala, et al., 2016)

1.4.4 EndoC-βH1

The immortalised human cell line, EndoC-βH1, was generated by targeted oncogenesis of the foetal pancreas, followed by lentiviral transduction of vectors expressing SV40LT, under the control of the rat insulin promoter. Foetal pancreas was then transplanted under the kidney capsule of recipient immune-incompetent SCID mice and expanded. The resultant insulinomas were subsequently transduced with human telomerase reverse

transcriptase (hTERT), again under the control of the rat insulin promoter, and transplanted into SCID mice and then harvested to expand *in vitro*, generating the EndoC-βH1 line (Ravassard, et al., 2011). The line possesses many comparable properties to islets, such as glucose-stimulated insulin secretion and expression of beta cell markers, although multiomic analyses revealed that these cells retain foetal progenitor islet cell signatures (Ravassard, et al., 2011, Hastoy, et al., 2018, Lawlor, et al., 2019). Despite these similarities, EndoC-βH1 are unable to be expanded from single clones, making genetic manipulation a challenge. The cell line can however, undergo siRNA-mediated KD and pooled CRISPR/Cas9-mediated KO to enable the study of diabetes-relevant genes, although a heterogeneous population of edited and unedited cells will be created (Thomsen, et al., 2016, Cardenas-Diaz, et al., 2019, Grotz, et al., 2019). EndoC-βH1 have also been effectively utilised to investigate the impact of individual T2D risk genes, such as *PAM* and *SLC30A8* (Thomsen, et al., 2018, Dwivedi, et al., 2019). However, to investigate patient-derived mutations, or to perform genetic manipulation at a single-cell level, an alternate cell model is required.

1.4.5 Human pluripotent stem cells

The recent advent of human pluripotent stem cells, encompassing embryonic and human induced pluripotent stem cells, has provided a realistic alternative to the EndoC-βH1 cell line to model diabetes. Human induced pluripotent stem cells, derived from differentiated adult somatic cells, are often utilised as they avoid the legal and ethical concerns that embryonic stem cells can entail (Takahashi, et al., 2007). Human iPSCs are typified by both their capacity to propagate indefinitely and their ability to differentiate into any cell type of the human body (Evans and Kaufman, 1981, Martin, 1981). They also exhibit a normal human diploid karyotype of 46 chromosome per cell, whereas other commonly used beta cell lines can have karyotypic abnormalities (Evans and Kaufman, 1981,

Jonnakuty and Gragnoli, 2007, Lawlor, et al., 2019). One of the main advantages over EndoC-βH1 cells is the ability to introduce or correct mutations into iPSCs on an isogenic background. Human iPSCs have already been used in several studies to investigate diabetes-associated mutations and uncover novel mechanisms of action (Balboa, et al., 2018, Cardenas-Diaz, et al., 2019). In contrast to EndoC-βH1 experiments, genes can be examined through each distinct stage of pancreatic development, with transplantation into immunocompromised mice facilitating substantial further maturation, demonstrated by human C-peptide secretion in response to glucose stimulation (Rezania, et al., 2014, Yoshihara, et al., 2016).

1.4.5.1 Human iPSCs can be differentiated to beta-like cells

The advancements *in vitro* differentiation protocols has enabled the generation of beta-like cells from hiPSCs. From the initial protocol, several modifications and improvements have been made by numerous different groups (Pagliuca, et al., 2014, Rezania, et al., 2014, Nostro, et al., 2015). Current *in vitro* differentiation protocols comprise seven distinct stages of endocrine pancreas development including definitive endoderm (DE), primitive gut tube (PGT), posterior foregut (PFG), pancreatic endoderm (PE), endocrine precursors (EP), endocrine-like (EN) and beta-like cells (BLC). An overview of the differentiation process is presented in Figure 1.4.1.

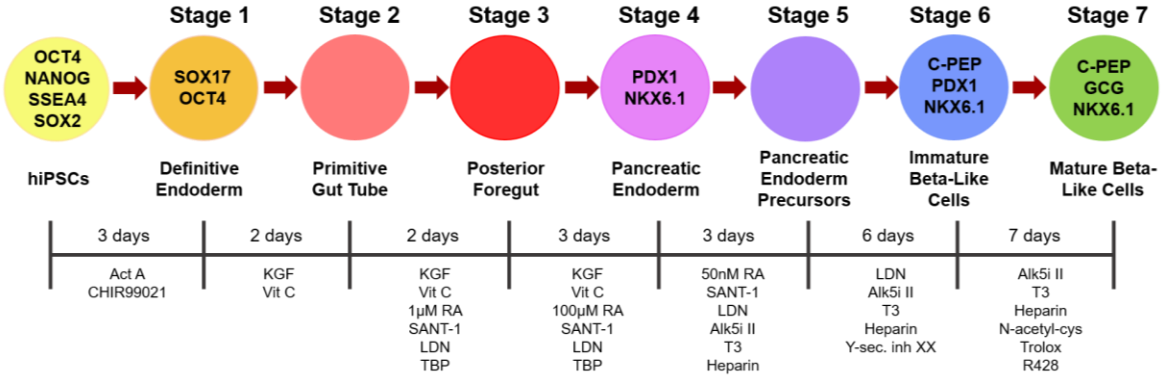


Figure 1.4.1 Overview of the *in vitro* beta cell differentiation protocol
Schematic of the step-wise *in vitro* beta cell differentiation protocol from hiPSCs to beta-like cells (Stage 7). Individual stages are represented by coloured circles. Appropriate addition of particular growth factors and small molecules (listed underneath) facilitate directed differentiation. Length of incubations are specified between stages. Expression of specific markers measured via flow cytometry at relevant stages are shown within the circles. Act A, Activin A; ALK5i II, Alk 5 Inhibitor II; CHIR99021, GSK-3 inhibitor; C-PEP, C-peptide; GCG, glucagon; hiPSCs, human induced pluripotent stem cells; KGF, keratinocyte growth factor; N-acetyl-cys, N-acetylcysteine; R428, receptor tyrosine kinase AXL inhibitor; RA, retinoic acid; T3, 3,3′,5-Triiodo-L-thyronine sodium salt; TBP, α-Amyloid Precursor Protein Modulator; Trolox, 6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid; Vit C, vitamin C; Y-sec. inh XX, ɣ-Secretase Inhibitor XX (Rezania, et al., 2014, Perez-Alcantara, et al., 2018)

More recently, differentiation protocols have concentrated on improving functionality, aiming to produce cells that are able to secrete insulin in response to glucose stimulation (Perez-Alcantara, et al., 2018, Nair, et al., 2019). To achieve this, it is crucial to have an abundant population of cells expressing both *PDX1* and *NKX6.1*. Further, using iPSCs which harbour fluorescent reporters in genes pertinent to beta cell development can enable isolation of relevant cell types from the heterogeneous population via flow assisted cell sorting (FACS) (Micallef, et al., 2012, Gupta, et al., 2018, Blochinger, et al., 2020). Moreover, the transplantation of pancreatic progenitors into immunocompromised rodents gives more functionally mature cells, illustrating that there may be external environmental signals required that are yet to be identified (Xie, et al., 2013). Studies have also demonstrated that utilising 3D culture in attempt to replicate the *in vivo*

environment can similarly improve differentiation efficiency (Pagliuca, et al., 2014, Liu, et al., 2020).

However, cell lines can exhibit different differentiation capacities (Kajiwara, et al., 2012, Rouhani, et al., 2014). Furthermore, current protocols invariably produce a heterogeneous population of immature polyhormonal cells, making functional assays a challenge (Hrvatin, et al., 2014, Pagliuca, et al., 2014, Yoshihara, et al., 2016). Despite the drawbacks associated with the functional immaturity of the beta-like cells, iPSCs are still a relevant platform for investigating disease-causing mutations, particularly with the latest innovations in genome engineering.

1.4.5.2 CRISPR/Cas9 editing in iPSCs

The emergence of Clustered Regularly Interspaced Palindromic Repeats (CRISPR) Cas9 nucleases has allowed the facilitation of introduction, or correction of, specific pathogenic mutations alongside the generation of isogenic cell lines which makes it possible to examine the same variant in multiple genetic backgrounds. Early gene editing technologies such as zinc finger nucleases (ZFN) and transcription activator like effector nucleases (TALEN) were prohibitive in their expense and complexity, with the arrival of CRISPR/Cas9 system offering a simpler and more affordable alternative (Urnov, et al., 2010, Hockemeyer, et al., 2011, Jinek, et al., 2012, Grobarczyk, et al., 2015).

The CRISPR/Cas9 system is based on the bacterial adaptive immune system response to invading pathogens (Jinek, et al., 2012, Cong, et al., 2013). Short DNA sequences are retained from previously invading bacteria as spacers in combination with a repeat element, forming pre-crRNA (Bolotin, et al., 2005, Mojica, et al., 2005). This pre-crRNA combines with a transactivating RNA (tracrRNA) to form a single guide RNA (sgRNA) which targets the Cas9 nuclease complex to the invading bacteria via recognition of the

spacer (Brouns, et al., 2008). The Cas9 then introduces a double strand break (DSB) by cleaving the pathogen DNA sequence, three nucleotides upstream of the protospacer adjacent motif (PAM) (Garneau, et al., 2010, Gasiunas, et al., 2012).

Custom 20 nucleotide synthetic sgRNAs have been successfully designed to be delivered alongside Cas9 to introduce genetic modifications to a variety of cell lines, including hiPSCs (Jinek, et al., 2012, Grobarczyk, et al., 2015). Repair of the DSB can then occur through one of two ways. Firstly, the non-homologous end joining (NHEJ) mechanism, which often leads to insertions/deletions (indels) causing frame shift mutations and thus often utilised for creating KO models. Secondly, the homology directed repair (HDR) pathway, where a repair template can be introduced to elicit single-nucleotide changes in the DNA sequence (Mali, et al., 2013).

1.4.5.3 Diabetes disease modelling using iPSCs

Human iPSCs have provided valuable mechanistic insights into monogenic forms of diabetes, both through obtaining samples from patients harbouring disease-causing mutations and through introduction of specific mutations within isogenic iPSCs lines (Teo, et al., 2013, Balboa, et al., 2018, Cardenas-Diaz, et al., 2019).

Human iPSCs have also been employed to investigate the function of transcription factors within human pancreatic development. This research demonstrated how iPSCs and CRISRP/Cas9 can be utilised together to uncover additional functions for genes, such as a role for *RFX6* in regulating the number of pancreas progenitors, complementing previous work in the EndoC-βH2 line which demonstrated a role for *RFX6* in insulin secretion (Chandra, et al., 2014, Zhu, et al., 2016).

CRISPR/Cas9 also gives the ability to correct iPSCs with patient-derived mutations and thus assess the effect of these mutations in comparison to corrected cells with the same

genetic background. For instance, a study on a patient with neonatal diabetes caused by an activating mutation in *STAT3* showed that the *STAT3* mutation instigated premature *NEUROG3* expression alongside insulin (*INS*) and glucagon (*GCG*). Further, correction of the mutation with CRISPR/Cas9 reversed the diabetic phenotype (Saarimaki-Vire, et al., 2017). Another study used HNF4A-MODY patient-derived iPSCs to investigate the effect of the mutation on both liver and pancreatic development. The results demonstrated that mutations in *HNF4A* lead to impaired foregut development and downregulation of pancreatic and liver-related genes (Ng, et al., 2019). Further examples of using iPSCs with *INS* and *HNF1A* mutations are covered later in this Chapter.

These studies demonstrate how stem cell-based diabetes models can be effective in identifying genes involved in beta cell differentiation and function, however modelling T2D variants that influence insulin secretion continues to be a challenge, as the cells lack functional maturity (Bruin, et al., 2014). Consequently, modelling of mutations using iPSCs is still in its relative infancy, as the vast majority of insulin-positive cells are polyhormonal and there are variations in the ability to produce insulin-positive cells depending on the cell lines and protocol used (Zhang, et al., 2009, Nostro, et al., 2011, Rezania, et al., 2014). Therefore, although iPSCs provide a useful approach for studying gene expression and preliminary exploration into genetic function, complementary work within the EndoC-βH1 line offers the opportunity to functionally assess different aspects of a gene's role within both pancreatic development and function.

1.5 *INS* mutations as a cause of monogenic diabetes

Mutations in the insulin gene (*INS*) were first identified in a patient with fasting hyperinsulinaemia where phenylalanine had been substituted for leucine in the insulin B-chain (p.F49L), although most patients with hyperinsulinemia do not develop overt

diabetes (Tager, et al., 1979). Hyperinsulinaemia occurs due to dramatically reduced receptor binding and biological activity of the mutated *INS* allele, prolonging its half-life. The wildtype allele then compensates for this defect by overproducing insulin (Tager, et al., 1980). Following this first patient, a spectrum of *INS* mutations were identified with variable effects (Tager, et al., 1979, Shoelson, et al., 1983, Shoelson, et al., 1983, Shibasaki, et al., 1985, Chan, et al., 1987, Yano, et al., 1992, Warren-Perry, et al., 1997). Although these patients had a clear association between their *INS* allele and hyperinsulinaemia or proinsulinaemia, many carriers were surprisingly normoglycaemic. These mutations cause defective receptor binding, improper cleavage of C-peptide from proinsulin or mis-sorting of proinsulin to the constitutive pathway where subsequent processing does not occur (Shoelson, et al., 1983, Gross, et al., 1989, Yano, et al., 1992, Warren-Perry, et al., 1997). However, affected individuals often secreted significantly more insulin compared to control subjects (Shoelson, et al., 1983, Nanjo, et al., 1986). With other factors, including aging, these individuals may develop insulin resistance and could be expected to develop diabetes (Nishi and Nanjo, 2011). However, more recent investigations have shown variants in *INS* as diabetes-causing, with more than 50 reported *INS* mutations identified as causing monogenic diabetes (Tager, et al., 1979, Stoy, et al., 2007, Liu, et al., 2015).Variants within *INS* cause a variety of phenotypes from mild adult onset diabetes (responsive to oral anti-diabetic drugs) to permanent neonatal diabetes, most often as a perturbation of wildtype insulin production (Tager, et al., 1979, Stoy, et al., 2007, Glaser, 2008). Variants within *INS* have been recognised as the second most common cause of autosomal dominantly inherited permanent neonatal diabetes mellitus (Edghill, et al., 2008). The vast majority of *INS* mutations are heterozygous autosomal dominant, however regulatory variants have also been identified which, when inherited in an autosomal recessive manner, can cause NDM. However,

functional studies have demonstrated that reduction of insulin biosynthesis rather than beta cell apoptosis is the molecular mechanism (Garin, et al., 2010). The wide range of phenotypes caused by variations in *INS* lends itself to gaining insights into different mechanisms by which the beta cell may become dysfunctional. Furthermore, specific pathways and therapeutic targets may be able to be identified which could be implemented in the treatment of more complex forms of diabetes such as T2D.

1.5.1 Previous investigations of diabetes-associated *INS* mutations

1.5.1.1 Monogenic *INS* diabetes as caused by the p.C96Y (Akita mouse) mutation

The most severe phenotypes due to variations in *INS*, such as that caused by the p.C96Y substitution, are a result of disruption of the formation of the disulphide bonds which stabilise the 3D structure of insulin, the B7-A7 and B19-A20 interchain bonds (Liu, et al., 2003). The heterozygous p.C96Y (c.287G>A; p.Cys96Tyr) variant was one of the first of the *INS* mutations to be clinically identified as a cause of permanent neonatal diabetes (Stoy, et al., 2007). The same mutation had already been previously described in a mouse model of diabetes, known as Akita (Wang, et al., 1999). Mice with this mutation develop diabetes within 3-4 weeks, with males more severely affected than females (Choeiri, et al., 2005). The mutation causes disruption in the B7-A7 bond, causing more hydrophobic interactions between the proinsulin molecules leading to aggregates of proinsulin (Wang, et al., 1999, Yoshinaga, et al., 2005). The resultant aggregates are high-molecular weight protein complexes which can engage with molecular chaperones, preventing WT proinsulin from being exported from the ER (Cunningham, et al., 2017). The exact mechanisms behind the pathogenesis of the disease are yet to be fully elucidated. As mice have two *INS* genes rather than one, in addition to the other physiological differences

between islets, translating findings from mouse to man must be done with some caution (Leroux, et al., 2001, Hay and Docherty, 2006).

1.5.2 Overexpression models of *INS* mutations

Mouse models have shown that knockout of one *INS* is compensated for by the expression of the second *INS* gene (Leroux, et al., 2001). This suggests that misfolding mutant proinsulin must be exerting its effect not only on mutant but wildtype proinsulin export from the ER, leading to ER stress and diabetes (Leroux, et al., 2001, Liu, et al., 2010). Transfection of mutant *INS*-containing vectors in appropriate cell lines have been incorporated in overexpression models to investigate effects on insulin synthesis and secretion. The studies have confirmed the dominant-negative effect of mutant proinsulin upon WT proinsulin and revealed folding differences between wildtype and mutant proinsulin (Liu, et al., 2005, Liu, et al., 2007, Hodish, et al., 2010). Studies in relevant cellular models such as Min6 and INS-1 832/13, alongside the use of transgenic mice models, have enabled different *INS* mutations to be investigated and assessed for their severity of effect and potential mechanism of action. These models have shown that mutations as a result of the disruption of disulphide bond formation trigger the dominant-negative blockade of proinsulin production from the wildtype *INS* allele, leading to the saturation of the activated UPR and consequently to apoptosis of the beta cell (Hodish, et al., 2010, Liu, et al., 2010).

1.5.3 Modelling of *INS* mutations in human stem cell models

Induced pluripotent stem cells have allowed the unique opportunity to use patient-derived *INS* mutations, alongside their isogenic corrected controls, to examine beta cells at each stage of development. Modelling of a patient-derived p.C96R (c.287T>C; p.Cys96Arg) mutation showed that there was increase in ER stress markers, such as *HSPA5*, compared

to the corrected control as revealed by single-cell RNA sequencing. This result was replicated upon transplanting the mutant differentiated cells into NOD-SCID-gamma mice and additionally demonstrated reduced insulin secretion without associated apoptosis. These results suggest that permanent neonatal diabetes could result from insufficient development of beta cell mass rather than beta cell apoptosis (Balboa, et al., 2018). A homozygous mutation of *INS* has also been modelled from a patient with no detectable insulin at birth. Differentiation of the mutant stem cells captured many features of the patient phenotype with insulin-negative endocrine stem cells that still expressed other maturity markers of beta cells. CRISPR/Cas9 correction of the mutation and transplantation into streptozotocin treated mice (causing destruction of the beta cells) led to restoration of insulin secretion and normal glucose homeostasis (Ma, et al., 2018). Despite the majority of differentiated beta-like cells being functionally immature, iPSC models have enabled investigation into the mechanisms underlying the pathogenesis of monogenic diabetes. No other method allows patient-derived mutations to be studied and provides the opportunity to develop personalised intervention and treatment. These two studies are the only ones thus far to have used iPSCs to model *INS* mutations. Therefore, there is still plenty of opportunity to investigate the ways in which monogenic diabetes *INS* variations exert their effects on both beta cell development and function.

1.6 *HNF1A*

1.6.1 *HNF1A* – a transcription factor regulating genes involved in glucose transport and lipid metabolism

HNF1A, encoded by hepatocyte nuclear factor 1 homeobox A (*HNF1A*), is a transcription factor that has a vital role in embryogenesis, including beta cell development, as well as being expressed in the adult pancreas, gut, liver and kidney (Mendel and Crabtree, 1991,

Wang, et al., 1998, D'Angelo, et al., 2010). *HNF1A* can bind to over 200 target genes in the liver and over 100 target genes in pancreatic islets, thereby regulating genes involved in glucose transport, such as GLUT1 and GLUT2, and lipid metabolism (Shih, et al., 2001, Odom, et al., 2004, Servitja, et al., 2009). *HNF1A* also directly regulates the expression of the sodium/glucose cotransporter (SGLT2) which is responsible for reabsorption of glucose in the kidneys (Pontoglio, et al., 2000). SGLT2 inhibitors are used in the treatment in T2D and have been trialled as a potential therapeutic avenue in the treatment of MODY (Hohendorff, et al., 2017).

1.6.2 *HNF1A* and its role in beta cell development

HNF1A does not exert its effects on pancreatic development in isolation. *HNF1A* is co-expressed with several other transcription factors such as *HNF1β* and *HNF4A* (Stoffers, et al., 1997, Edghill, et al., 2006). Together with *NEUROG3*, *HNF1A* activates *PAX4*, an essential gene in the formation of both beta and delta cells (Smith, et al., 2003). Besides GLUT1 and GLUT2, *HNF1A* together with *FOXA2* also controls the expression of *MGAT4A*, which is involved in glucose entry within the beta cell, with *HNF1A* additionally helping maintain normal glycaemia via induction of the expression of *ACE2*, encoding angiotensin-converting enzyme 2 (Luni, et al., 2012, Pedersen, et al., 2013). Several of the genes downregulated in *Hnf1a* knockout mouse islets, such as Glucose-6-phosphatase 2 (*G6pc2*) and transmembrane protein 27 (*Tmem27*), have their own essential regulatory functions within the pancreatic islets meaning that *Hnf1a* likely exerts its effects via multiple mechanisms (Massa, et al., 2004, Akpinar, et al., 2005, Bouatia-Naji, et al., 2008, Servitja, et al., 2009). Homozygous knockout of *Hnf1a* in mice has suggested that *Hnf1a* may regulate beta cell growth, with mice having a smaller proportion of beta cells, although this was only a qualitative analysis (Pontoglio, et al., 1998). Dominant-negative *Hnf1a* mutations also led to a reduced beta cell mass but this

may be the result of additional unspecified regulatory functions, as overexpression of wildtype *Hnf1a* causes a comparable phenotype (Hagenfeldt-Johansson, et al., 2001, Yamagata, et al., 2002, Luco, et al., 2006). Analysis of the islets from an individual with a heterozygous MODY-causing *HNF1A* mutation showed no difference in beta cell mass, albeit the data were only from one patient (Haliyur, et al., 2019).

1.6.3 *HNF1A* as a cause of MODY

Variants in *HNF1A* are the most frequent cause of MODY, with UK testing of probands revealing that up to 52% of MODY cases can be attributed to mutations in *HNF1A*, with over 400 heterozygous mutations identified so far (Yamagata, et al., 1996, Stride, et al., 2005, Shields, et al., 2010). Patients with *HNF1A* mutations exhibit beta cell dysfunction and hyperglycaemia as a result of decreased insulin secretion in response to increased blood glucose levels (Dukes, et al., 1998, Pontoglio, et al., 1998, Ellard and Colclough, 2006). Individuals with HNF1A-MODY (often referred to as MODY3) can often be misdiagnosed with other forms of diabetes but levels of C-reactive protein, usually used to identify inflammation, which is under transcriptional control by *HNF1A*, have been found to be significantly lower than in other forms of diabetes (Owen, et al., 2010, McDonald, et al., 2011, Thanabalasingham, et al., 2011). Patients typically present before 30 years of age and can have diagnosis confirmed via genetic testing, with positive diagnosis meaning that treatment can often be switched to oral sulphonylureas (Shepherd, et al., 2009, Colclough, et al., 2013). The ability to be able to distinguish HNF1A-MODY from T2D is important in providing appropriate treatment and monitoring of the disease, showing the benefit of combining a clinical and genetic approach.

1.6.4 *HNF1A* and its role in Type 2 Diabetes

Variations in *HNF1A* have been implicated in T2D, with different *HNF1A* variants influencing T2D risk depending upon ethnic background. The missense p.E508K (c.1522G>A) variant was associated with T2D risk in Latino populations, whereas the p.G319S variant has only been identified in Canadian Oji-Cre populations, where it is linked to early-onset T2D (Hegele, et al., 1999, Consortium, et al., 2014). Meta-analysis of more common variants has shown that p.I27L and p.A98V lead to a mild increase in T2D risk and *in vivo* studies of the two variants combined demonstrated a modest but significant impairment in glucose stimulated insulin secretion (Holmkvist, et al., 2006, Gaulton, et al., 2015). Other functional investigations have shown that 0.44% of the population harbour *HNF1A* variants that cause a significant increase in the risk of developing diabetes (Najmi, et al., 2017). A recent approach has developed an analytic framework that incorporates standard functional assays to stratify *HNF1A* variants along a phenotypic spectrum, with the aim to better categorise patients and improve the management and treatment of their diabetes. This approach demonstrated successful separation of causal MODY variants from benign and T2D risk variants. In addition, it enabled the identification of individuals with conflicting functional genotype and clinical phenotype (Althari, et al., 2020). Therefore, understanding the way in which monogenic *HNF1A* mutations affect beta cell function could help gain insight into the molecular mechanisms underpinning T2D.

1.6.5 *HNF1A* in animal models

One of the challenges in utilising mouse models of HNF1A is that mutant mice do not differ phenotypically from their wildtype counterparts making modelling of heterozygous mutations in mice redundant (Lee, et al., 1998, Boj, et al., 2010). This may be due to the different expression profiles of *HNF1A*, with three isoforms of *HNF1A* in humans

compared to one in rodents (Harries, et al., 2009). Recently, one homozygous hypomorphic *HNF1A* variant has been identified in insulin-treated family members, however, no other homozygous *HNF1A* mutations have been reported in humans, suggesting that most homozygous mutations are incompatible with life (Misra, et al., 2020). However, there have been numerous studies with *Hnf1a*-/- knockout mice. Loss of *Hnf1a* function in mice leads to hyperglycaemia caused by impaired insulin secretion and altered gene expression of islet-enriched transcription factors such as *Pdx1* and *Hnf4a*, as well as impairment of expression of genes involved in mitochondrial metabolism (Lee, et al., 1998, Shih, et al., 2001, D'Angelo, et al., 2010). Human *HNF1A* deficiency most severely affects the beta cells but abnormalities in other tissues have been identified with homozygous mutant mice. HNF1A null mice exhibit hepatic and renal dysfunction, leading to traits typical of Fanconi syndrome in humans (Pontoglio, et al., 1996). Taken together, this demonstrates the likelihood of cell-specific responses to *HNF1A* haploinsufficiency. This, in combination with the limitations of capturing the patient phenotype in mouse models, has led to alternative ways of modelling *HNF1A* mutations being sought.

1.6.6 *HNF1A* in human models

The only way to study primary human beta cells is to isolate cells from a cadaveric donor. In this respect, a recent human islet study aimed to characterise the pancreatic tissue and isolated islets from an individual with Type 1 Diabetes. However, the islets did not exhibit typical T1D characteristics and consequent sequencing of the donor's DNA serendipitously revealed a previously identified heterozygous *HNF1A* mutation (c.779C>T, p.Thr260Met, p.T260M) (Glucksmann, et al., 1997, Haliyur, et al., 2019). The beta cells exhibited the expected transcript and protein levels of beta cell identity markers such as *PDX1* and *NKX6.1* but reduced expression of transcription factors with a

known involvement in mature beta cell function such as *MAFA* and *RFX6*. The study confirmed the association to known *HNF1A* targets (e.g. *HNF4A, G6PC2* and *FOXA2*) in addition to identifying new *HNF1A* targets such as *ABCC8* and *KCNJ11*, which together cause disruption in GSIS. ATP-sensitive channels were reduced, with alterations in GSIS-specific glucose metabolism and ATP production pathways. Further, alpha cells secreted lower levels of glucagon in response to potent alpha cell stimuli such as low glucose (1.7mM) and exhibited an inhibitory response to membrane depolarization by KCl. Almost half of the genes differentially expressed in the mutant beta cells were also altered in alpha cells (Haliyur, et al., 2019).

An iPSC model introducing a heterozygous or homozygous *HNF1A* knockout revealed ablation of *HNF1A* leads to a gene expression signature that is more akin to alpha cells, with abnormal expression of genes related to beta cell function and development. Both HET and KO cell lines failed to increase insulin secretion in response to high glucose, although this finding was not replicated in an *HNF1A* KO EndoC-βH3 cell line which may suggest that prolonged *HNF1A* loss is required for defects in GSIS. The data also indicated that *HNF1A* inactivation impedes the development of a normal islet gene signature due to downregulation of *PAX4* with resultant upregulation of *ARX* expression preventing repression of alpha cells genes. Furthermore, there may be a dosage effect of *HNF1A* on mitochondrial function as only HET cells exhibited any respiratory defect (Cardenas-Diaz, et al., 2019). Despite being the most common cause of MODY, there has yet to be a study utilising patient-derived *HNF1A* iPSCs to study the influence of *HNF1A* on beta cell development and function. Therefore, the function of *HNF1A* in beta cells is yet to be fully elucidated.

1.7 Aims of thesis

Mutations in *HNF1A* and *INS* are recognised causes of monogenic diabetes but the cellular mechanisms driving the association to their disease remains to be determined (Stride, et al., 2005, Stoy, et al., 2007). Previous work has demonstrated that variations in *HNF1A* can cause impaired insulin secretion and beta cell dysfunction but how this occurs is not completely understood (Cardenas-Diaz, et al., 2019, Haliyur, et al., 2019). Variations in *INS* have been shown to cause misfolding of proinsulin, leading to its retention in the ER and subsequent ER stress and beta cell dysfunction (Wang, et al., 1999, Liu, et al., 2010, Cunningham, et al., 2019). However, the mechanism behind this has yet to be fully elucidated. Human islets are difficult to obtain, with limited genetic manipulation possible, and rodent islets differ in several aspects to their human counterparts. In this regard, hiPSCs were utilised, alongside CRISPR/Cas9 to artificially introduce specific coding mutations, the only human beta cell model where this is currently possible. Furthermore, these cells could be differentiated toward the pancreatic endocrine lineage, enabling the investigation of the potential developmental effects of *HNF1A*. For *HNF1A*, I compared my data to hiPSCs derived from a patient with HNF1A-MODY. To assess the effects of impaired function in a more mature cellular system, I complemented my work with both overexpression (*INS*) and knockdown models (*HNF1A*) within the EndoC-βH1 cell line. This provided the opportunity to confirm genes from my hiPSC work whilst providing additional insights into gene function. For my *INS* work, I used the *INS*-KO EndoC-βH1 line as a novel method for investigating *INS* mutations which served to avoid the confounding effects of endogenous *INS* expression. Together, these studies aim to provide data that is directly translatable to more complex forms of diabetes.

The overall aims of my thesis are to:

1. Generate and characterise a diabetes causing *INS* mutation in a hiPSCs model
2. Characterise a diabetes causing *INS* mutation in an EndoC-βH1 cell model
3. Characterise a compound heterozygous *HNF1A* knockout cell line and compare my results to those already generated for a HNF1A-MODY patient cell line

2

Materials and Methods

This Chapter gives an overview of the experimental methods used to generate the data presented in this thesis.

2.1 Cell Culture

2.1.1 General maintenance of cell lines

All cell lines were maintained at 37°C and 5% CO_2, with a MycoAlert Assay Kit (Lonza) used regularly to ensure cultures were mycoplasma free.

2.1.1.1 Maintenance of HEK293 cells

HEK293 cells were grown in Dulbecco's modified Eagle's medium (DMEM, Sigma Aldrich, D6429) supplemented with 10% FBS (foetal bovine serum) and 100U/ml penicillin-streptomycin (P/S, Gibco™, 15140122). Cells were dissociated using TrypLE™ Select (Thermo Fisher Scientific, 12563011) and passaged at a 1:20 ratio weekly.

2.1.1.2 Maintenance of EndoC-βH1 cells

The immortalised EndoC-βH1 cell line, derived from human foetal pancreatic tissue, was developed by Ravassard et al. (Ravassard, et al., 2011). In summary, the foetal pancreatic buds were transduced under the rat insulin promoter (*RIP*) with oncogene simian virus 40 tumour antigen (*SV40LT*) and transplanted into severe combined immunodeficient (SCID) mice. Insulinomas were isolated and transduced by human telomerase reverse transcriptase (*RIP-hTERT*) and these were transplanted once more into SCID mice to improve the expansion of the cells. These secondary insulinomas were excised and cultured *in vitro* as EndoC-βH1 cells (Ravassard, et al., 2011).

EndoC-βH1 cells were cultured and plated in cell dishes coated with DMEM (high glucose, Gibco™, 41965) supplemented with 1% Extracellular Matrix (Sigma Aldrich,

E1270), 2µg/ml fibronectin (Sigma Aldrich, F1141) and 100U/ml P/S. The culture media contained DMEM (low glucose, Gibco™, 31885) supplemented with 2% bovine serum albumin (BSA, Fraction V Fatty acid free, ROCHE 10775835001), 50µM 2-mercaptoethanol (Gibco™, 31350010), 10nM nicotinamide (Sigma Aldrich, N3376), 5.5µg/ml transferrin (Sigma Aldrich, T8158), 6.6ng/ml sodium selenite (Sigma Aldrich, 214485), 100U/ml P/S and 2mM L-glutamine (Gibco™, 25030081). Cells were passaged weekly ($1.2x10^6$ cells/T25 flask) using Trypsin-EDTA solution (Gibco™, 15400054).

2.1.1.3 Maintenance of hiPSCs

The human induced pluripotent stem cell (hiPSC) line SB Ad3.1 was acquired from the StemBancc consortium (www.stembancc.org) via the Human Biomaterials Resource Centre, University of Birmingham (http://www.birmingham.ac.uk/facilities/hbrc). The hiPSC were generated from human skin fibroblasts, procured from a commercial source (Lonza CC-2511, tissue acquisition number 23447). The donor was a 36 year old female Caucasian with no reported diabetes. Ethical approval was obtained from the National Research Ethics Service South Central Hampshire research ethics committee (REC 13/SC/0179). Fibroblasts were reprogrammed to pluripotency as outlined previously (van de Bunt, et al., 2016). Human iPSCs were maintained in cell culture plates coated with DMEM/F12 (Sigma Aldrich, D6429) supplemented with Matrigel diluted according to the manufacturer's instructions (Corning, 354277). Cells were grown in mTeSR1 basal medium (Stemcell Technologies, 05851) supplemented with mTeSR1 5x supplement (Stemcell Technologies, 05852), except for the NKX6.1-GFP hiPSC line, which was grown in TeSR-E8 (E8) (Stemcell Technologies, 05991), supplemented with TeSR-E8 25x supplement (Stemcell Technologies, 05992) and 100U/ml P/S. The NKX6.1-GFP reporter line was derived from the SB Ad3.1 line as outlined previously, whereby the GFP gene was inserted following the endogenous *NKX6.1* locus (Gupta, et al., 2018).

Cells were passaged as aggregates every 3-4 days or when confluent using ReLeSR™ (Stemcell Technologies, 05872). When single cells were needed, cells were released via TrypLE™ Select (Thermo Fisher Scientific, 12563011) or ACCUTASE™ (Stemcell Technologies, 07920) and media was supplemented with 10μM ROCK inhibitor Y-27632 (ROCKi, Y-27632 dihydrochloride, Abcam ab120129) for 24h post-passaging.

2.2 Genomic DNA extraction

For gDNA extraction, cells were incubated in 100μl cell lysis buffer containing 0.1% SDS, 0.04mg/ml RNase A (Sigma Aldrich, R5125) and 0.4mg/ml Proteinase K (Bioline, BIO-37037) per well of a 96 well plate at 37°C overnight. The following day, a 1/10th volume of 8M LiCl (Sigma Aldrich) and 1x volume of isopropanol were added per well and then placed on a plate shaker for 15min. Plates were centrifuged for 20min at 2,500rpm at RT, with the supernatants then tipped off. Following this, 150μl EtOH was added per well and plates were spun once more for 15min at 2,500rpm at RT. Supernatants were tipped off and pellets left to dry for 10min before resuspension in 40μl 0.1% TE buffer.

2.2.1 DNA quantification

DNA was quantified using the NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific), which measures the concentration of DNA based on sample absorbance at a wavelength of 260nm. Sample purity was assessed by measuring the 260nm/280nm absorbance ratio which provided an indication of any contamination by protein or organic compounds. DNA deemed as “pure” had a 260/280 ratio of greater than 1.8. Only DNA midiprep samples with concentrations above 100 ng/μl and that were of ideal purity were subsequently used for transfections.

2.3 Gene expression analysis

2.3.1 RNA extraction

RNA was extracted from cells using the RNeasy mini kit (Qiagen) or TRIzol® reagent (Thermo Fisher Scientific) according to manufacturer's guidelines. For TRIzol® extraction, after phase separation and isolation, RNA precipitation from the aqueous layer with isopropanol was carried out with overnight incubation at -20°C. Glycogen (R0551, Life Technologies) was added during the incubation step to increase the efficiency of recovering RNA from small amounts of starting material. RNA pellets were resuspended in RNase-free H_2O (Ambion, Thermo Fisher Scientific) and quantified using a NanoDrop 1000 spectrophotometer (Thermo Scientific) and stored at -80°C.

2.3.2 cDNA synthesis

Complementary DNA (cDNA) was synthesised from RNA using the GoScript™ Reverse Transcription System, (Promega, A5001) according to the manufacturer's instructions. Reactions were performed using MicroAmp™ Optical 8-Cap Strip PCR tubes (Applied Biosystems) in a Tetrad 2 thermal cycler (Bio-Rad Laboratories, Inc). 100ng to 1µg of input RNA was used. cDNA was stored at −20°C.

2.3.3 Real-time TaqMan® quantitative PCR

Gene expression was performed using TaqMan® real-time PCR assay run using the SDS v2.3 software (Applied Biosystems) on either the 7900HT Fast Real-Time PCR System (Thermo Fisher Scientific) or the QuantStudio™ 7 Flex Real-Time PCR System, 384-well desktop (Applied Biosystems). Details of the TaqMan® assays used are detailed in their respective results Chapters and cycling conditions are outlined in Table 2.3.1. Samples were assayed in triplicate on 384-well plates (4titude Ltd), alongside housekeeping gene TBP (TATA-box binding protein), with non-template controls used

for each assay. Each cDNA sample was analysed in triplicate wells and raw Ct values were analysed by the $2^{-\Delta\Delta Ct}$ method, which ascertains the fold change in gene expression relative to housekeeping genes and control samples.

Table 2.3.1 Real-time TaqMan® quantitative PCR conditions

Step	Temperature	Duration	Cycles
Initial denaturation and activation of Taq polymerase	50°C	2min	1
	95°C	10min	1
Denaturation	95°C	15sec	40
Primer annealing and elongation	60°C	1min	

2.4 Transcriptome profiling

2.4.1 RNA extraction, sequencing and quantification

Cells were harvested using pre-warmed ACCUTASE™, pelleted, resuspended in Buffer RLT (Qiagen) and stored at -80°C. RNA was extracted as described in Section 2.3.1. Library preparation and sequencing was undertaken by the Oxford Genomic Centre (WCHG). Polyadenylated transcripts were isolated with the NEBNext PolyA mRNA Magnetic Isolation Module (New England Biolabs, E7490). Library preparation was done using the NEBNext Ultra Directional RNA Library Kit with 12 cycles of PCR and custom 8bp indexes (New England Biolabs, E7420). Several lanes of the Illumina HiSeq4000 were utilised to allow libraries to be multiplexed and sequenced as 75-nucleotide paired-end reads. An in-house RNA-Seq pipeline was used to process FASTQ files obtained from the genomics core. The reads were mapped to the human genome build hg19 using STAR v.2.5 producing BAM (Binary Alignment Map) output files. GENCODE v19 (https://www.gencodegenes.org/releases/19.html) was used as the transcriptome reference (Harrow, et al., 2012, Dobin, et al., 2013). The featureCounts

from the Subread package v.1.5 (http://www.subread.sourceforge.net/) was used to complete gene-level quantification (Liao, et al., 2014). This produced a count table with each column representing a sequenced RNA library alongside each row denoting an Ensembl gene. The unnormalised raw number of sequencing reads were represented by count values, which were then mapped to their respective gene in each library.

Principal component analysis of the RNA-Seq datasets only included genes expressed at >1 count per million. Counts were transformed to the logarithmic scale to base 2 (log_2) and normalised in respect to library size employing the variance stabilising transformations (VST) of the DESeq2 package, which removes the dependencies of the variance on the mean.

2.4.2 Differential gene expression analysis

Software packages implemented for differential gene expression analysis are outlined in more detail in the subsequent sections. Custom scripts were written in R version 3.6.1 (https://www.r-project.org).

2.4.2.1 DESeq2

The Bioconductor package DESeq2 (Love, et al., 2014) was used to perform the differential analysis of RNA-Seq count data, according to the vignette listed on November 30, 2016 (Love et al. 2016). A DESeqDataSet was created from the unnormalised count table with a table of sample information, including genotype and developmental stage. The DESeq function works by estimating the size factors, accounting for differences in library sizes of individual samples, then estimates the dispersion for each gene and fits a negative binomial generalised linear model. Results are comprised of (a) the average of the normalised count values, divided by size factors, taken across all samples, (b) the effect size estimate (log_2FC), which represents how

much the gene expression in, for example a KO sample, has changed compared to the WT gene expression (value is reported on a logarithmic scale to base 2), (c) the standard error estimation for the log_2FC estimate, (d) a p-value, determined with the Wald test, and (e) an adjusted p-value (padj) which has been corrected for multiple testing using the Benjamini-Hochberg method with a false discovery rate (FDR) of <1% designated as a significant threshold.

2.5 Protein expression studies

2.5.1 Cell Lysis

For protein expression studies, cells were released using TrypLE™ Select, Trypsin-EDTA solution or ACCUTASE™, washed with 1xPBS and pelleted. All steps were carried out on ice. For cell lysis and protein extraction, cells were resuspended in pre-chilled RIPA buffer and 1x proteinase inhibitor cocktail (PIC, Sigma Aldrich, P8340) (Table 2.5.1), vortexed and incubated on an end-over-end rotator at 4°C for 30min. Samples were spun at 13,000rpm for 30min in a pre-chilled benchtop centrifuge. Supernatants were transferred into 1.5ml tubes and stored at -20°C and quantified by DC (detergent compatible) assay (BioRad, 5000111).

Table 2.5.1. Composition of RIPA lysis buffer used to generate whole cell extracts from pellet cells. RIPA buffer was made and stored at 4°C for up to 3 months. Protease inhibitor cocktail was added at 1:100 to working stocks of RIPA and was stored at 4°C for up to a week. All reagents purchased from Sigma Aldrich.

Reagent	Final concentration
Tris pH7.4	50mM
NaCl	150mM
Triton X-100	1%
Sodium dodecyl sulphate (SDS)	0.1%
Sodium deoxycholate	0.5%
Distilled water	-

2.5.2 Protein Quantification

A DC Protein Assay (BioRad, 5000111) was used to quantify the protein concentration of cell lysates. The assay is based on the reaction of the protein with an alkaline copper tartrate solution and Folin reagent. The reaction causes a characteristic blue colour with maximum absorbance at 750 nm and minimum absorbance at 405 nm. Absorbance was measure using a Versamax microplate reader alongside SoftMax® Pro v5 software (Molecular Devices). A BSA (Sigma Aldrich, A7517) standard curve was used to calculate protein concentrations.

2.5.3 Western Blotting

Protein lysates were separated by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) on a Bio-Rad Criterion TGX stain-free 4-20% pre-cast gel (mini: 15-well; midi: 18 or 26-well). 1x Tris/Glycine/SDS (TGS) running buffer comprised of 25mM Tris, 192mM glycine and 0.1% SDS pH 8.3 (Bio-Rad Laboratories, Inc). Samples containing 10-20µg whole-cell extract were combined with 4x Laemmli Buffer (BioRad Laboratories) and 10% β-mercaptoethanol (Sigma Aldrich), and denatured by heating at 80°C for 10min. Following this, samples were spun briefly and loaded on a gel alongside a molecular weight marker (Precision Plus All Blue standard, Bio-Rad Laboratories). Mini gels were run for 20min at 250V and midi gels were run at 250V for 30min before activation and imaging using the ChemiDoc MP Imaging System and ImageLab 5 software (Bio-Rad Laboratories). Proteins were transferred to a 0.2µm polyvinylidene difluoride (PVDF) membrane at 2.5A for 7min using the Trans-Blot Turbo transfer system (Bio-Rad Laboratories). Proper transfer of proteins was confirmed by subsequent imaging. Prior to addition of primary antibody, membranes were blocked for at least 30min at room temperature (RT) in 5% non-fat milk powder in phosphate buffered saline (PBS, Sigma Aldrich), supplemented with 0.1% Tween-20 (Fisher

Scientific (PBST)). Primary antibodies used are listed in their relevant results Chapters. Blots were incubated overnight at 4°C with primary antibodies, diluted in 5% milk/PBST. Following incubation, blots were washed in PBST for 3x5min and then incubated with secondary antibody for 1h at RT. Secondary antibodies (listed in Table 2.5.2) were made up in 5% milk/PBST. Membranes were washed again for 3x5min in PBST and protein bands detected using Clarity Western Enhanced Chemiluminescence Substrate (Bio-Rad Laboratories). Visualisation of the protein bands was performed using a ChemiDoc MP and subsequently quantified for signal intensity using Image Lab™ (Bio-Rad Laboratories) and adjusted for surrounding background signal. The intensity of the protein of interest was normalised to the signal intensity of the loading control on the same blot.

Table 2.5.2. Secondary antibodies used for western blot. All antibodies were reconstituted in 50% glycerol in distilled water and stored at -20°C. Antibodies were diluted in 5% milk/PBST.

Antibody	Source	Dilution
Anti-mouse HRP	Thermo Fisher, 31450	1:2500
Anti-rabbit HRP	Thermo Fisher, 31460	1:2500

2.5.4 Transient Transfections

For transfection experiments, cells were plated in parallel to DNA/transfection reagent preparation and cultured in antibiotic-free media. In summary, FuGENE® 6 Transfection Reagent (Promega) was mixed with Opti-MEM reduced serum-free medium (Gibco™, 31985062) and incubated for 5min before plasmid DNA was added. The mix was then incubated for an additional 15-20min and pipetted dropwise onto cells (transfection reagent:DNA mixture ratios of 1:3 for HEK293 and EndoC-βH1 cells, and 1:5 for hiPSCs). Transfected cells were collected 48h post-transfection for downstream analysis.

2.5.5 RNA Interference

Gene silencing was performed in EndoC-βH1 cells using small interfering RNAs (siRNAs) in combination with lipid-based reverse transfection. ON-TARGETplus siRNAs were purchased from Dharmacon (GE Healthcare) as a pool of four oligonucleotides (SMART pool). The siRNAs used are detailed in their relevant chapters. siRNAs were diluted in Opti-MEM reduced serum-free medium (Gibco™, 31985062) to a final concentration of 20nM, mixed with 0.4% Lipofectamine RNAiMAX (Thermo Fisher Scientific, 13778150) and incubated for 15-20min at RT. Transfection complexes were added prior to addition of cells resuspended in culture media. Silencing efficiency was measured 96h later by RT-qPCR and/or WB analysis.

2.5.6 Flow cytometry

Human iPSCs and differentiated cells were treated with TrypLE™ to create a single-cell suspension and then fixed with either 4% paraformaldehyde (Thermo Fisher Scientific) or BD Cytofix™ fixation buffer. Cells were then permeabilised using BD Phosflow Perm Buffer III (Biosciences) or BD Perm/Wash™ buffer and stained with different cell surface or intracellular markers. Isotype antibodies were used to assist in determining gating. Details for antibodies used, including isotypes are listed Table 2.5.3. Samples were run on either the SH800 Cell Sorter (Sony) or the BD LSRFortessa™ (Biosciences).

Table 2.5.3 FACS antibodies including isotype controls

Antibody	Dilution	Conjugate	Cat #
NANOG	1:10	PE	560589
OCT4	1:10	PerCP-Cy™ 5.5	560589
SOX2	1:10-1:40	AF-647	560589
SSEA-4	1:10	FITC	560126
IgG_1, κ isotype	1:10	PE	560589
IgG_1, κ isotype	1:10	PerCP-Cy™ 5.5	560589
IgG_{2a}, κ isotype	1:10	AF-647	560589
IgG_3, κ isotype	1:10	FITC	555578
CXCR4	1:40	PE	FAB173P
C-peptide	1:200	AF-647	565831
NEUROD1	1:40	PE	563001
NKX6.1	1:40	AF-647	563338
NKX6.1	1:40	PE	563023
PDX1	1:40	PE	562161
SOX17	1:40	AF-488	562205
IgG_{2B} isotype	1:40	PE	IC0041P
IgG1, κ isotype	matched	PE	554680
IgG1, κ isotype	matched	AF-488	557721
IgG1, κ isotype	matched	AF-647	557732

All antibodies derived from mice. Top, pluripotency markers; bottom, differentiation stage-specific expression markers; matched, amount of isotype control was matched to amount of antibody used. All antibodies purchased from BD Pharmingen™, except for CXCR4 and IgG_{2B} isotype, which were purchased from R&D Systems.

All genome-engineered hiPSCs were assessed for pluripotency markers using the BD Human Pluripotent Stem Cell Transcription Factor Analysis Kit (BD, Biosciences, 560589) following the manufacturer's instructions. Markers included OCT4, NANOG, SOX2 and stage-specific embryonic antigen (SSEA-4).

The efficiency of *in vitro* differentiation was assessed using flow cytometry by using stage-specific expression markers for endocrine pancreas development. Developmental markers included SYR-box 17 (SOX17) and C-X-C chemokine receptor type 4 (CXCR4) for definitive endoderm (DE); the transcription factors PDX1 and NKX6.1 for pancreatic endoderm (PE); NKX6.1 and C-peptide for beta like cells (BLCs).

2.6 CRISPR/Cas9 genome editing in hiPSCs

The generation of hiPSC models by CRISPR/Cas9 genome editing is summarised in Figure 2.6.1. The overall genome editing pipeline consisted of three different steps which can be outlined as (1) design, preparation and evaluation of the sgRNAs (2) gene editing in both SB and NKX6.1-GFP reporter lines, and (3) quality control (QC) checks carried out on the correctly edited hiPSC lines. Correctly edited lines were differentiated to beta-like cells (BLC) if they passed QC. Cell lines were then characterised via cellular phenotyping (Chapters 3 and 5) and transcriptome profiling (Chapters 4 and 5) at distinct stages of endocrine development (see Chapter specific methods).

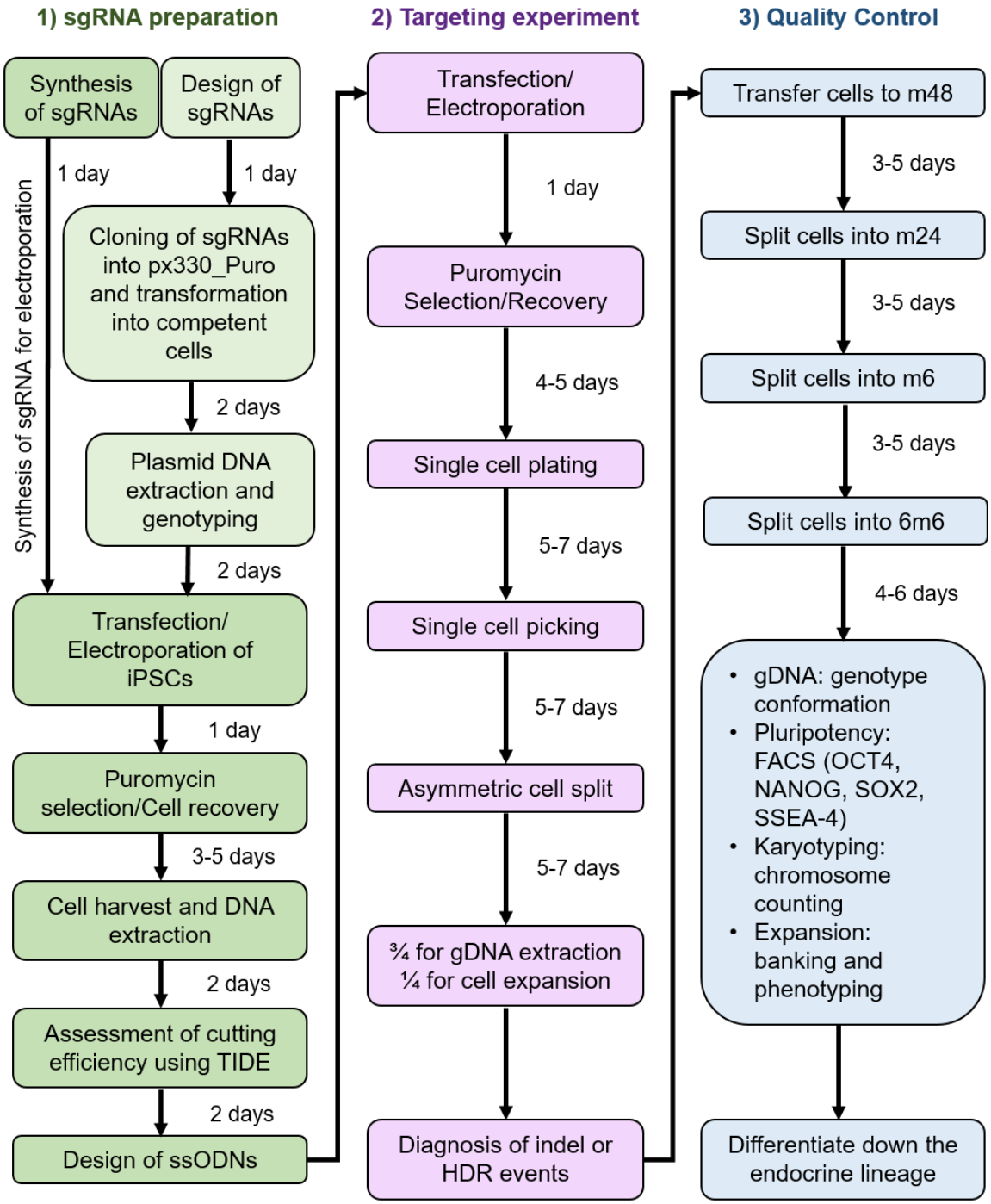


Figure 2.6.1 Schematic: Generation of hiPSC lines

Workflow summarising the steps performed to prepare and choose sgRNAs targeting the gene of interest. For gene KO two sgRNAs were selected, for point mutations one sgRNA and an ssODN repair template were designed and manufactured. Either lipid-based transfection or electroporation of hiPSCs was used, with transfected cells selected with puromycin and electroporated cells left to recover. Cell clones of the required genotype underwent quality control (QC) and hiPSC lines that passed QC were differentiated and characterised. The time required to complete each step is indicated between boxes if applicable. sgRNA, single guide RNA; ssODN, single-stranded oligodeoxynucleotide; HDR, homology directed repair; TIDE, Tracking of Indels by Decomposition.

2.7 *In vitro* differentiation of hiPSC towards beta-like cells

Genome-edited hiPSC lines were differentiated along the endocrine lineage into BLCs. For differentiation, $1x10^6$ ($0.5x10^6$ for NKX6.1-GFP lines) hiPSC were seeded as single cell suspension in mTeSR™1 supplemented with 10μM ROCKi per well onto 12-well plates (Corning CellBind, 3336) pre-coated with GFR (growth factor reduced)-Matrigel diluted 1:30 (Corning, 356230). 8-10h after plating, medium was changed to fresh mTeSR™1 or TeSR™-E8™ (E8). *In vitro* differentiation was initiated ~24h following plating, according to the protocol detailed in Perez-Alcantara et al. 2018. To summarise, media was supplemented at specific stages with small molecules and recombinant growth factors which enabled sequential generation of cells representing seven essential developmental stages of the endocrine pancreas: DE (definitive endoderm), PGT (primitive gut tube), PFG (posterior foregut), PE (pancreatic endoderm), EP (endocrine precursors), EN (endocrine-like cells) and BLC (beta-like cells) (Figure 2.7.1). Tables 2.7.1 and 2.7.2 document the basal differentiation media composition in addition to stage-specific growth factor information. Genetically engineered hiPSC lines were differentiated in parallel using the same culture and differentiation media. Cells remained in planar culture for the entire 28 days.

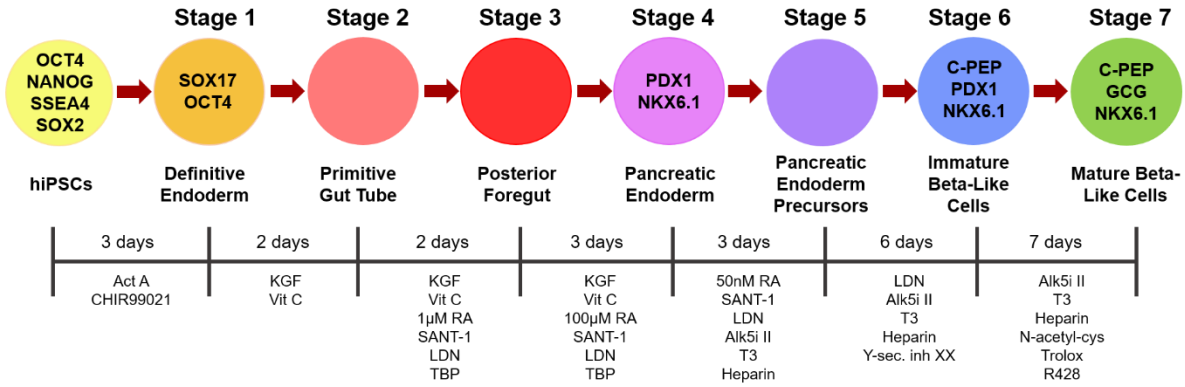


Figure 2.7.1 Summary of the *in vitro* beta cell differentiation protocol
Schematic of the step-wise *in vitro* beta cell differentiation protocol from hiPSCs to BLCs. Individual stages are represented by coloured circles. Appropriate addition of particular growth factors and small molecules (listed underneath) facilitate directed differentiation. Length of incubations are specified between stages. Expression of specific markers measured via flow cytometry at relevant stages are shown within the circles. Act A, Activin A; ALK5i II, Alk 5 Inhibitor II; CHIR99021, GSK-3 inhibitor; C-PEP, C-peptide; GCG, glucagon; hiPSCs, human induced pluripotent stem cells; KGF, keratinocyte growth factor; N-acetyl-cys, N-acetylcysteine; R428, receptor tyrosine kinase AXL inhibitor; RA, retinoic acid; T3, 3,3′,5-Triiodo-L-thyronine sodium salt; TBP, α-Amyloid Precursor Protein Modulator; Trolox, 6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid; Vit C, vitamin C; Y-sec. inh XX, γ-Secretase Inhibitor XX (Rezania, et al., 2014, Perez-Alcantara, et al., 2018).

Table 2.7.1. Basal Differentiation Media

Stage 1-2	Stage 3-4	Stage 5-7	Manufacturer/Supplier	Cat #
MCDB131	MCDB131	MCDB131	Thermo Fisher Scientific	10372019
0.1% PBS	0.1% PBS	0.1% PBS	Sigma Aldrich	P0781
1.5g/l $NaHCO_3$	2.5g/l $NaHCO_3$	1.5g/l $NaHCO_3$	Thermo Fisher Scientific	25080060
1x Glutamax	1x Glutamax	1x Glutamax	Thermo Fisher Scientific	35050038
10mM Glucose	10mM Glucose	20mM Glucose	Thermo Fisher Scientific	A2494001
0.5% BSA	2% BSA	2% BSA	Roche	1077585001
	1:200 ITS-X	1:200 ITS-X	Thermo Fisher Scientific	51500056
		10μM Zinc sulfate	Sigma Aldrich	Z0251

ITS-X, Insulin-Transferrin-Selenium-Ethanolamine

Table 2.7.2 Formulation of beta cell differentiation media including specific growth factors and small molecules added each day during differentiation. These components were added at the corresponding day to the basal medium described in Table 2.7.1.

Stage	Days	Medium	Factors	Concentration	Supplier	Cat #
Definitive Endoderm (DE)	1	MCDB131 (Stage 1-2)	Activin A	100ng/ml	PeproTech	120-14
			CHIR	3μM	Axon Medchem	1368
	1	MCDB131 (Stage 1-2)	Activin A	100ng/ml	PeproTech	120-14
			CHIR	0.3μM	Axon Medchem	1368
	1	MCDB131 (Stage 1-2)	Activin A	100ng/ml	PeproTech	120-14
Primitive Gut Tube (PGT)	2	MCDB131 (Stage 1-2)	KGF	50ng/ml	PeproTech	100-19
			Ascorbic acid	0.25mM	Sigma Aldrich	A4544
Posterior Foregut (PFG)	2	MCDB131 (Stage 3-4)	KGF	50ng/ml	PeproTech	100-19
			Ascorbic acid	0.25mM	Sigma Aldrich	A4544
			Retinoic acid	1μM	Sigma Aldrich	R2625
			Sant-1	0.25μM	Sigma Aldrich	S4572
			LDN193189	100nM	Stemgent	04-0074
			PKC act V (TBP)	200nM	Merck	565740
Pancreatic Endoderm (PE)	3	MCDB131 (Stage 3-4)	KGF	2ng/ml	PeproTech	100-19
			Ascorbic acid	0.25mM	Sigma Aldrich	A4544
			Retinoic acid	0.1μM	Sigma Aldrich	R2625
			Sant-1	0.25μM	Sigma Aldrich	S4572
			LDN193189	200nM	Stemgent	04-0074
			PKC act V (TBP)	100nM	Merck	565740

CHIR, GSK-3 inhibitor; KGF, keratinocyte growth factor; T_3, 3,3,5-Triiodo-L-thyronine sodium salt (thyroid hormone); R428, receptor tyrosine kinase AXL inhibitor. Table continued on the following page

Stage	Days	Medium	Factors	Concentration	Supplier	Cat #
Endocrine Precursors (EP)	3	MCDB131 (Stage 5-7)	Retinoic acid	0.05µM	Sigma Aldrich	R2625
			Sant-1	0.25µM	Sigma Aldrich	S4572
			LDN193189	100nM	Stemgent	04-0074
			ALK5 Inhibitor II	10µM	Enzo Life Sciences	ALX-270-445
			T3	1µM	Sigma Aldrich	T6397
			Heparin sodium salt	10µg/ml	Sigma Aldrich	H3149
Endocrine-like cells (EN)	7	MCDB131 (Stage 5-7)	LDN193189	100nM	Stemgent	04-0074
			ALK5 Inhibitor II	10µM	Enzo Life Sciences	ALX-270-445
			T3	1µM	Sigma Aldrich	T6397
			Heparin sodium salt	10µg/ml	Sigma Aldrich	H3149
			ɣ-Secretase Inhibitor XX	100nM	Merck	565789
Beta-like cells (BLC)	7	MCDB131 (Stage 5-7)	ALK5 Inhibitor II	10µM	Enzo Life Sciences	ALX-270-445
			T3	1µM	Sigma Aldrich	T6397
			Heparin sodium salt	10µg/ml	Sigma Aldrich	H3149
			N-acetyl-Cys	1µM	Sigma Aldrich	A9165
			Trolox	19µM	Merck	648471
			R428	2µM	Selleck Chemicals	S2841

CHIR, GSK-3 inhibitor; KGF, keratinocyte growth factor; T3, 3,3,5-Triiodo-L-thyronine sodium salt (thyroid hormone); R428, receptor tyrosine kinase AXL inhibitor

2.8 Statistical Analyses

Statistical analyses were performed in GraphPad Prism 8 (v8.3) with results from multiple experiments expressed as mean ±SEM. For two unmatched groups, the p-value was determined using a two-tailed unpaired t-test. Comparison of groups larger than two was done with a one-way ANOVA followed by Sidak's multiple comparisons test. A p-value of <0.05 was considered statistically significant, with n representing the number of biologically independent experiments. Further details of statistical testing are outlined in the relevant chapters or figure legends.

3

Generation and characterisation of a diabetes causing *INS* mutation in an hiPSC model

Concepts or data presented in this chapter have been published previously or contributed to the following manuscript, which is provided in full in the Appendices.

Wesolowska-Andersen, et al. Analysis of Differentiation Protocols Defines a Common Pancreatic Progenitor Molecular Signature and Guides Refinement of Endocrine Differentiation. *Stem Cell Reports* 2020; 14(1):138-153

3.1 Introduction

Monogenic diabetes, caused by mutations in a single gene, represents around 1-3% of all diabetes cases (Hattersley, et al., 2006, Shields, et al., 2010, Pihoker, et al., 2013, Shepherd, et al., 2016). The majority of monogenic diabetes can be classified as neonatal diabetes mellitus (NDM) or maturity-onset diabetes of the young (MODY) (Misra and Owen, 2018). NDM usually presents in the first six months of life, although some cases have also been identified within the first year, with 30 genes currently known to cause NDM (Edghill, et al., 2006, Rubio-Cabezas, et al., 2012, De Franco, et al., 2015, Johnson, et al., 2017, De Franco, et al., 2020). NDM mutations are typically autosomal dominant, with mutations in *ABCC8*, *KCNJ11* or *INS* being the most common cause (Gloyn, et al., 2004, Babenko, et al., 2006, Flanagan, et al., 2006, Proks, et al., 2006, Greeley, et al., 2011). However, compound heterozygous or homozygous loss of function mutations in *GCK* can also cause permanent NDM (Raimondo, et al., 2014). MODY is similarly often autosomal dominant but is generally diagnosed later in childhood (<25 years) and is most often caused by mutations in *GCK* or *HNF1A* (Froguel, et al., 1992, Yamagata, et al., 1996, Ellard, et al., 2008). MODY has also been diagnosed in patients with *ABCC8* and *KCNJ11* mutations, demonstrating that variation within a single gene can cause a spectrum of phenotypes (Klupa, et al., 2009, Liu, et al., 2013). However, in consanguineous families, autosomal recessive mutations are the most common cause of permanent neonatal diabetes as a result of Wolcott-Rallison syndrome. This rare recessive disorder is caused by mutations in *EIF2AK3* and results in not just diabetes but skeletal dysplasia and renal and hepatic dysfunction (Rubio-Cabezas, et al., 2009). Wolfram syndrome, caused by mutations in *WSF1*, encoding wolframin can be a rare autosomal dominant cause of NDM, in addition to recessive loss-of-function mutations causing early-onset diabetes (Gunn, et al., 1976, De Franco, et al., 2017). Wolframin has been

shown to protect against ER stress by targeting *ATF6* for degradation, indicating that variations within *WFS1* will trigger ER stress, leading to beta cell apoptosis (Fonseca, et al., 2005, Yamada, et al., 2006, Fonseca, et al., 2010).

Monogenic diabetes provides the opportunity to understand the ways in which single genes can impact on both pancreatic development and function. Furthermore, this knowledge can lead to optimising of treatments, as seen in patients with mutations in K_{ATP}-channels, where sulphonylureas rather than insulin are used in managing the disease (Gloyn, et al., 2004, Pearson, et al., 2006). Investigating monogenic diabetes can also help understand more complex forms of diabetes, as common variation in several monogenic diabetes genes is associated with altered T2D risk (e.g. *HNF1A*, *GCK* and *NEUROD1*). There can be a considerable overlap between the phenotype of T2D and monogenic diabetes which can make them challenging to differentiate (Fuchsberger, et al., 2016).

INS mutations have been implicated in both MODY and permanent neonatal diabetes (Shoelson, et al., 1983, Stoy, et al., 2007). Several of these mutations cause misfolding of proinsulin, leading to its retention in the ER and subsequent ER stress and beta cell dysfunction (Wang, et al., 1999, Liu, et al., 2010, Cunningham, et al., 2019). However, the mechanism behind this has yet to be fully elucidated. Previous investigations into *INS* mutations have often focussed on the p.C96Y mutation, also known as the Akita mutation (Wang, et al., 1999, Liu, et al., 2007, Cunningham, et al., 2017). Akita mice, which carry an autosomal dominant missense mutation in *Ins2*, display many features of diabetes shown in humans, characterised by increased expression of ER chaperones, ER morphological changes and decreased number of secretory granules (Yoshioka, et al., 1997, Kayo and Koizumi, 1998, Wang, et al., 1999, Izumi, et al., 2003, Zuber, et al., 2004). These findings have also been demonstrated in the Munich *Ins2*^{C95S} mutant mouse

(Herbach, et al., 2007). However, dissimilarities in mouse pancreas development and architecture, alongside differences in rodent insulin promoters to that of humans, means that results from animal experiments on *INS* must be interpreted with caution (Hay and Docherty, 2006, Steiner, et al., 2010, Jennings, et al., 2015). Recently, human embryonic stem cells, alongside hiPSCs, have emerged as a viable alternative to model human endocrine pancreas development, with the parallel development of CRISPR/Cas9 enabling the precise introduction and/or correction of disease-causing mutations to be investigated (Stepniewski, et al., 2015, Braverman-Gross, et al., 2018). Furthermore, unlike embryonic stem cells, hiPSCs provide the opportunity to derive cells from patients with disease (Saarimaki-Vire, et al., 2017, Balboa, et al., 2018, De Franco, et al., 2020). CRISPR/Cas9 can also be utilised to introduce clear markers of maturation in iPSCs in the form of reporter lines. These reporter lines facilitate the isolation and purification of selected populations in generating several different specialised cell types (Calatayud, et al., 2019, Sone, et al., 2019, Fiedorowicz, et al., 2020). This has included the development of reporter lines that facilitate the enrichment of a beta-like cell population at the end point of differentiation (Gupta, et al., 2018, Blochinger, et al., 2020).

3.2 Experimental aims

To date, very few *INS* mutations have been modelled in an iPSC cell line (Balboa, et al., 2018, Ma, et al., 2018). To investigate the mechanism by which *INS* mutations may cause beta cell dysfunction, I used CRISPR/Cas9 to introduce the p.C96Y *INS* mutation into an iPSC line (SB). The original plan was to study p.C96Y, as it is the same mutation as in the Akita mouse, allowing comparison between different models. However, I was unable to isolate any clones containing the p.C96Y mutation and therefore employed an alternative strategy. I targeted exon 1 of *INS* with several sgRNAs and subsequently chose the sgRNA with the highest cutting efficiency. I selected *INS* mutations which were

close to the sgRNA cutting site and which captured a spectrum of clinical phenotypes. I aimed to introduce the mutations (p.C43G, p.R46Q and p.F49L) into both the SB line and the parental line (SB) with a reporter marker (NKX6.1-GFP). I was unable to successfully introduce the p.R46Q mutation and only obtained one heterozygous clone for p.F49L. However, this approach generated an isogenic *INS* p.C43G cell line, alongside WT controls, in both the SB and NKX6.1-GFP cell lines. Both cell lines were differentiated along the endocrine lineage and characterised by cellular phenotyping at stages 1, 4, 6 and 7 of beta cell development. The overall aims of my experiments were to:

- Generate a hiPSC line with a disease causing *INS* mutation using a parental line with a reporter marker to facilitate purification of beta-like cells
- Establish whether the mutation alters *INS* gene expression at the RNA and protein level
- Establish whether the mutation alters ER stress and cell death in this model

3.3 Methods

3.3.1 Genome Editing

3.3.1.1 sgRNA design and cloning into pX330-Puro-Cas9 plasmid

As a first strategy, the MIT CRIPSR online design tool (http://crispr.mit.edu) was used to design sgRNAs targeting p.C96Y (c.287G>A) in exon 2 of the insulin gene. Oligonucleotides targeting a region close to the position of the p.C96Y substitution were purchased from Eurofins Genomics, annealed and sub-cloned into the BsbI restriction enzymes sites in plasmid pX330-Puro-Cas9, creating pX330-Puro-*INS*-Exon2, which contained both Cas9 and a puromycin selection cassette (Cong et al. 2013). The sequence of each sgRNA and its cutting efficiency is given in Table 3.3.1. Each sgRNA was assessed for cutting efficiency using TIDE (Tracking of Indels by Decomposition) analysis (hhtps://tide.deskgen.com) (Brinkman, et al., 2014). TIDE is a bioinformatics tool that quantifies the frequency of non-homologous end joining (NHEJ) DNA repair and identifies the most frequent type of insertions or deletions (indels) in the DNA from a pool of cells transfected or electroporated with the relevant sgRNA. The TIDE web tool requires input of both the targeted and control PCR-amplified DNA sequences, in addition to the sequence of the sgRNA. For TIDE analysis, the genomic region around the CRISPR/Cas9 target sites was amplified by PCR primers located outside of the sgRNA binding areas. Primer pairs and conditions used for PCR amplification and sequencing are listed in Tables 3.3.2 and 3.3.3. The sgRNA with the highest Cas9 nuclease cutting efficiency was chosen for subsequent gene editing experiments.

3.3.2 Transfection of hiPSCs

SB hiPSCs were plated at a density of 200,000 per well on pre-coated 24 well plates. Cells were transfected with Cas9-sgRNA-expressing plasmids and ssODNs (exon 1 or 2) using FuGENE 6 (Promega) and grown in mTeSR1 containing 10μM ROCK inhibitor (YM27632 dihydrochloride, Abcam ab120129) for 24h at 37°C and 5% CO_2. For selection of successfully transfected cells, cells were maintained in mTeSR1 containing 400ng/ml puromycin for 48h. Once cells reached 80% confluency, they were plated at a very low density (2,000 cells/60mm dish), expanded and single colonies were picked ~7 days layer and maintained in individual wells of a 96-well plate. As soon as they reached confluency, cells were split into two new pre-coated 96-well plates, one for expansion and one for gDNA extraction and subsequent indel/HDR evaluation.

Table 3.3.1 Single guide RNA (sgRNA) sequences targeting *INS* (NM_001185098.2) exon 1 or 2 and their cutting efficiency as assessed by TIDE

Targeting region	Forward Primer (5' to 3')	Reverse Primer (3' to 5')	Cutting Efficiency
Exon 1	AGCCTCGTTCC CCGCACACT	AGTGTGCGGGG AACGAGGCT	14%
Exon 1	AAGCTCTCTAC CTAGTGTGC	GCACACTAGGT AGAGAGCTT	58%
Exon 1	CAAGGGCCTTT GCGTCAGGT	ACCTGACGCAA AGGCCCTTG	76%
Exon 1	CCACCAGGTGT GAGCCGCAC	GTGCGGCTCAC ACCTGGTGG	22%
Exon 1	CCCTAATGGGC CAGGCGGC	GCCGCCTGGCC CATTAGGG	6%
Exon 1	TCTACACACCC AAGACCCGC	GCGGGTCTTGG GTGTGTAGA	15%
Exon 1	CACAGGTGTTG GTTCACAA	TTGTGAACCAA CACCTGTG	11%
Exon 1	GGGCAGGAGGC GCATCCAC	GTGGATGCGCC TCCTGCCC	41%
Exon 1	ATCAGAAGAGG CCATCAAGC	GCTTGATGGCC TCTTCTGAT	4%
Exon 1	GTTCACAAAGG CTGCGGCT	AGCCGCAGCCT TTGTGAAC	38%
Exon 1	CTAGGTAGAGA GCTTCCACC	GGTGGAAGCTC TCTACCTAG	18%
Exon 2	CCACAATGCCA CGCTTCTGC	GCAGAAGCGTG GCATTGTGG	4%
Exon 2	CCTGCAGAAGC GTGGCATTG	CAATGCCACGC TTCTGCAGG	8%
Exon 2	AGTAGTTCTCC AGCTGGTAG	CTACCAGCTGG AGAACTACT	6%

Table 3.3.2 Primer pairs for *INS* (NM_001185098.2) PCR and Sanger sequencing

Gene/Primer	Forward primer (5' to 3')	Reverse primer (3' to 5')
*INS*_Exon1	CTCGTGAAGCATGTGGGG	AGGACGTGACCAAGAGAACT
*INS*_Exon2	TGACTGTGTCCTCCTGTGTC	CATACTGGACCCTGAGCCAC

Table 3.3.3 *INS* (NM_001185098.2) PCR conditions

Step	Temperature	Duration	Cycles
Initial Denaturation	94°C	10min	1
Denaturation	94°C	1min	32
Annealing	62°C	1min	
Elongation	72°C	1min	
Final extension	72°C	10min	1

3.3.3 Design of ssODN repair templates

The single-stranded oligodeoxynucleotide (ssODN) repair templates (PAGE-purified, Eurogentec) were designed with homologous genomic arms flanking both sites of the target region containing the genetic variant, a new restriction site and a silent mutation in the protospacer adjacent motif (PAM) sequence to block the re-ligation of CRISPR/Cas9.

To generate the *INS* p.C96Y hiPSC line, an ssODN repair template was designed and co-transfected with pX330-Puro-*INS*-Exon2. Transiently puromycin-resistant cells were genotyped via PCR amplification and digested with AluI, with successfully digested PCR products sent for Sanger sequencing (Eurofins Genomics, Germany). Table 3.3.4 gives the sequence of the ssODN template used.

A second strategy, employing several oligonucleotides targeting exon 1 of *INS* were similarly designed and subcloned into pX300-Puro-Cas9. Table 3.3.1 outlines the sequences of the sgRNAs tested. The sgRNA with the highest cutting efficiency was

selected and three ssODN repair templates were designed to introduce the p.C43G (c.127 T>G), p.R46Q (c.137G>A) or p.F49L (c.148 C>G) mutation. A wildtype ssODN was provided alongside the relevant *INS* mutation, to minimise the chance of indels. Table 3.3.4 details the sequences for each ssODN. Of the CRISPR/Cas9 edited clones, the p.C43G mutation was most commonly detected and taken forward for downstream experiments. Consequently, a new ssODN template was designed to introduce a silent mutation (p.C43C – referred to as SB sham) in the SB line to create a CRISPR sham. For downstream experiments, the selected sgRNA was synthesised using the EnGen® sgRNA synthesis kit (NEB, E3322V) to form a ribonucleoprotein complex with Cas9.

Table 3.3.4 ssODN design for generation of *INS* hiPSC clones

Mutation	Sequence (5' to 3')
C43C_sham	CCCAGCCGCAGCCTTTGTGAACCAACACCTGTGCGGCTCACA CCTGGTGGAAGCTCTCTACCTAGTGTGTGGAGAACGAGGCTT CTTCTACACACCCAAGACCCGCCGGGAGGCAGAGGACCTGC AGGGTGAGCCAAC
C43G	GACCCAGCCGCAGCCTTTGTGAACCAACACCTGTGCGGCTC ACACCTGGTGGAAGCTCTCTACCTCGTGGGCGGAGAACGAG GCTTCTTCTACACACCCAAGACCCGCCGGGAGGCAGAGGAC CTGCAGGGTGAGCCAAC
R46Q	AGCCTTTGTGAACCAACACCTGTGCGGCTCACACCTGGTGGA AGCTCTCTACCTAGTGTGCGGAGAACAAGGCTTCTTCTACAC ACCCAAGACCCGCCGGGAGGCAGAGGACCTGCAGGGTGAG CCAACTGCCCATTG
F49L	GAACCAACACCTGTGCGGCTCACACCTGGTGGAAGCTCTCT ACCTAGTGTGCGGAGAACGAGGCTTCTTGTACACACCCAAG ACCCGCCGGGAGGCAGAGGACCTGCAGGGTGAGCCAACTGC CCATTGCTGCCCCTGG
C96Y	GTGCAGGCAGCCTGCAGCCCTTGGCCCTGGAGGGGTCCCTG CAGAAGCGTGGCATTGTGGAACAATGCTATACTAGCATCTG CTCTCTCTACCAACTGGAGAACTACTGCAACTAGACGCAGCC CGCAGGCAGCCCCAC
WT	CTTTGTGAACCAACACCTGTGCGGCTCACACCTGGTGGAAGC TCTCTACCTAGTGTGCGGAGAACGAGGCTTCTTCTACACACC CAAGACCCGCCGGGAGGCAGAGGACCTGCAGGGTGA

Mutations introduced are underlined

3.3.3.1 Electroporation of ribonucleoprotein (RNP) complexes

For electroporation experiments, the sgRNA targeting exon 1 with the highest cutting efficiency was re-designed and ordered so it could be synthesised as a RNP complex - a purified Cas9 nuclease duplexed with a sgRNA (Eurofins, Germany). Electroporation of the RNP complex was used in preference to transfection due to the higher editing efficiency, lower chance of off-target effects and quicker recovery of the cells (Hendel, et al., 2015). The sgRNA sequence 5'-TTCTAATACGACTCACTATAGAAGCTCTCT ACCTAGTGTGCGTTTTAGAGCTAGA-3'was synthesised using the EnGen® sgRNA synthesis kit (NEB, E3322V). Both the EnGen 2X sgRNA Reaction Mix, *S. pyogenes* and the target-specific DNA oligo (1µM) were thawed on ice. Each component was mixed and spun in a microfuge prior to use and placed on ice before assembling the reaction at room temperature (Table 3.3.5). The reaction was mixed thoroughly and incubated at 37°C for 2 hours before transferring to ice.

Table 3.3.5 Components of the RNP mix

Reagent	Volume
Nuclease-free water	3µl
EnGen 2X sgRNA Reaction Mix, *S. pyogenes*	10µl
Target-specific DNA Oligo (1µM)	5µl
EnGen sgRNA Enzyme Mix (add last)	2µl

The mix was then DNase treated using 20µl transfection mix, 28µl nuclease-free water and 2µl DNaseI and incubated at 37°C for 15min. The sgRNA was purified from the reaction using the RNA Clean and Concentrator Kit® (Zymo Research, R1015) and stored at -20°C.

The sgRNA was defrosted on ice and then an electroporation transfection mix was prepared and incubated at RT for 15min to allow for RNP complex formation (Table 3.3.6).

Table 3.3.6 Components of the electroporation transfection mix

Reagent	Volume
sgRNA	1µl (20µM/681ng)
Cas9	0.6ul (20µM)
Buffer R (Invitrogen, #MPK1096R)	4.4µl

During the incubation time, iPSCs were passaged using ACCUTASE® in 0.5ml mTeSR or E8 media with 0.5µl ROCK inhibitor, spun and resuspended in 15µl buffer R (Neon™ Transfection System 10µl Kit, Invitrogen, MPK1096) to a final amount of $2x10^5$ cells/tube. The electroporation complex mix was added and incubated for 5min on ice. Electroporation was performed using the Neon™ Transfection System Kit with the parameters: 1,200 volts, 30 ms, 2 pulses. Cells were seeded on a Matrigel-coated 48 well plate, with 200µl mTeSR or E8 with ROCKi. After 48h, 2,000 cells were plated on 60mm dishes, supplemented with SMC4 media (Table 3.3.7).

Table 3.3.7 Composition of SMC4 media. Small molecule identity, concentration and target of each component in SMC4. Targeted cellular pathway is based on the product data sheet. Small molecules purchased from Stemcell Technologies.

Small molecule	Concentration	Targeted cellular pathway	Cat. number
Thiazovivin	5µM	Rho/ROCK	72252
CHIR99021	1µM	GSK3	72052
PD0325901	0.4µM	MEK/ERK	72182
SB431542	2µM	Activin/BMP/TFG-β	72232

3.3.3.2 Analysis of NHEJ and HDR events

For genotyping, the genomic regions around the CRISPR/Cas9 target sites were amplified by PCR primers located outside of the sgRNA binding areas or outside of the ssODN template (Table 3.3.2). The PCR amplified DNA fragments were visualised via agarose gel electrophoresis. PCR products that were successfully digested, indicating ssODN template integration, or where a digestion strategy could not be performed, were purified using *Exo1* and rSAP and sent for Sanger sequencing using the same primer pairs that were used for PCR amplification (Table 3.3.2). For heterozygous mutations, PCR products were subcloned into pGEM®-T (Promega) following the manufacturer's guidelines. Plasmid DNA was transformed into XL-1 Blue Competent cells (Agilent, 200130), then extracted with the PureYield™ Plasmid Miniprep kit (Promega) and sent for Sanger sequencing to verify the sequence of individual alleles.

3.3.4 Differentiation experiments

Section 2.7 describes a seven stage *in vitro* differentiation protocol applied to direct the hiPSC lines towards the pancreatic endocrine lineage.

3.3.5 Cell sorting using FACS

NKX6.1-GFP cells lines were collected for fluorescence-activated cell sorting (FACS) at stages 6 and 7 of differentiation. Samples were dissociated with TrypLE™ Select (Thermo Fisher Scientific), quenched with MCDB131 media (Thermo Fisher) and pelleted before being resuspended in 1xPBS. Samples were stored on ice prior to sorting. The NKX6.1-GFP reporter line has been shown to accurately capture *NXK6.1* expression, without disrupting the endogenous *NKX6.1* loci (Gupta, et al., 2018). Sorting of the NKX6.1-GFP lines was performed using a FACS SH800 Cell Sorter (Sony) with GFP

fluorescence acting as a proxy for NKX6.1 expression. A negative control, the SB line, without GFP expression, was used to establish correct gating.

3.3.6 Quantitative real-time PCR (RT-qPCR)

For quantification of gene expression, RT-qPCR was performed as outlined in section 2.3.3. The specific Taqman probes used for gene expression analysis in this Chapter are detailed in Table 3.3.8. Data was analysed using the ddCT method, whereby gene expression of a test sample is compared to a control sample on the same plate.

Table 3.3.8 TaqMan® gene expression assays used for real-time PCR and their target amplification regions. All assays purchased from Thermo Fisher Scientific.

Target genes	Assay number	Target region
DDIT3	Hs99999172_m1	Spans exon boundary
HSPA5	Hs99999174_m1	Spans exon boundary
INS	Hs00355773_m1	Spans exon boundary
NKX6.1	Hs00232355_m1	Spans exon boundary
TBP	Hs00427620_m1	Spans exon boundary

3.3.7 Western Blotting analysis

Cells were collected, lysed in RIPA buffer and quantified using a DC assay (BioRad). Samples were denatured by boiling at 80°C for 10 minutes in 4xLaemmli buffer (BioRad) with β-mercaptoethanol and then separated using SDS-PAGE (BioRad Criterion). 10μg protein was loaded per lane with blots blocked in 5% PBST. Antibodies used for Westerns are summarised in Table 3.3.9 and Table 3.3.10 and were incubated either overnight at 4°C or for 1-2hrs at RT. Tubulin was used as a loading control. Protein bands were detected using Clarity Western Enhanced Chemiluminescence Substrate (BioRad) and visualised on a ChemiDoc MP. Signal intensities of bands were quantified by densitometry analysis using Image Lab™ (BioRad). The intensity of protein of

interest was normalised to the intensity of tubulin on the same blot. Differences in normalised protein expression were evaluated a two-way analysis of variance (ANOVA).

Table 3.3.9 Primary Antibodies used for Western Blotting. Antibodies were diluted in 5% milk/PBST.

Antibodies	Recommended dilution	Species	Manufacturer	Cat#
Anti-insulin	1:1000	mouse	Santa Cruz	sc-377071
Anti-DDIT3	1:1000	rabbit	Abcam	ab179823
β-Tubulin (E-10)	1:2000	mouse	Santa Cruz	sc-365791

Table 3.3.10 Secondary Antibodies used for Western Blotting

Antibodies	Recommended dilution	Manufacturer	Cat#
Anti-mouse HRP	1:2500	Thermo Fisher Scientific	31450
Anti-rabbit HRP	1:2500	Thermo Fisher Scientific	31460

3.3.8 Statistical Analyses

Statistical analyses were performed in GraphPad Prism 8 (v8.3) with results from multiple experiments expressed as mean ±SEM. For two unmatched groups, the p-value was determined using a two-tailed unpaired t-test. Comparison of groups larger than two was done with a one-way ANOVA followed by Sidak's multiple comparisons test. A p-value of <0.05 was considered statistically significant, with n representing the number of biologically independent experiments.

3.4 Results

3.4.1 Technical challenges in generating an hiPSC p.C96Y model

Both the SB and NKX6.1-GFP lines were sequenced covering the *INS* region prior to gene editing to ensure both were wildtype. To generate the hiPSC *INS* p.C96Y line, I tested three sgRNA targeting exon 2 in the SB line. To activate the HDR pathway, allowing precise introduction of the relevant mutation, an ssODN repair template was delivered alongside the sgRNA with the highest cutting efficiency. As well as the desired allele, the repair template contained two silent synthetic mutations: one that destroyed the protospacer adjacent motif (PAM) sequence to abolish cleavage by Cas9 after ssODN-directed repair, and one that introduced a new restriction enzyme site (AluI) for future evaluation of CRISPR clones.

Initial TIDE analysis of the three sgRNAs gave low cutting efficiency of 5-10%. The sgRNA with the highest cutting efficiency was taken forward for transfection alongside the ssODN repair template (Figure 3.4.1). PCR amplification of the 497bp fragment, followed by restriction enzyme digest by AluI gave no clones with successful introduction of p.C96Y ssODN-directed repair. In addition, re-evaluation of the pool of cells showed a cutting efficiency of below 6%, suggesting the initial results could have been the consequence of poor Sanger sequence read quality creating background sequence noise. The experiment was repeated with no improvement in cutting efficiency or introduction of the p.C96Y mutation. Due to the difficulties in targeting exon 2, an alternative approach was needed to target a different region of *INS*.

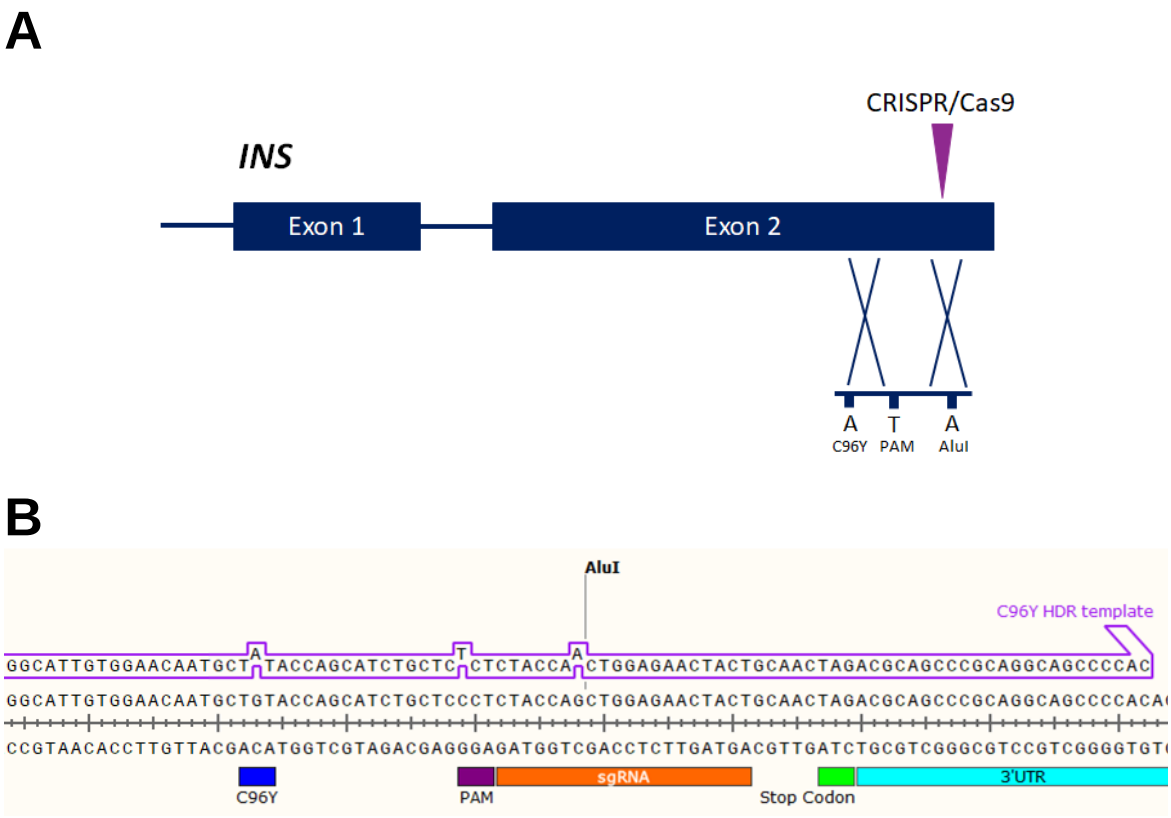


Figure 3.4.1 Genome editing strategy to generate *INS* p.C96Y hiPSC line
(A) Schematic of *INS* highlighting the location of the sgRNA used to generate the *INS* p.C96Y hiPSC line. The SB Ad3.1 hiPSC line was genetically engineered to introduce a heterozygous p.C96Y mutation. PAM – protospacer adjacent motif. (B) The position of the *INS* p.C96Y mutation on exon 2 is indicated by the blue box. The sgRNA sequence is shown as an orange box with the PAM in purple. A section of the *INS* p.C96Y homologous directed repair template sequence is shown above the nucleotide sequence with a purple outline. The introduction of the p.C96Y mutation introduces a new restriction enzyme site (AluI) for future evaluation of CRISPR/Cas9 editing

3.4.2 Generation of hiPSC *INS* models

A second strategy was employed to introduce an *INS* mutation. Several sgRNA targeting exon 1 were designed (see Table 3.3.1) with the sgRNA with the highest cutting efficiency taken forward for downstream experiments. Three *INS* mutations close to the sgRNA with differing clinical phenotypes and that were amenable to CRISPR/Cas9 and ssODN design were selected: p.C43G (c.127 T>G), p.R46Q (c.137G>A) and p.F49L (c.148 C>G). To generate these hiPSCs *INS* cell lines, the relevant ssODN repair template was provided with the sgRNA targeting exon 1 (Table 3.3.4) and electroporated

into the SB line alone in the first instance. As with the p.C96Y strategy, besides the desired allele, the repair templates contained silent synthetic mutations: one that destroyed the PAM sequence to abolish cleavage by Cas9 after ssODN-directed repair, and one that introduced a new restriction enzyme site for future evaluation of CRISPR clones (Table 3.4.1). In some repair templates, there was not a silent mutation that enabled introduction of a new restriction enzyme site. Of the CRISPR/Cas9 edited clones, the p.C43G mutation was most commonly detected and taken forward for downstream experiments. Consequently, a new ssODN template was designed to introduce a silent mutation (p.C43C) in the SB line to create a CRISPR sham. A CRISPR sham was included as increased passage number from undergoing the CRISPR pipeline can lead to acquisition of large copy number variants which may affect differentiation potential (Merkle, et al., 2017).

Table 3.4.1 List of hiPSC clones with mutation, PAM (protospacer adjacent motif) and restriction enzyme sites introduced

Clone	Mutation	PAM mutation	Restriction site introduced (nucleotide change)
C43C_sham	T to C	G to A	None introduced
C43G	T to G	G to A	BssSαI (A to C)
R46Q	G to A	G to A	None introduced
F49L	C to G	G to A	BsrGI (C to G)
C96Y	G to A	C to T	AluI (G to A)
WT	N/A	G to A	N/A

3.4.3 Genotyping of edited *INS* hiPSC lines

PCR amplification of the 442bp fragment, followed by Sanger sequencing, revealed that no clones had incorporated the R46Q ssODN template. Clones which were successfully digested with the relevant restriction digest enzymes and therefore harbouring either the

p.F49L or p.C43G ssODN-repair template were sent for Sanger sequencing to verify their genotype. As the p.C43G mutation had the highest rate of success, the same process was also performed in the NKX6.1-GFP reporter line. A summary of the results for the generation of *INS* mutations are outlined in Table 3.4.2.

Table 3.4.2 Summary of results for the generation of *INS* mutations

Clone	Number of clones picked	Number of clones with correct genotype
NKX6.1-GFP C43G	96	3
NKX6.1-GFP CRISPR sham	96	6
SB C43G	96	3
SB C43C (CRISPR sham)	96	2
SB R46Q	96	0
SB F49L	96	1

3.4.4 Quality control of *INS* hiPSC models

Clones with successful integration of either the *INS* p.C43G or *INS* p.C43C mutation, alongside the parental cell lines SB and NKX6.1-GFP, were assessed for characteristic hiPSC features, such as pluripotency and genomic integrity, to ensure that genome editing had not had an effect on hiPSC properties (Table 3.4.3). Flow cytometry was used to evaluate expression of the pluripotency markers OCT4, NANOG, SOX2 and SSEA-4 (Figure 3.4.2).

Clones were also assessed for genomic stability with chromosomal counting performed by Dr Daniela Moralli at the Chromosome Dynamics Core (WCHG). Metaphases were DAPI stained, counted and inspected for chromosomal rearrangements. These rearrangements include aneuploidy (addition or absence of a chromosome), insertions or deletions of a few megabases, and translocations (Trask, 2002). A sample had to have a median count of 46 chromosome and no metaphase containing more than 46

chromosomes to pass karyotyping. Lower chromosome counts were likely due to technical issues, for instance loss of a chromosome during wash steps. Table 3.4.3 shows the number of clones that passed karyotyping using these criteria. Overall, the three selected SB clones had a mean of 54.3±7.2% diploid cells, whilst the three selected NKX6.1-GFP clones had a mean of 64.8±14.2% diploid cells (Figure 3.4.2).

Table 3.4.3 Summary of quality control of edited *INS* hiPSC clones

Clone	Picked	Correct genotype	Pluripotent	Diploid
SB C43G	96	3	3	1
SB C43C CRISPR sham	96	2	1	1
NKX6.1-GFP C43G	96	3	2	1
NKX6.1-GFP CRISPR sham	96	6	3	2

Pluripotent for markers OCT4, NANOG, SOX2 & SSEA-4 as measured by flow cytometry, diploid karyotype of 46 chr/cell

To investigate the effects of the *INS* p.C43G mutation on beta cell development and function, one WT, one CRISPR sham, and one p.C43G clone for each of the SB and NKX6.1-GFP cell lines were taken forward for *in vitro* beta cell differentiation.

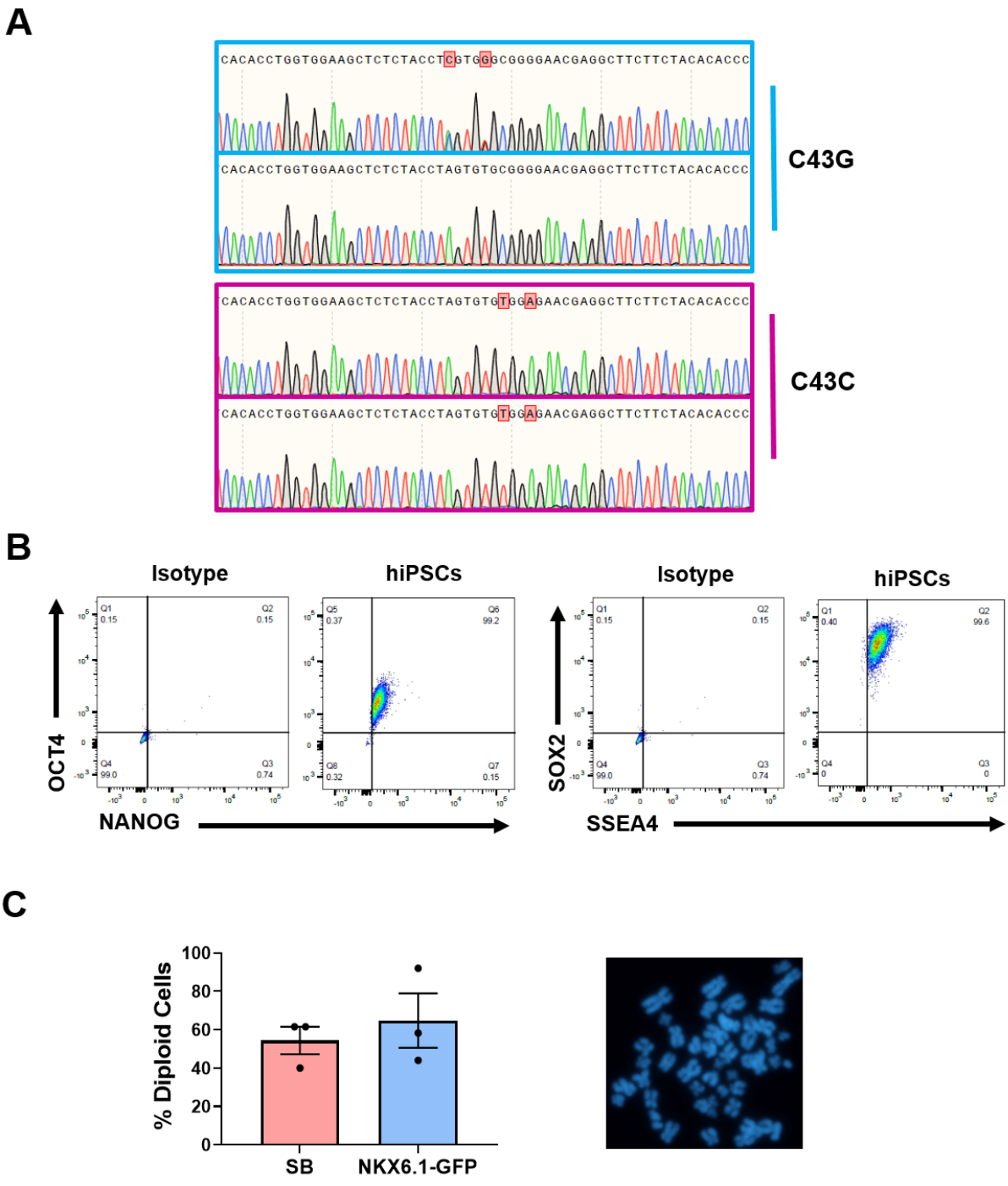


Figure 3.4.2 Characterisation of *INS* hiPSC lines for stem cell properties
(A) Sanger sequencing traces of the *INS* p.C43G and p.C43C (SB CRISPR sham) hiPSC lines. Sequencing of the p.C43G line revealed heterozygosity for c.127 T>G and one silent mutation at c.123 A>C (ssODN-introduced BssSαI restriction enzyme site). The p.C43C line was homozygous for c.129 C>T with one silent mutation at c.132 G>A (ssODN-introduced PAM modifying mutation) (B) Edited hiPSC lines were assessed for expression of common pluripotency markers including OCT4, NANOG, SSEA-4 and SOX2. (C) Karyotyping was done via chromosome counting of DAPI-stained metaphases (blue). SB n=3, NKX6.1-GFP n=3, values are displayed as mean±SEM and analysed using two-tailed unpaired t-test

3.5 Differentiation of SB *INS* p.C43G and NKX6.1-GFP *INS* p.C43G hiPSC models and associated control lines into pancreatic beta-like cells and cellular phenotyping at selected stages of development

To model pancreatic endocrine development a directed step-wise protocol was undertaken (see Section 2.7). Flow cytometry analysis was performed to assess whether the p.C43G genotype had any effect on differentiation efficiency. In addition, RNA and protein lysates were collected at stages 6 and 7. As the SB line has been differentiated numerous times within our group previously, it was compared to the NKX6.1-GFP line, which had not been extensively differentiated within our laboratory to ensure that the introduction of the NKX6.1-GFP reporter had not had any adverse effect on differentiation efficiency. Further to this, for each cell line, data was compared between each group of wildtype, CRISPR sham and mutant.

3.5.1 Evaluation of developmental markers of beta-cell differentiation in *INS* p.C43G and WT clones

As outlined in Section 1.4.5.1 and Section 2.7, *in vitro* endocrine development is characterised by expression of particular beta-cell development markers at specific stages. To assess whether there was any variation in differentiation efficiency between groups, flow cytometry was performed to compare stage-specific developmental markers.

Stage 1 – Definitive endoderm

Mouse and human ESCs initiate formation of DE via the transcription factor *SOX17*, marked by expression of the chemokine receptor CXCR4 (Yasunaga, et al., 2005, D'Amour, et al., 2006, Qu, et al., 2008, Seguin, et al., 2008). Assessment of the expression of CXCR4 using flow cytometry showed no significant differences between groups (WT, CRISPR sham and mutant) in either the SB (p=0.357) or NKX6.1-GFP lines

(p=0.888) (Figure 3.5.1). CXCR4 expression levels for all groups for both the SB and NKX6.1-GFP lines were comparable to those in other published studies (Bruin et al. 2014). Similarly, there was no significant difference in the levels of SOX17 for all SB (p=0.692) and NKX6.1-GFP (p=0.910) cell lines, with the cell lines expressing similar levels of SOX17 (p=0.0935).

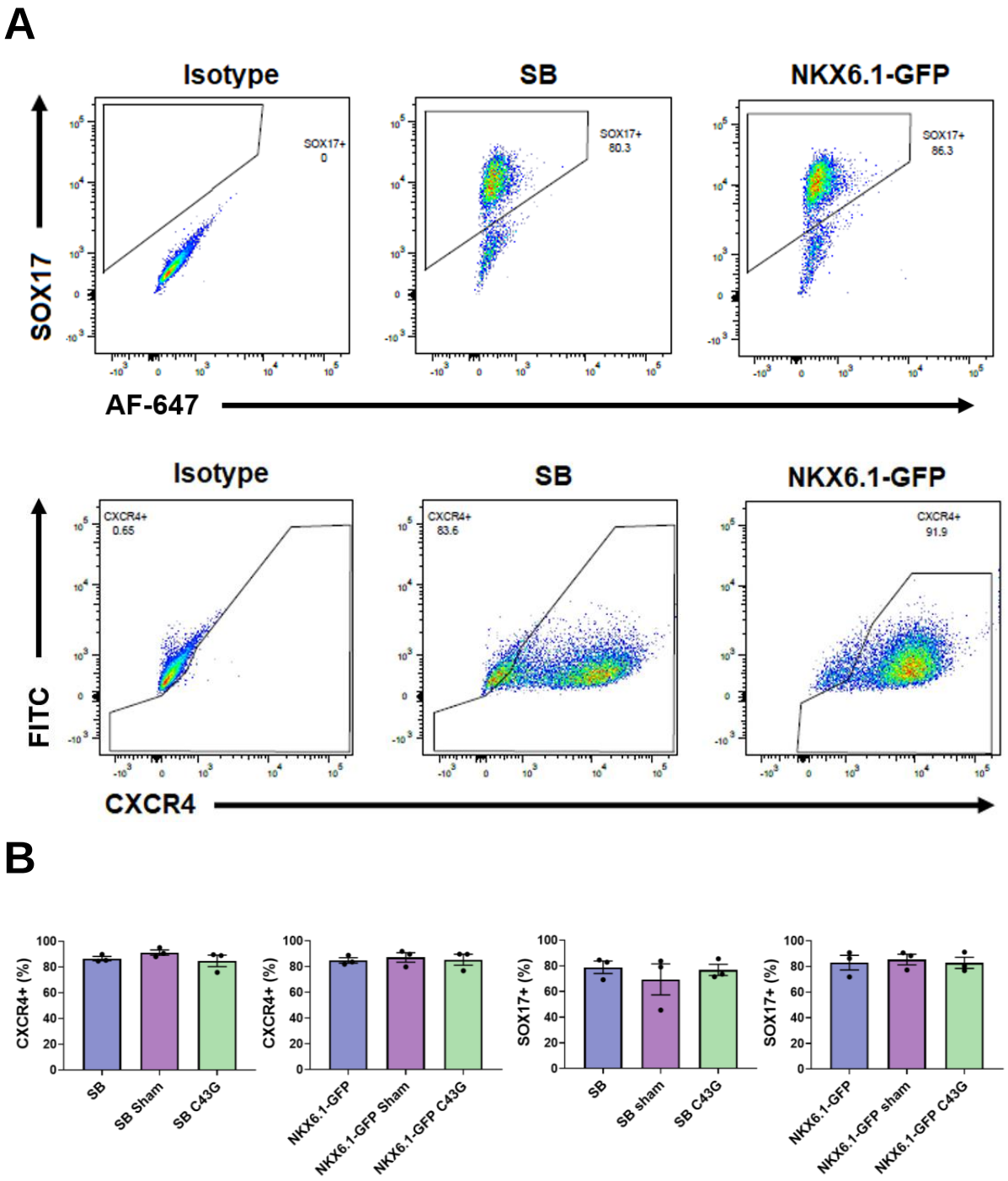


Figure 3.5.1 Assessment of markers for stage 1 (DE) of differentiation
Differentiated *INS* clones were evaluated for expression of SOX17 and the chemokine cell receptor CXCR4. (A) Representative flow cytometry plots for SOX17 and CXCR4 for the WT SB and NKX6.1-GFP cell lines. (B) There were no significant differences in the expression of CXCR4 or SOX17 in either the SB or NKX6.1-GFP cell lines. SB lines n=3, NKX6.1-GFP lines n=3. Plots display mean±SEM. Data analysed using one-way ANOVA.

Stage 4 - Pancreatic Endoderm

Levels of NKX6.1 expression were low to barely detectable in most of the clones at stage 4 (PE), compared to markedly higher levels as seen by others using a similar protocol for differentiating hiPSCs (Rezania et al. 2014). PDX1 expression was more robust in the SB clones with 45.1±3.2%, 43.2±4.5 and 54.1±4.9% in WT, CRISPR sham and p.C43G clones respectively (p=0.238). Although there was no significant difference in PDX1 expression between groups in the NKX6.1-GFP cell lines (p=0.664), there was more variability between the differentiations (Figure 3.5.2). There was a significant difference between SB WT and the other two SB lines (sham and mutant) for co-expression of NKX6.1 and PDX1 (p=0.0293) but similar levels for both markers in the NKX6.1-GFP lines (p=0.352) (Figure 3.5.2). The SB lines and NKX6.1-GFP lines had similar levels of expression for NKX6.1 and PDX1 singularly and combined.

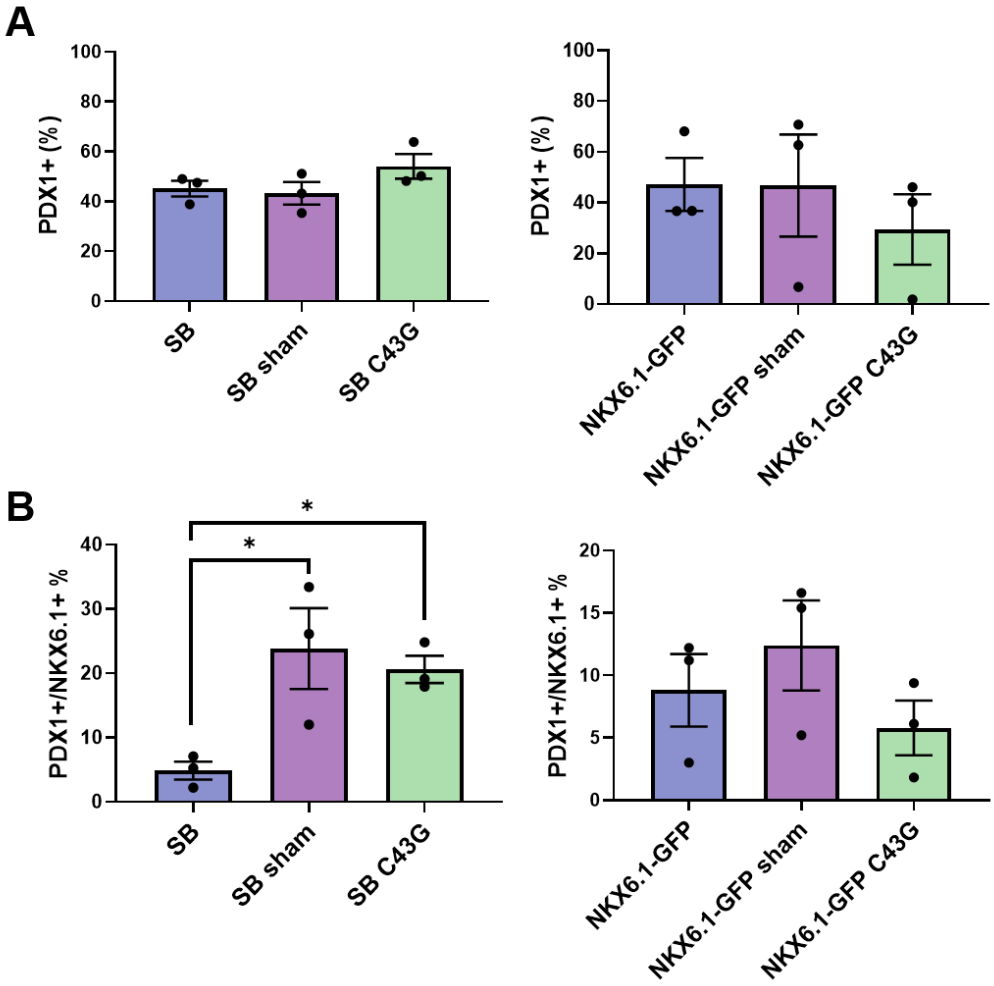


Figure 3.5.2 Assessment of markers for stage 4 (PE) of differentiation
Differentiated *INS* clones were evaluated for expression of PDX1 and PDX1/NKX6.1 combined. (A) Representative flow cytometry plots for PDX1 for the SB and NKX6.1-GFP cell lines. There were no significant differences in the expression of PDX1 in either the SB or NKX6.1-GFP cell lines (B) There was a significant difference between SB WT and the other two SB lines (sham and mutant) for co-expression of NKX6.1 and PDX1 but similar levels for both markers in the NKX6.1-GFP lines. SB lines n=3, NKX6.1-GFP lines n=3. Plots display mean±SEM *p<0.05 for one-way ANOVA

Stage 6 – Endocrine-like cells

Endocrine-like cells were assessed for differentiation efficiency by testing the expression levels of the hormones insulin and glucagon, as well as NKX6.1 via flow cytometry. Due to the presence of ITS-X (Insulin-Transferrin-Selenium-Ethanolamine) in the differentiation culture media and potential uptake of insulin into the cells, C-peptide was used as a surrogate of insulin expression (Rajagopal, et al., 2003, Hansson, et al., 2004). C-peptide, a by-product of insulin synthesis, is secreted in equimolar amounts to insulin and as such was used to quantify insulin positive cells in flow cytometry analysis. The SB clones had broadly similar levels of C-peptide expression ($p=0.111$), although these levels were lower than have been previously reported in studies using a similar differentiation protocol (Figure 3.5.3) (Wesolowska-Andersen, et al., 2020). In the NKX6.1-GFP clones, there was a significant higher expression of C-peptide in the NKX6.1-GFP CRISPR sham clones ($p=0.0031$), although this difference was lost when clones were compared for co-expression of glucagon and insulin ($p=0.0660$) (Figure 3.5.3). The expression of C-peptide was significantly higher in the SB clones than the NKX6.1-GFP clones ($p=0.0112$) (Figure 3.5.4). This suggests that the significantly higher expression of C-peptide in the NKX6.1-GFP CRISPR sham clones is likely due to technical variation but it cannot be ruled out that there may be undetected off-target effects acquired through genome engineering that may affect the differentiation ability of individual clones. Glucagon expression in the SB and NKX6.1-GFP clones was also lower than expected at this stage. Co-expression of both C-peptide and glucagon was low in both the SB and NKX6.1-GFP clones, indicating that the clones may not be as functionally mature as expected at this stage.

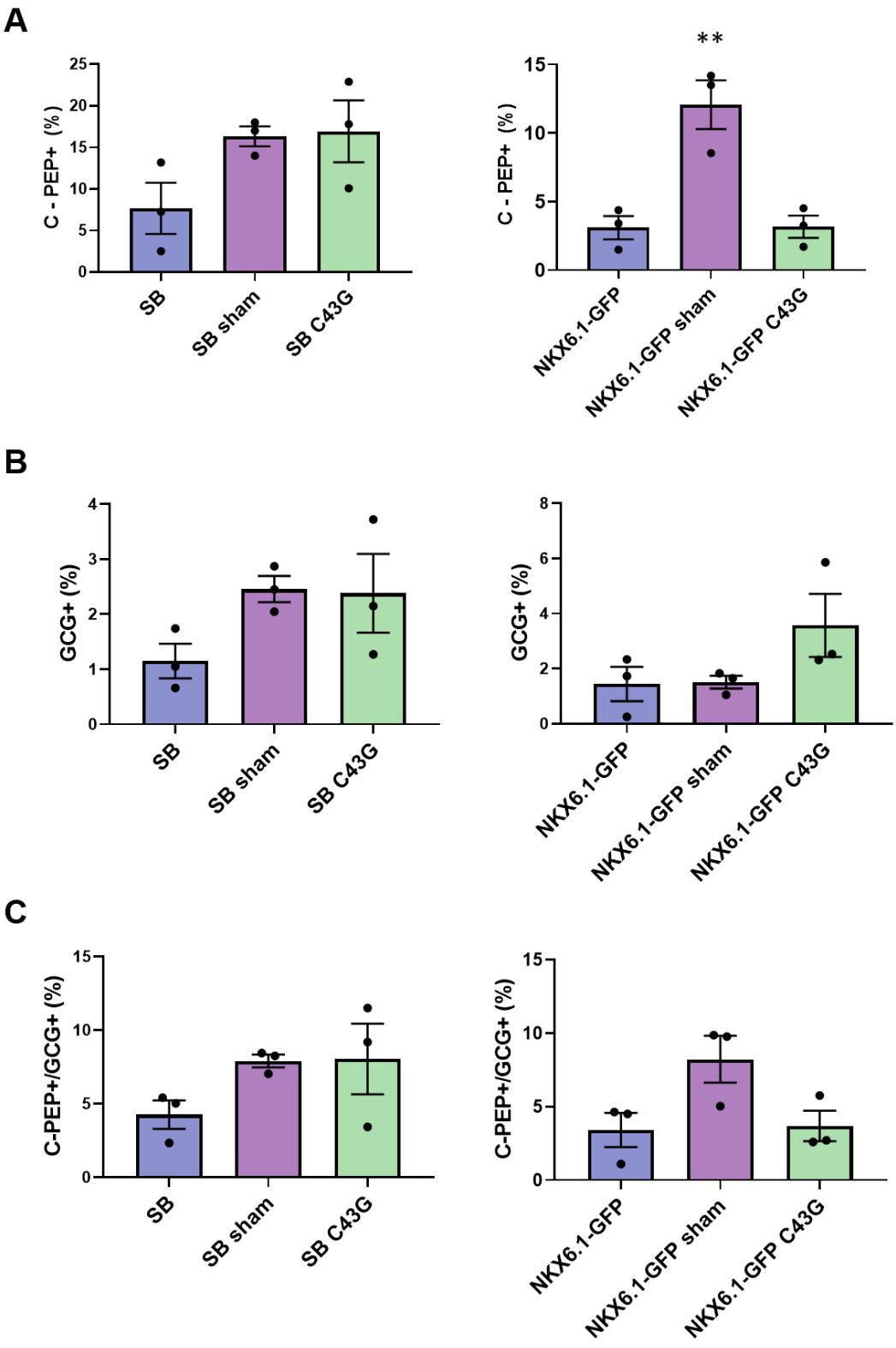


Figure 3.5.3 Expression of C-peptide in NKX6.1-GFP clones at stage 6
Differentiated *INS* clones were evaluated for expression of (A) C-peptide, (B) glucagon and (C) co-expression of C-peptide and glucagon. There was a significantly higher expression of C-peptide alone in the NKX6.1-GFP CRISPR sham. C-PEP, C-peptide; GCG, glucagon. SB lines n=3, NKX6.1-GFP lines n=3. Plots display mean±SEM, **p<0.01 for one-way ANOVA

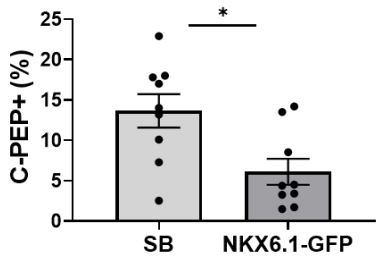


Figure 3.5.4 Significantly higher levels C-peptide in the SB clones vs the NKX6.1-GFP clones
Differentiated *INS* clones were evaluated for expression of C-peptide via FACS. There was a significantly higher expression of C-peptide in the SB clones compared to the NKX6.1-GFP clones. C-PEP=C-peptide. SB lines n=3, NKX6.1-GFP lines n=3. Plots display mean±SEM, *p<0.05 for two-tailed unpaired t-test

Stage 7 – Beta-like cells

Beta-like cells (BLCs) were assessed for differentiation efficiency in the same way as stage 6. Once again, SB clones showed similar levels of C-peptide expression (p=0.0621). As with stage 6, the NKX6.1-GFP CRISPR sham clones had significantly higher expression of C-peptide (p=0.0169), with all clones having a significantly higher expression of C-peptide than at stage 6 (p=0.0266) (Figure 3.5.5). Glucagon expression increased from stage 6 in all the SB and NKX6.1-GFP clones, with no significant differences between groups (p=0.171 and p=0.162, respectively). However, the NKX6.1-GFP lines expressed significantly more glucagon than the SB lines (p=0.0158), indicating that the NKX6.1-GFP lines may have differentiated into a more alpha cell phenotype (Figure 3.5.5). The results suggest that the p.C43G mutation had no influence on the expression levels of insulin or glucagon within my clones.

Overall, the differentiation efficiency was comparable between clones, albeit with insulin and glucagon expression being lower than expected at the end point of differentiation. In addition, the NKX6.1-GFP CRISPR sham had a higher proportion of C-peptide at stage 6 and 7 than the mutant and WT clone. However, the proportion of C-peptide was similar to that of the SB sham and SB p.C43G line, suggesting that the differences in proportions of C-peptide were a result of technical variation between differentiations.

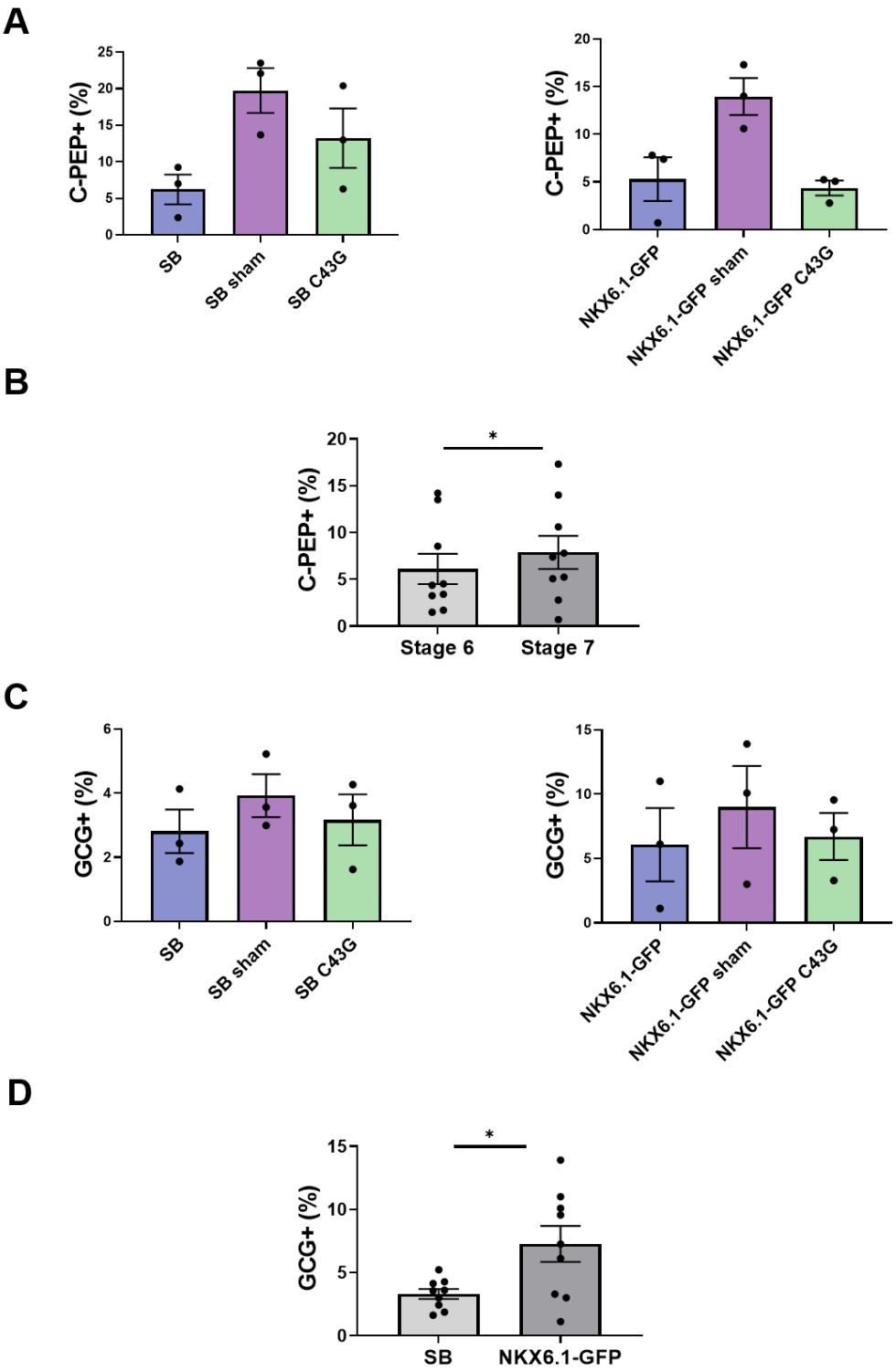


Figure 3.5.5 Analysis of expression of C-peptide and glucagon at stage 7
Analysis of differentiated *INS* clones via FACS examining expression of (A) C-peptide for the SB and NKX6.1-GFP lines, (B) C-peptide for the NKX6.1-GFP clones at stage 6 and 7, (C) glucagon for the SB and NKX6.1-GFP lines, and (D) comparing expression of glucagon between SB and NKX6.1-GFP lines. C-PEP=C-peptide; GCG=glucagon.
(A) and (C) SB lines n=3, NKX6.1-GFP lines n=3, (B) Stage 6 and 7, NKX6.1-GFP lines n=9, (D) SB lines n=9, NKX6.1-GFP lines n=9. All plots display mean±SEM.
Data analysed using one-way ANOVA for (A) and (C), and analysed using a two-tailed unpaired t-test for (B) and (D). *$p<0.05$

3.6 Analysing transcript levels of *INS* p.C43G hiPSC lines differentiated along the endocrine lineage

Having successfully differentiated both the SB and NKX6.1-GFP clones into BLCs, I next wanted to examine whether I could enhance the mature cell population by using FACS to sort the NKX6.1-GFP reporter line. I then assessed the transcript level of relevant markers for unsorted SB and sorted NKX6.1-GFP at both stages 6 and 7.

3.6.1 Separation of NKX6.1-GFP+ population using FACS

To facilitate purification of BLCs, the NKX6.1-GFP reporter line was sorted via FACS at stages 6 and 7. Previous experiments have demonstrated that NKX6.1 expression is accurately captured by GFP labelling (Gupta, et al., 2018). FACS allowed for successful separation of NKX6.1 positive cells with cells having a significantly higher expression of *NKX6.1* transcript than their GFP- counterparts at stage 6 (p=0.0012) and stage 7 (p=0.0028) (Figure 3.6.1).

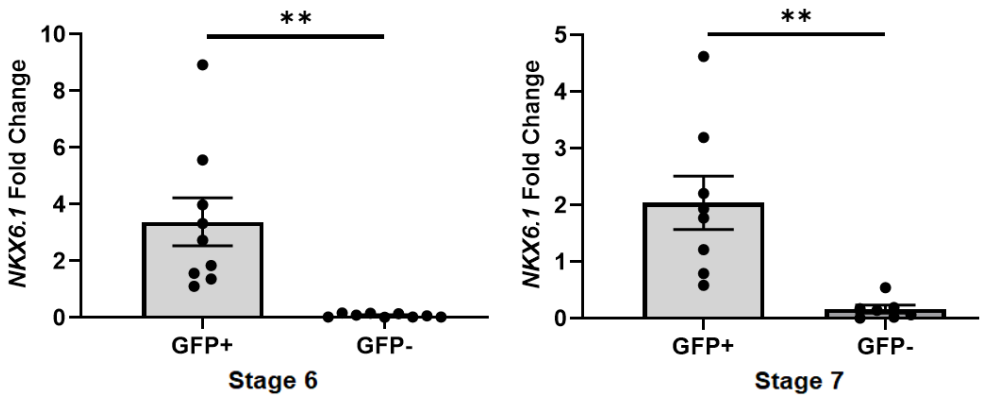


Figure 3.6.1 FACS sorting of NKX6.1-GFP cell lines by GFP expression at stage 6 and 7 with subsequent transcript analysis of *NKX6.1*
Cells were sorted by GFP fluorescent at stages 6 and 7 with transcript levels of *NKX6.1* then assessed using RT-qPCR. NKX6.1-GFP n=8-9, mean±SEM; **p<0.01 for two-tailed unpaired t-test.

3.6.2 No differences in expression of insulin at stage 6 or 7 between groups

In SB clones, there was no significant difference between groups in transcript levels of *INS* at stage 6 (p=0.294) or stage 7 (p=0.296) (Figure 3.6.2). For NKX6.1-GFP clones, separating by NKX6.1 expression via flow cytometry, and quantifying via RT-qPCR, showed that there was a significant increase in *INS* expression at stage 6 (p=0.0294) and stage 7 (p=0.0267) in the NKX6.1+ cells, indicating that sorting did enrich the population with insulin-producing beta-like cells (Figure 3.6.3). However, there was no difference in *INS* expression levels between groups after FACS at either stage 6 (p=0.280) or stage 7 (p=0.331) when measured using RT-qPCR (Figure 3.6.3). Assessment of insulin protein expression levels was hampered by the low differentiation efficiency of some samples, making visualisation a challenge, as some control samples did not exhibit insulin expression, meaning a comparison between groups could not be made (Figure 3.6.3). Due to the difficulties in obtaining enough cells from FACS, all samples for Western Blot were from unsorted cells.

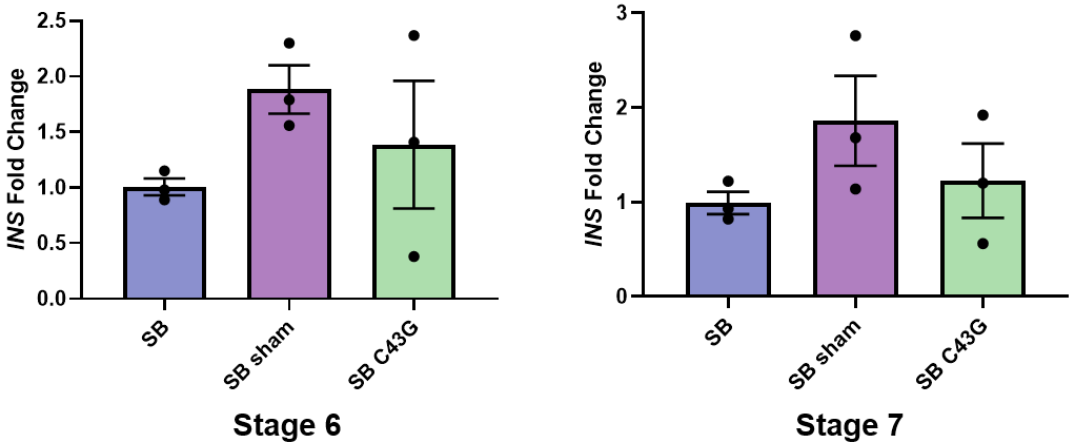


Figure 3.6.2 No significant differences in expression of *INS* at stage 6 or 7 between the SB clones
Evaluation of *INS* expression using RT-qPCR in SB clones at stage 6 and stage 7 of differentiation. SB, SB sham, SB C43G all n=3; plots display mean±SEM. Data analysed using one-way ANOVA

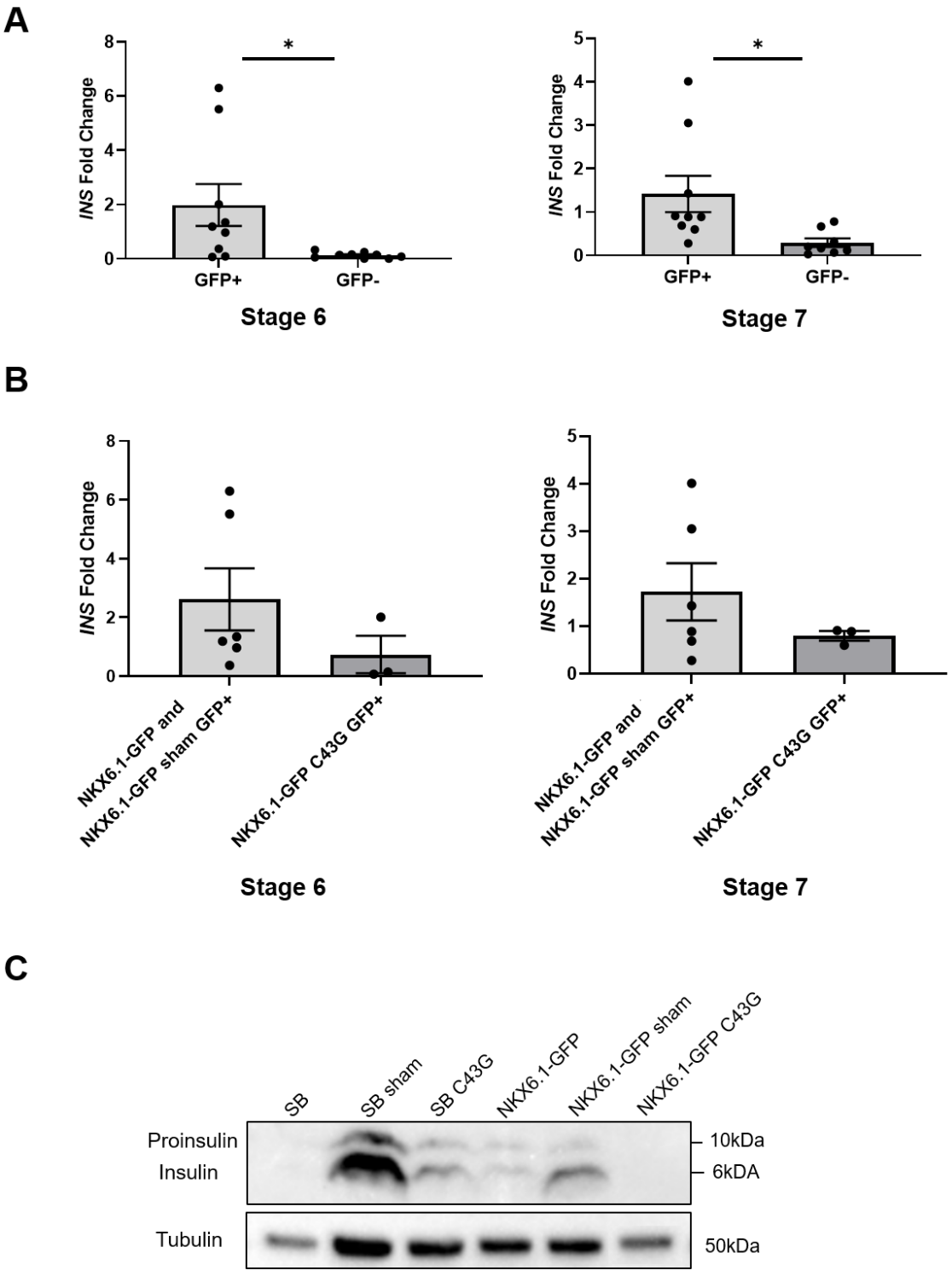


Figure 3.6.3 Separating by NKX6.1 expression allowed for enrichment of *INS* expressing cells but protein levels of insulin were difficult to visualise.

(A) Cells were sorted by GFP fluorescent at stage 6 and 7, with *INS* expression levels quantified via RT-qPCR**.** NKX6.1-GFP n=9, mean±SEM; *p<0.05 for unpaired t-test, (B) There was no difference in *INS* expression levels between the mutant and control lines after FACS sorting at either stage 6 or stage 7 when measured using RT-qPCR. NKX6.1-GFP lines all n=3, mean±SEM. Data analysed using two-tailed unpaired t-test, (C) Difficulties in visualisation of insulin protein levels. Representative blot of n=3.

3.6.3 No difference in levels of ER stress markers between groups

I looked at selected markers to assess whether p.C43G had an impact on ER stress. In the SB clones, there was no significant difference in the expression of either *HSPA5* or *DDIT3* at stage 6 (p=0.235 and p=0.846, respectively) or in *DDIT3* at stage 7 (p=0.635). For the NKX6.1-GFP clones, there was no difference in *DDIT3* expression at stage 7 (p=0.590) (Figure 3.6.4). Due to low amounts of RNA, it was not possible to test any further markers of ER stress. To assess levels of *DDIT3* at the protein level, expression of CHOP was examined at stage 6, with no difference between groups in either the SB line (p=0.618) or NKX6.1-GFP lines (p=0.258) (Figure 3.6.5). Due to technical difficulties, alongside low levels of protein, it was not possible to evaluate expression of CHOP at stage 7.

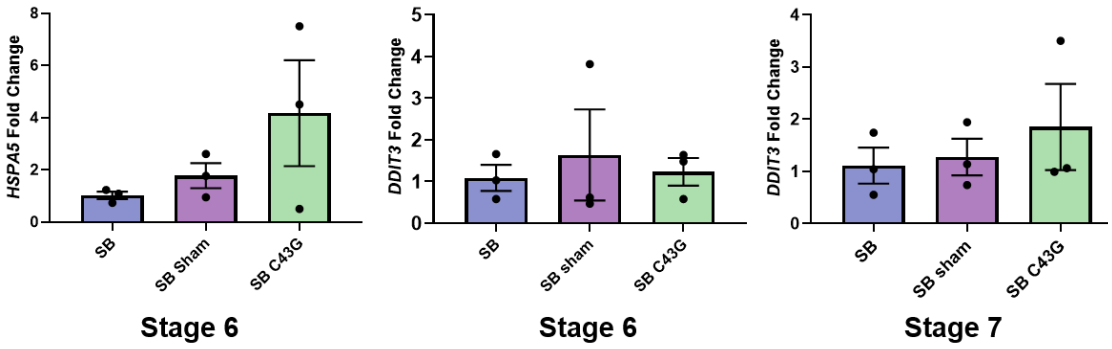


Figure 3.6.4 No significant differences in expression of selected markers of ER stress for SB clones at stage 6 and 7
Evaluation of *HSPA5* and *DDIT3* expression using RT-qPCR in SB clones at stage 6 (*HSPA5* and *DDIT3*) and stage 7 (*DDIT3* only) of differentiation. SB, SB sham, SB C43G, all n=3. Plots display mean±SEM. Data analysed using one-way ANOVA.

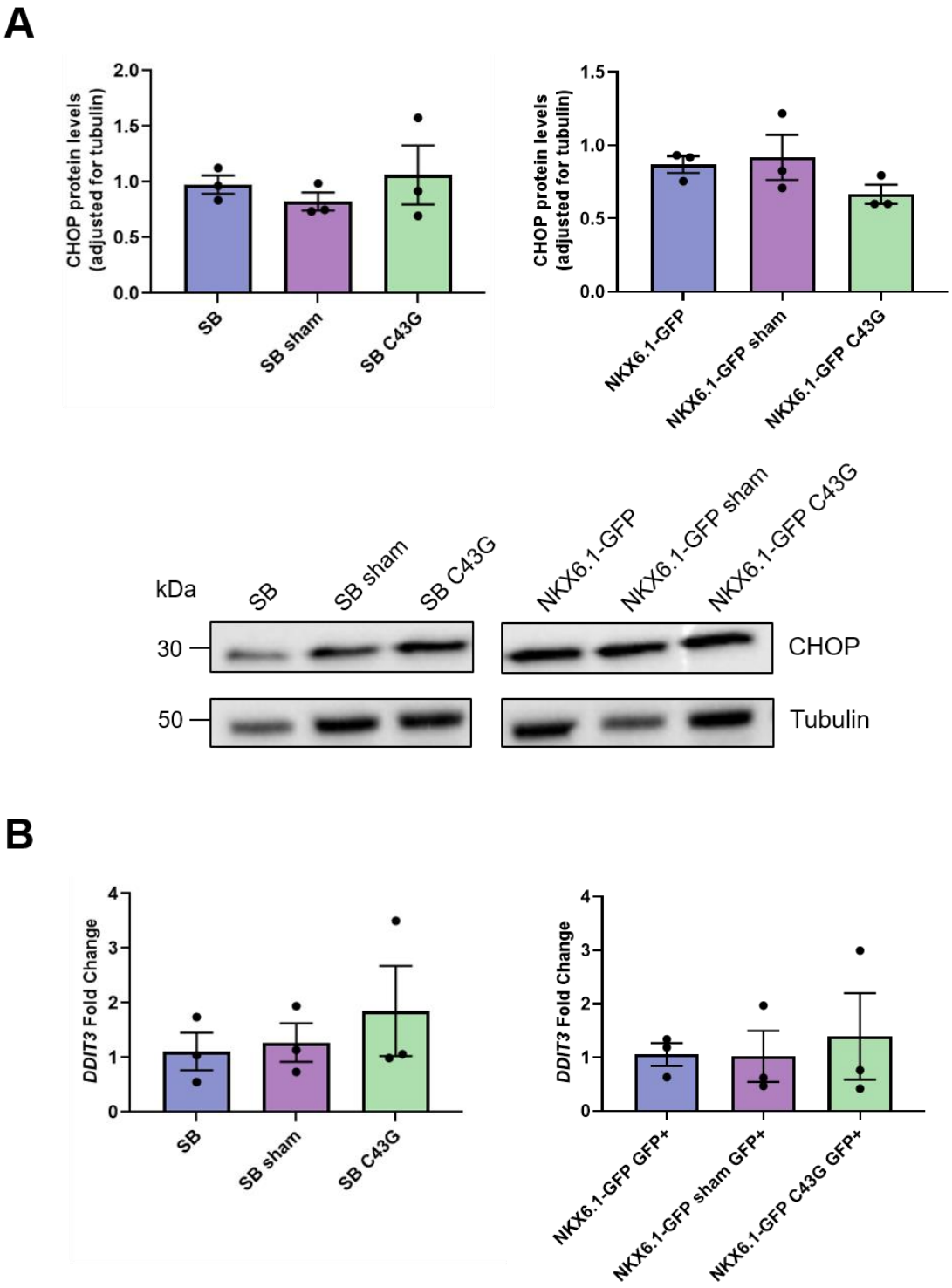


Figure 3.6.5 No significant differences in expression of *DDIT3* (CHOP) at stage 6 or 7
(A) Evaluation of CHOP expression in SB and NKX6.1-GFP clones at stage 6 of differentiation. Blot is representative of 3 independent experiments. (B) *DDIT3* expression in SB unsorted and NKX6.1-GFP FACS sorted GFP+ lines at stage 7 of differentiation, as quantified using RT-qPCR. SB lines all, n=3; NKX6.1-GFP lines all, n=3. Plots display mean±SEM. All data analysed using one-way ANOVA.

3.7 Discussion

This Chapter describes the different strategies employed to generate mutant *INS* hiPSC lines and successful generation and differentiation of the p.C43G line to BLCs. For all but the NKX6.1-GFP CRISPR sham, the levels of key development markers at stages 1, 4, 6 and 7 were comparable. The NKX6.1-GFP CRISPR sham had significantly higher expression of C-peptide at stages 6 and 7 of differentiation but the levels of C-peptide were comparable to the SB sham and SB p.C43G clones, indicating a high variability between clones and increasing the risk of evaluating an artefact rather than a real effect. Of note, the expression of the key development markers were below the expected level which may be indicative of inefficient differentiation.

There was no difference in the expression of *INS* or ER stress markers between groups

I saw no differences in the levels of *INS* expression or ER stress markers, *DDIT3* and *HSPA5*, in my mutant cells. Overexpression models of *INS* mutations, including p.C43G, that cause permanent neonatal diabetes found the mutations caused significantly increased levels of *DDIT3* (Rajan, et al., 2010). An iPSC model of a neonatal-diabetes causing *INS* mutation, with a similar mechanism of action to p.C43G, has shown that mutant *INS* expression can lead to upregulation of *INS* and other insulin secretion related genes (Balboa, et al., 2018). However, this change was not detectable using RT-qPCR, with the difference only revealed with single cell transcriptome analysis and comparing the beta cell populations in isolation. Although this model by Balboa et al. showed no significant increase in the expression of *DDIT3*, there was a significant upregulation of chaperone genes such as *HSPA5*. The increased expression of ER stress markers remained following transplantation of the *INS* mutant beta-like cells into

immunocompromised mice (Balboa, et al., 2018). An Akita mouse model similarly showed an upregulation of *INS* but also of *DDIT3*, with targeted disruption of the gene delaying diabetes onset (Oyadomari, et al., 2002). Most likely I could not detect a difference in ER stress due to the lack of functional maturity of my cells, as both 3D suspension culture and transplantation of differentiated cells into immunocompromised mice has been demonstrated to be required to produce more functionally mature cells which can secrete C-peptide in response to glucose stimulation (Pagliuca, et al., 2014, Rezania, et al., 2014, Balboa, et al., 2018).

To examine whether I could enhance the mature cell population I used FACS to sort the NKX6.1-GFP reporter line by GFP expression. Transcript analysis demonstrated that I had successful separation of NKX6.1 positive cells with cells having a significantly higher expression of *NKX6.1* transcript than their GFP- counterparts at stage 6 and stage 7. There was a significant increase in *INS* expression at stage 6 and stage 7 in the NKX6.1+ cells, indicating that sorting did enrich the population with insulin-producing beta-like cells. However, isolating a more mature beta cell population did not allow me to detect any differences in *INS* expression or ER stress between groups, further supporting the hypothesis that transplantation is required to allow proper evaluation of beta-cell function.

Limitations of investigating *INS* mutations in an hiPSC differentiation model

In recent years, hiPSC models have emerged as a popular method of modelling monogenic diabetes *in vitro* (Teo, et al., 2016, Balboa, et al., 2018, Cardenas-Diaz, et al., 2019, Ng, et al., 2019). Although these models have advanced our understanding of the pathogenesis behind diabetes development, there are still several limitations. Differentiation efficiency can vary greatly, not just between protocols but between

different hiPSC lines due to the heterogeneity of genetic backgrounds (Rouhani, et al., 2014, Kyttala, et al., 2016, Wesolowska-Andersen, et al., 2020). Although some of this variation can be overcome by examining mutations generated on an isogenic background, the same cell lines can still have variations in differentiation efficiency between experiments. In addition, potential off-target effects of the CRISPR/Cas9 editing may cause modification to genes involved in differentiation efficiency. This can make it difficult to disentangle whether the phenotypic effect is a result of the genotype or due to technical issues. My differentiations were performed on monolayer, which have the tendency to increase the proportion of polyhormonal cells that often display limited glucose-simulated insulin secretion (D'Amour, et al., 2006, Bruin, et al., 2014). Recent advances has seen improved differentiation efficiency through 3D-suspension culture (Toyoda, et al., 2015, Xu, et al., 2019). The implantation of these 3D-suspension *in vitro* cultures into immunocompromised mice enables generation of glucose-responsive insulin-secreting beta cells and can even ameliorate hyperglycaemia (Pagliuca, et al., 2014, Rezania, et al., 2014, Yabe, et al., 2017). Use of a 3D-suspension culture, alongside *in vivo* transplantation could provide a means to study *INS* mutations in a more functionally relevant model. Furthermore, the recent development of an INSULIN-H2B-Cherry reporter line enables FACS separation of insulin positive cells without impairment of insulin processing. Although the fluorescent protein is not localised with insulin, it can be viewed both via immunofluorescence and live cell imaging, which would allow me to visualise where misfolded protein aggregates (Blochinger, et al., 2020). Such an experimental technique, alongside more sensitive analysis tools such as single-cell RNA-Sequencing, would enable both enrichment of the population of interest alongside the ability to identify specific cell populations which could be obscured by bulk transcriptome analysis.

Conclusions

Differentiation of the *INS* p.C43G lines did not show an influence of p.C43G on selected ER stress markers or *INS* expression, even with enrichment of the mature cell population in the NKX6.1-GFP line. This is the first time that the p.C43G mutation has been modelled in an iPSC line and further characterisation of this mutation could be performed by utilising 3D-suspension cultures or transplantation of *in vitro* culture into immunocompromised mice. Enrichment of an insulin positive population could also be achieved through using an insulin reporter line.

Due to the functional immaturity and heterogeneity of beta-like cells, I wanted to see if I could investigate the effects of the p.C43G mutation in an already differentiated cellular system. In Chapter 4, I discuss the use of the mature human beta cell line EndoC-βH1, which has undergone CRISPR/Cas9 genome editing, alongside transient overexpression of a pCVM6_*INS*_C43G vector.

4

Characterisation of the p.C43G *INS* mutation in the EndoC-βH1 cell line

4.1 Introduction

Studying monogenic diabetes offers the chance to uncover how single genes can influence pancreatic development and function. The results from such investigations can help improve treatments, an example being patients with mutations in K_{ATP}-channels, where sulphonylureas can be used in place of insulin (Gloyn et al. 2004; Pearson et al. 2006). Findings from monogenic diabetes can also be translated to more complex forms of diabetes, as demonstrated in monogenic diabetes-causing genes such as *HNF1A* and *GCK*, which are also implicated in altered T2D risk (Fuchsberger, et al., 2016).

INS mutations have been shown to be cause of both MODY and permanent neonatal diabetes (Shoelson, et al., 1983, Stoy, et al., 2007). The majority of these mutations lead to misfolding of proinsulin, preventing its transport from the ER, resulting in ER stress and beta cell dysfunction (Wang, et al., 1999, Liu, et al., 2010, Cunningham, et al., 2019). However, the mechanism behind this has yet to be fully determined.

Rodent cellular models and transgenic mouse models have provided a means for exploring the effect of *INS* mutations (Wang, et al., 1999, Hartley, et al., 2010, Rajan, et al., 2010). These studies have demonstrated that mutations affecting the Cys residues on the A7-B7 or A20-B19 disulphide bonds trigger the dominant-negative blockade of proinsulin production from the wildtype (WT) *INS* allele. Both misfolded mutant and WT proinsulin can accumulate in the ER, activating the UPR which could lead to apoptosis of the beta cell (Leroux, et al., 2001, Liu, et al., 2005, Liu, et al., 2007, Hodish, et al., 2010, Liu, et al., 2010). However, there are several differences in both the architecture and physiology of the islet between mice and humans, as well as mice having two *INS* genes rather than one, which means that results from animal experiments on *INS* must be interpreted with caution (Hay and Docherty, 2006, Steiner, et al., 2010, Jennings, et al.,

2015). Therefore, given these dissimilarities, human beta cell lines are more appropriate to model *INS* mutations.

In Chapter 3, I described how an hiPSC-based model was employed to model a monogenic diabetes variant during human beta cell development. However, *in vitro* hiPSC-derived beta-like cells are invariably functionally immature, with even the best protocols only achieving around 60% insulin positive cells that commonly only elicit a robust glucose response post-transplantation into mice (Bruin, et al., 2014, Pagliuca, et al., 2014, Russ, et al., 2015, Vethe, et al., 2017). The lack of functional maturity makes modelling of monogenic diabetes variants particularly difficult, as phenotypes linked with insulin secretion may be lost due to the immaturity and polyhormonal nature of beta-like cells. Therefore, to investigate the functional effect of *INS* mutations, a cell line with a more mature phenotype is required. In this respect, the immortalised human cell line EndoC-βH1 provides a more physiologically relevant beta cell model, with the ability to secrete insulin in response to glucose stimulation (Ravassard, et al., 2011, Andersson, et al., 2015, Gurgul-Convey, et al., 2015, Hastoy, et al., 2018).

Experimental Aims

In Chapter 3, I demonstrated that the *INS* p.C43G mutation does not affect the levels of *INS* expression or selected ER stress markers at stage 6 or 7 of *in vitro* differentiation. To assess whether the expression of the p.C43G *INS* mutation altered mature beta cell function, I chose the human beta cell line EndoC-βH1 with a CRISPR/Cas9-mediated stable KO of *INS* recently generated by Dr Antje Grotz in our laboratory (Grotz, et al., 2019). This cell line, in addition to providing a human background, has the advantage of being able to study *INS* expression without the confounding effects of endogenous *INS* expression. I first aimed to confirm the KO of *INS* within the cells and then establish a

transfection system for investigating *INS* mutations. Having validated my model, I aimed to determine whether over-expression of the *INS* p.C43G mutations resulted in transcriptional upregulation of genes associated with the ER stress response.

Overall my aims were to:

- Confirm KO of the *INS* gene within the EndoC-βH1 *INS*-KO line
- Establish a transfection system for studying selected *INS* mutations in a human beta cell model
- Evaluate the re-expression level of the constructs within the *INS*-KO line
- Investigate whether the p.C43G mutant had a dominant negative effect on WT expression
- Investigate whether the p.C43G mutation led to defects in ER stress

In this Chapter, I present and discuss my results in comparison to those I observed in my iPSC derived model (Chapter 3).

4.2 Methods

4.2.1 Cloning of *INS* expression plasmids

Due to the difficulties in establishing a transfection system, several different plasmids were trialled. In the first instance, human *INS* cDNA (NM_000207.2) was provided as *Escherichia coli* (*E. Coli*) culture in LB broth with 8% glycerol and chloramphenicol antibiotic (Dharmacon). A small sample of this pDNR-LIB vector stock was transformed and amplified using DH5α competent cells (Life Technologies) and DNA was extracted using PureYield™ Plasmid Mini or Midiprep kits (Promega). The plasmid was verified for the correct *INS* sequence via Sanger sequencing (Table 4.2.1). The cDNA *INS* sequence was amplified using overhanging PCR primers (Eurofins) which included recognition sites for the restriction enzymes EcoRI and NotI, allowing digestion of the *INS* cDNA fragment and subsequent ligation into the pCMV6_Entry vector to create a wildtype *INS* pCVM6 plasmid (designated pCMV6_*INS*_WT_Myc) with an in-frame C-terminal Myc-FLAG tag (Table 4.2.1) (Figure 4.2.1).

Table 4.2.1 Primer pairs for Sanger sequencing and PCR

Primer	Sequence (5' to 3')
T7	TAATACGACTCACTATAGG
M13R	CAGGAAACAGCTATGAC
pDNR_LIB	GGTGAGAATCCAAGCACTAGTC
Forward PCR sequence (pCMV6_*INS*_WT_Myc)	GGAATTCATGGCCCTGTGGATGCGCCT
Reverse PCR sequence (pCMV6_*INS*_WT_Myc)	ATAAGAATGCGGCCGCCCATAGTTGCAGTAGT TCTCCAGC
hProCpepGFP	CTCAAGCTTGGAATTCGCGG
Reverse PCR sequence (pCMV6_*INS*_WT)	ATAAGAATGCGGCCGCCCCTAGTTGCAGTAGT TCTCCAGC

The hProCpepGFP plasmid was procured from Peter Arvan's group (Liu et al. 2007). DNA was then extracted using the PureYield™ Midiprep kit (Promega) and sent for Sanger sequencing to ensure retrieval of the plasmid had been successful (Table 4.2.1).

I introduced a stop codon to the pCVM6_*INS*_WT_Myc plasmid to prevent expression of the C-terminal Myc-FLAG tag by amplifying the *INS* sequence using overhanging PCR primers (Eurofins) (Table 4.2.1). The newly introduced stop codon was confirmed via Sanger sequencing and the PCR product purified using a QIAquick® PCR Purification kit (Qiagen). The purified product underwent a blunting reaction, was vortexed briefly and incubated at 70°C for 5min before being chilled on ice and ligated using the CloneJET PCR Cloning Kit (Thermo Fisher Scientific). The ligation mixture was transformed into competent DH5α cells, purified and sent for sequencing confirmation (Table 4.2.1). The purified plasmid was digested using the restriction enzymes EcoRI and NotI with the resultant product run on an agarose gel. The correct 342bp PCR product size was excised from the gel and purified using a QIAquick® Gel Extraction Kit (Qiagen), then ligated into the pCMV6_Entry plasmid and transformed into DH5α cells. DNA was extracted using PureYield™ Plasmid Mini or Midiprep kits (Promega). The plasmid was verified for the correct *INS* sequence via Sanger sequencing (Table 4.2.1).

The mutation corresponding to c.127 T>G, p.C43G was introduced via site-directed mutagenesis (SDM) using the QuickChange II Site-Directed Mutagenesis Kit (Agilent Biotechnologies) according to the manufacturer's guidelines. SDM primers comprised of C43G_F (5'-CTCTCTACCTAGTGGGCGGGGAACGAGGC-3') and C43G_R (5'-GCCTCGTTCCCCGCCCACTAGGTAGAGAG-3'). The plasmid pCMV6_Entry vector (pCMV6_EV) acted as a control. The pCMV6 plasmid DNA was transformed and amplified in DH5α competent cells and DNA was extracted using PureYield™ Plasmid Mini- or Midiprep kits following the manufacturer's instructions (Promega). Correct

insertion of *INS* and subsequent SDM was verified by Sanger sequencing (Eurofins Genomics, Germany) (Table 4.2.1). Due to the troubleshooting required for this Chapter, several plasmids were used, with vector maps for all plasmids given in Figure 4.2.1.

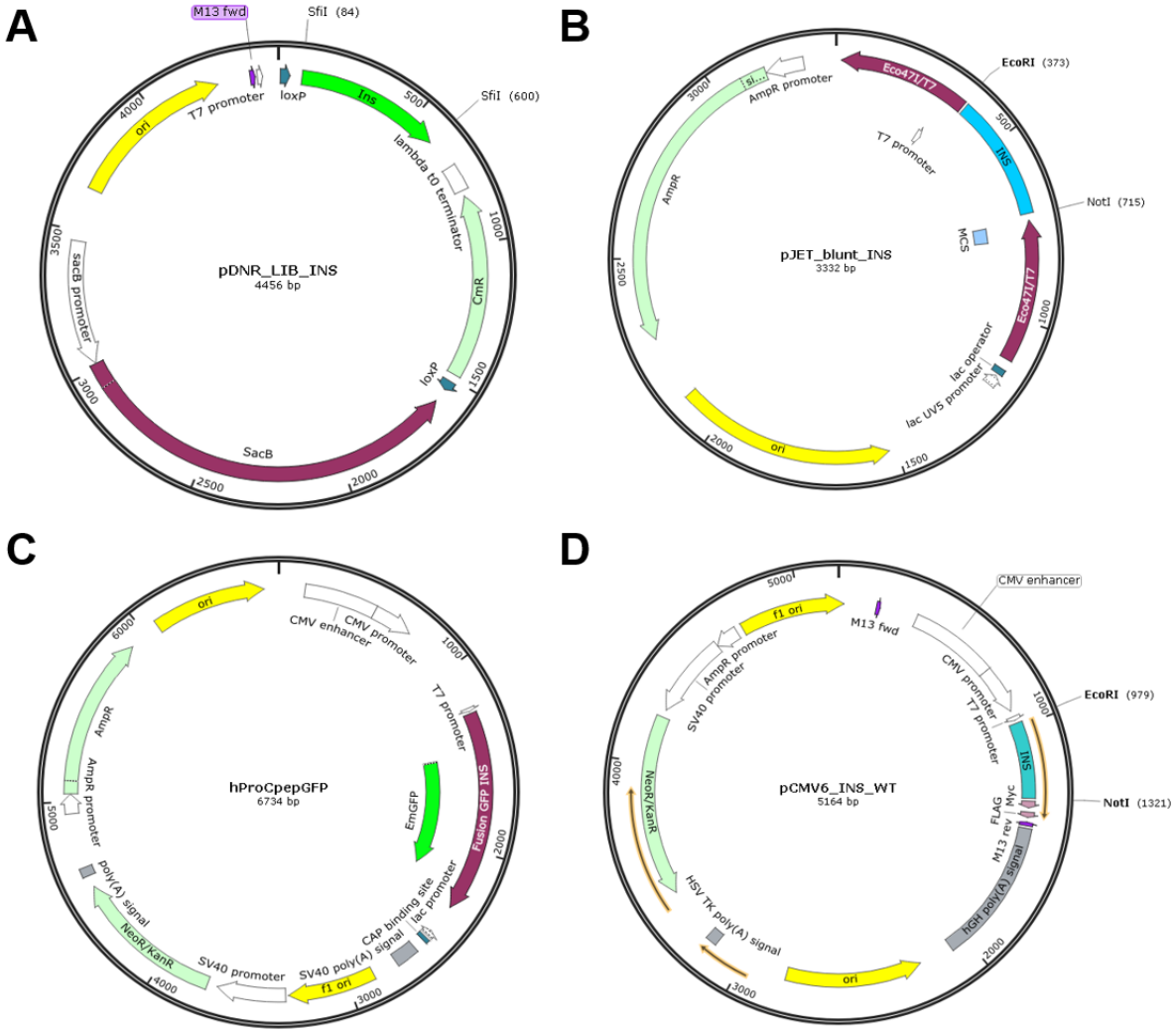


Figure 4.2.1 Vector maps for the human cDNA clones for (A) pDNR_LIB_*INS*, (B) pJET_blunt_*INS* (C) hProCpepGFP and (D) pCMV6_*INS*_WT_Myc

Maps including the cloning sites (A) SfiI, (B) and (D) EcoRI and NotI

4.2.1 Transient transfections

Transient transfections were carried out as described in Section 2.6.2. Tunicamycin (5μg/ml) (Sigma Aldrich) in DMSO was included as a positive control to induce ER stress (Tsonkova, et al., 2018). Transfected cells were assessed for RNA and protein expression 48h post-transfection.

4.2.2 Quantitative real-time PCR (RT-qPCR)

For quantification of gene expression, RT-qPCR was performed as outlined in Section 2.2.3. The specific TaqMan® probes used for gene expression analysis in this Chapter are detailed in Table 4.2.2. Data was analysed using the ddCT method, whereby relative gene expression of a test sample is compared to a control sample on the same plate.

Table 4.2.2 TaqMan® gene expression assays used for real-time PCR and their target amplification regions. All assays purchased from Thermo Fisher Scientific.

Target genes	Assay number	Target region
DDIT3	Hs99999172_m1	Spans exon boundary
INS	Hs00355773_m1	Spans exon boundary
TBP	Hs00427620_m1	Spans exon boundary
XBP1	Hs03929085_g1	Spans exon boundary

4.2.3 Immunofluorescence staining

For immunostaining of cells, EndoC-βH1 were plated on pre-coated coverslips (Thermo Fisher Scientific) and fixed 48hr post-transfection with 4% paraformaldehyde (Thermo Fisher Scientific) in PBS for 20min at RT. Cells were washed with 3xPBS for 5min each before being permeabilised using 0.1% Triton X-100 (Sigma Aldrich) in PBS for 15min. Cells were then incubated in 5% goat serum in PBS for 30min at RT to block non-specific binding. Following this, primary antibodies were added, either at 4°C overnight

or for 2hr at RT. The antibodies used and their dilutions are listed in Table 4.2.3. Afterward, cells were washed with 3xPBS and subsequently incubated with secondary antibodies for 1hr at RT in the dark (Table 4.2.4). Coverslips were washed again and then incubated in Red dot2, a far-red nuclear stain (Biotium, 40061) at 1:200 in goat serum for 10min at RT. Cells were further washed 3x5min in PBS and rinsed with distilled water before being mounted using Vectashield mounting medium (Vector Laboratories, H-1200). Coverslips were sealed using commercially available nail polish. Slides were stored at 4°C, protected from light prior to imaging.

Table 4.2.3. Primary antibodies used for immunostaining Antibodies were diluted in 5% goat serum and added dropwise to coverslips

Antibody	Source	Dilution
Anti-Chromogranin A (rabbit polyclonal)	Abcam, ab15160	1:100
Anti-insulin (guinea pig polyclonal)	Dako, A0564	1:750
Red dot2	Biotium, 40061	1:200

Table 4.2.4. Secondary antibodies used for immunostaining Antibodies were diluted in 5% goat serum and added dropwise to coverslips

Antibody	Source	Dilution
Anti-mouse Alexa Fluor 488	Invitrogen, A11029	1:250
Anti-guinea pig Texas Red	Vector labs, TI-7000	1:300

4.2.4 Confocal microscopy

A Bio-Rad Radiance 2100 confocal microscope was used for visualisation of the immunostained cells, using the 60x 1.0 N.A. water immersion objective. Images were acquired consecutively for three channels (green, red and far-red) using LaserSharp 2000 software. Laser settings were optimised for each sample and channel and hence fluorescent intensities have not been compared across samples. The LSM Image Browser 4.2 (Carl Zeiss) was used to export image files and processed with ImageJ 2.1.0 (NIH).

4.2.5 Western Blotting

Western Blotting was performed as outlined in Section 2.3.5. Details of the primary antibodies used in this Chapter are outlined in Table 4.2.5. Secondary antibodies used have previously been summarised in Table 2.3.2.

Table 4.2.5 Primary antibodies used for WB. Antibodies were diluted in 5% milk/PBST.

Antibody	Source	Dilution
Anti-GFP (mouse monoclonal)	Santa Cruz, sc-9996	1:1000
Anti-insulin (mouse monoclonal)	Santa Cruz, sc-377071	1:1000
Anti-β tubulin (mouse monoclonal)	Santa Cruz, sc-365791	1:2000

4.2.6 Generation of an *INS*-KO EndoC-βH1 cell line

The stable *INS*-KO EndoC-βH1 cell line was generated using a pooled lentiviral CRISPR/Cas9 approach by Dr Antje Grotz (Gloyn lab, OCDEM, Oxford). To summarise, single guide RNAs (sgRNAs) targeting *INS* exon 2 were taken from the TKO Library v3 (Hart, et al., 2017) and cloned into the lentiviral vector plentiCRISPRv2, as outlined previously (Shalem, et al., 2015). Successful sgRNA integration into the vector was confirmed by Sanger sequencing. Subsequent lentivirus production, was performed by Dr Antje Grotz, with assistance by Elena Navarro Guerrero at the Target Discovery Institute. To generate a stable *INS*-KO line, EndoC-βH1 cells were transduced at a multiplicity of infection (MOI) of 8 with constructs containing two sgRNAs targeting *INS* (exon 2) and the Cas9 enzyme (Table 4.2.6). Transduced cells were selected for by treating with puromycin (3µg/µl, Sigma Aldrich) for 7 days. After selection, I maintained cells in EndoC-βH1 culture media and passaged weekly. Media was supplemented with puromycin (3µg/µl) every four to six passages to ensure propagation of the transduced

cells. A detailed guide of the CRISPR/Cas9 genome editing pipeline has been published previously (Grotz, et al., 2019).

Table 4.2.6. Single guide RNA (sgRNA) sequences used in the generation of *INS*-KO EndoC-βH1 cells by CRISPR-Cas9 genome editing. Generation of the cell line was performed by Dr Antje Grotz (Grotz, et al., 2019).

sgRNA	Target exon	Sequence (5' to 3')
1	2	CACAATGCCACGCTTCTGCA
2	2	CATCTGCTCCCTCTACCAGC

4.2.7 Transcriptome profiling

For transcriptome profiling, RNA was extracted and processed from EndoC-βH1 cells following transfection of: pCMV6_GFP, pCVM6_*INS*_WT, pCMV6_*INS*_C43G or tunicamycin (5μg/ml) 48h post-transfection as described in Section 2.5.4 and Section 2.4.1. The software package DESeq2 in R version 3.6.1 was used to analyse the data (Love, et al., 2014). Gene ontology and pathway enrichment analysis was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8 (https://david.ncifcrf.gov/) (Huang, et al., 2007). Differentially expressed genes were submitted as a custom gene list and *Homo sapiens* was selected as background species. The following key DAVID categories with significant enrichment were selected for analysis: Online Mendelian Inheritance in Man (OMIM) disease, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway, and gene ontology (GO) terms for biological process, cellular component and molecular function. Results which returned p-values less than 0.05 were considered statistically significant. Individual analysis methods are described in more detail in Section 2.4.

4.2.8 Statistical Analyses

Statistical analyses were performed in GraphPad Prism 8 (v8.3) with results from multiple experiments expressed as mean ±SEM. For two unmatched groups, the p-value was determined using a two-tailed unpaired t-test. Comparison of groups larger than two was done with a one-way ANOVA followed by Sidak's multiple comparisons test. A p-value of <0.05 was considered statistically significant, with n representing the number of biologically independent experiments. Further details of statistical testing are outlined in the relevant figure legends.

4.3 Results

4.3.1 Characterisation of the *INS*-KO EndoC-βH1 cell line

As already described, the *INS*-KO cell line was generated by Dr Antje Grotz and characterised by her prior to me performing further downstream experiments. Transcript analysis showed that there was a trend towards reduced *INS* expression (p=0.057). Although the transcript expression was not significantly decreased, Western blot analysis showed that there was complete loss of insulin protein. In addition, insulin was not detected when insulin content was assessed using a sensitive AlphaLISA-based method (p<0.0001) (Grotz, et al., 2019) (Figure 4.3.1).

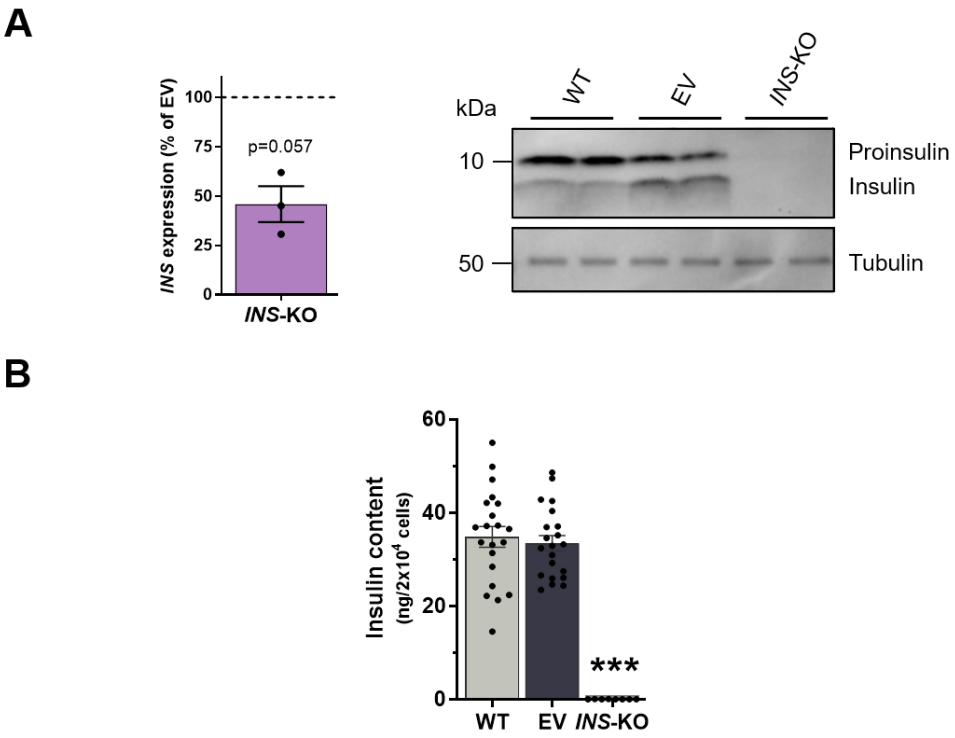


Figure 4.3.1 Characterisation of the *INS*-KO line
(A) Reduced *INS* transcript expression and complete loss of insulin protein in the *INS*-KO cell line. (B) Insulin content assessed in *INS*-KO in comparison with WT and EV EndoC-βH1 cell lines using AlphaLISA. ***p<0.0001 vs EV. Data analysed using (A) one sample t-test and (B) one-way ANOVA with Sidak's multiple comparison test. EV, empty-vector; KO, knockout. Data from RT-qPCR is the mean ± SEM from three independent experiments with fold changes log transformed for statistical analysis. Experiments and data analysis performed by Dr Antje Grotz (OCDEM, Oxford) (Grotz et al. 2019).

4.3.1.1 Immunofluorescence reveals that *INS*-KO cells do not express insulin, with no alteration in the marker of the trafficking pathway

Prior to establishing my transfection system and given the residual *INS* transcriptomic signal, I first wanted to ensure that any insulin signal detected via immunofluorescence was from my constructs alone. I also wanted to see if the knockout of *INS* in the *INS*-KO line had led to any obvious differences in the appearance or distribution of the secretory vesicles and hence I also stained for Chromogranin A, a marker of the trafficking pathway in neuroendocrine cells. Initial staining, against insulin and Chromogranin A, was performed separately to ensure there was no cross-reactivity. Immunofluorescence confirmed the initial characterisation of the *INS*-KO line by Western blot, with no expression of insulin within the *INS*-KO cells. Chromogranin A expression was similar to EndoC-βH1 cells, indicating that loss of *INS* had not caused any obvious defects in the distribution or localisation of the secretory vesicles. (Figure 4.3.2). These results confirmed the *INS*-KO line as a suitable cellular system for examining the effect of overexpressing *INS* constructs, as expression and processing of the *INS* constructs should not be impaired by aberrant cellular machinery.

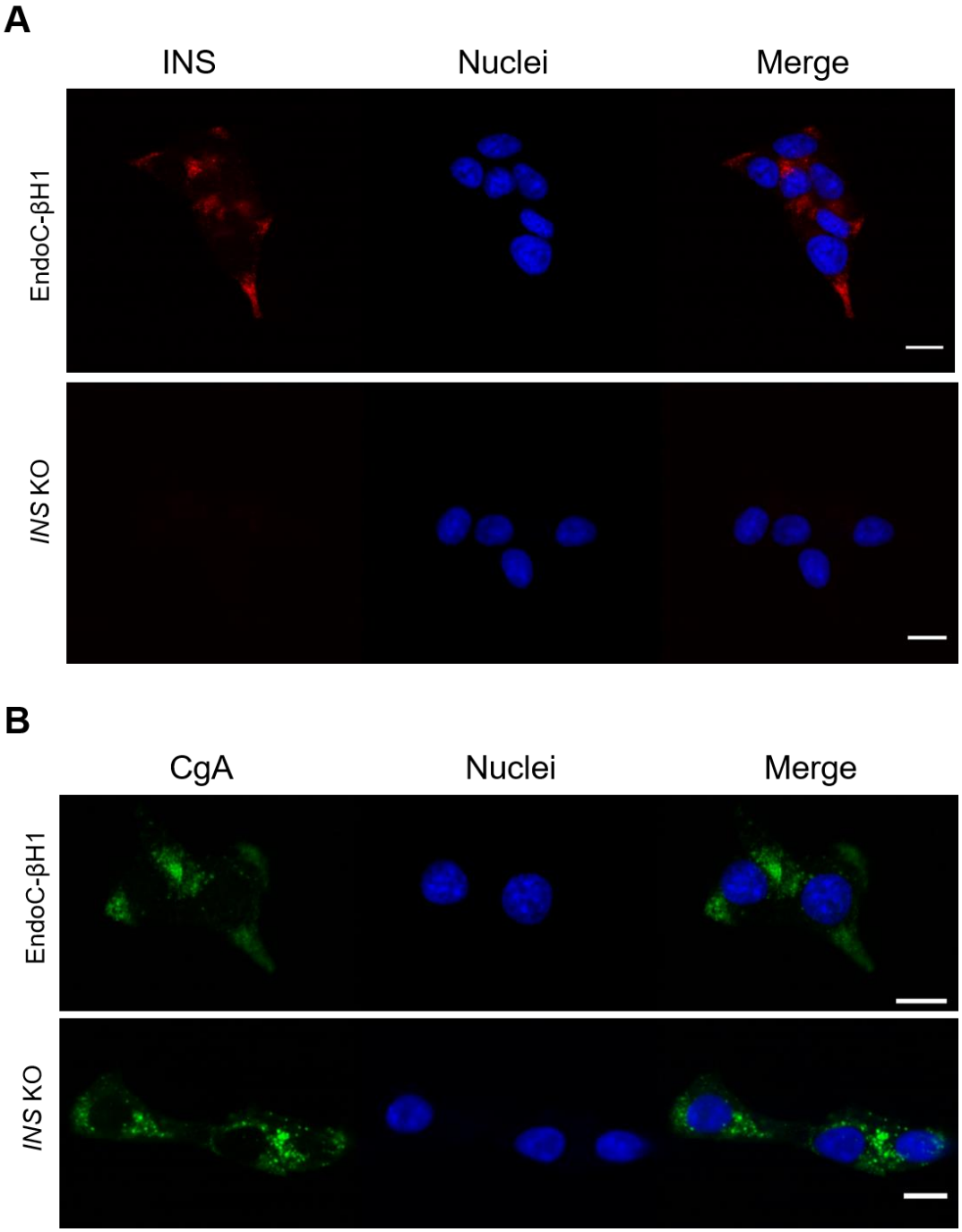


Figure 4.3.2 *INS*-KO cells do not express insulin, with no detectable differences in Chromogranin A expression
Fixed EndoC-βH1 and *INS*-KO cells were immunostained with (A) insulin (red) and (B) Chromogranin A (green) with reddot2 (blue) used as a nuclear marker. Images were visualised using BioRad Radiance 2100 confocal microscope and processed using ImageJ. Scale bar is 10μm

4.3.2 Establishing a re-expression system in the *INS*-KO cell line

Alongside the *INS*-KO line, HEK293 cells were utilised, due to both their lack of endogenous insulin expression and their high transfection efficiency. Both HEK293 and *INS*-KO cells were transfected with pCMV6_*INS*_WT_Myc or pCMV6_GFP, as a transfection control. HEK293 cells showed efficient transfection and expression of both pCMV6_*INS*_WT_Myc and pCMV6_GFP, with a dose-dependent effect. Whilst *INS*-KO cells could express GFP, there was no signal for insulin when assessed via Western Blot, with insulin expression in the EndoC-βH1 line validating that the antibody worked (Figure 4.3.3). The expression of GFP demonstrates that the *INS*-KO cell line is amenable to transfection but that the ability to detect expression of insulin at the protein level could require additional optimisation or that the processing of proinsulin was impaired. Transfection was undertaken with different incubation times (24hr-72hr) and using both FuGENE® and Lipofectamine® 3000 based transfection methods. Different ratios of transfection reagent to plasmid were also trialled with no effect on expression. Additional troubleshooting was undertaken, with transfection commenced both simultaneously to cell plating and 24hr later. Due to the low transfection efficiency of EndoC-βH1 cells, a higher quantity of cells were transfected to increase the amount of protein collected to ensure potential low levels of insulin could be detected. Each Western blot was visualised for up to 6min to ensure any weak signal was able to be visualised. The difficulties in transfecting the pCMV6_*INS*_WT_Myc construct could be due to the Myc-FLAG tag present at the end of the *INS* sequence. The Myc-FLAG tag could disrupt the folding of the proinsulin molecule, preventing exit from the ER. Previous work, where a firefly luciferase was fused with the C-terminus of proinsulin demonstrated that correct targeting to the secretory pathway was a relatively rare event and may have only been possible as a result of aggregates being largely composed of native proinsulin, allowing correct

packaging and exit to the Golgi (Pouli, et al., 1998). As my *INS*-KO cells do not contain any native proinsulin, it would suggest that my constructs would be unable to form aggregates that could exit the ER. Due to being unable to successfully transfect my construct into my *INS*-KO cells, I needed an alternative approach.

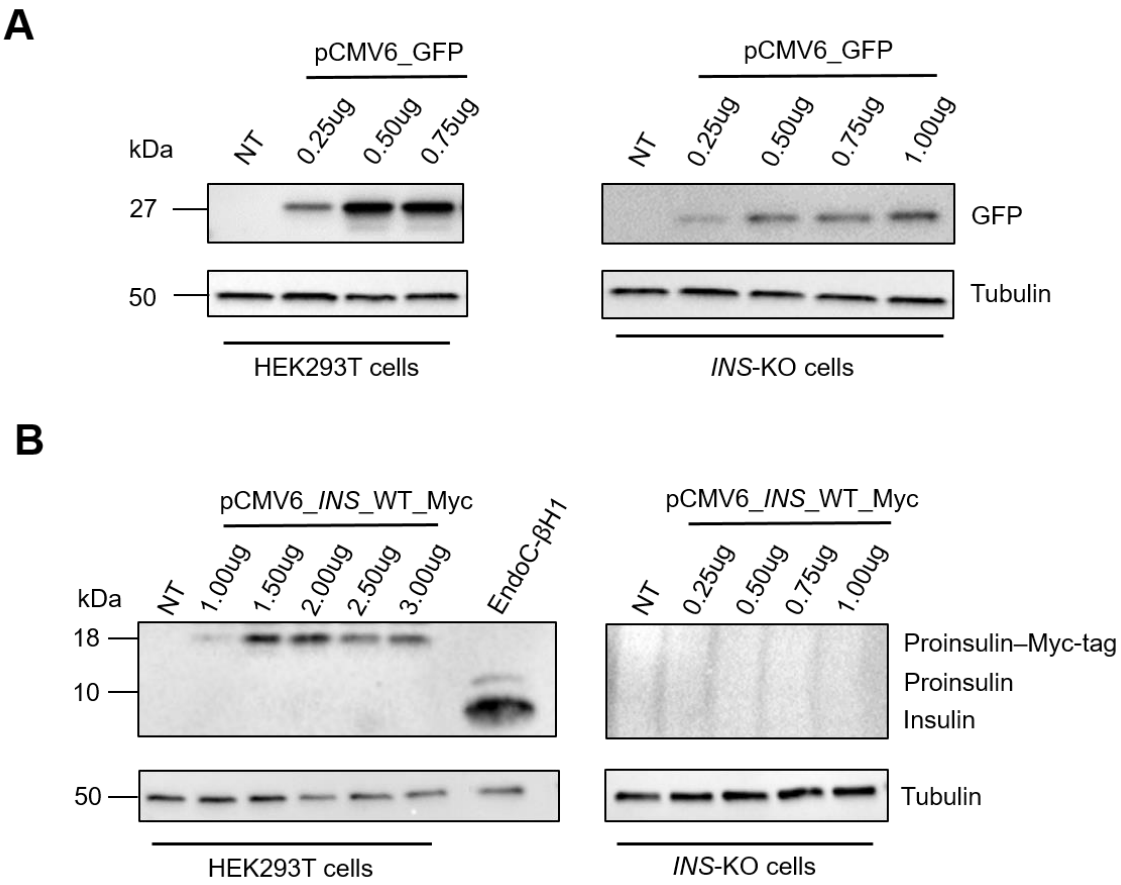


Figure 4.3.3 *INS*-KO cells can be successfully transfected with pCMV6_GFP but not pCVM6_*INS*_WT_Myc
Western Blot analysis of transfection of HEK293 and *INS*-KO cells with (A) pCMV6_GFP and (B) pCVM6_*INS*_WT_Myc. The EndoC-βH1 cell line was used as a positive control for insulin expression. Blots were stained for GFP (A) and insulin (B) with tubulin as a loading control. Blots are representative of n=3

4.3.3 Establishing a transfection system of *INS*-KO cells with a previously published plasmid

Due to the difficulties in transfecting *INS*-KO cells believed to be due to the C-terminal position of the Myc-FLAG tag, I tried a published plasmid which had the fluorescent protein GFP fused within the C-peptide region. The plasmid, hProCpepGFP, procured from Dr Peter Arvan's laboratory (University of Michigan), has previously been used successfully for *INS* overexpression experiments in INS-1 832/13 rat cells (Liu, et al., 2007).

4.3.4 Transfection of hProCpepGFP shows incomplete processing of insulin

EndoC-βH1 and *INS*-KO cells showed efficient transfection and expression of pCMV6_GFP via Western Blot. Whilst *INS*-KO cells could successfully express GFP, there was only a singular band detected which was the equivalent size of proinsulin-GFP for cells transfected with hProCpepGFP. There was no band of equivalent size to C-peptide-GFP, indicating that proinsulin had not been successfully cleaved into insulin and C-peptide (Figure 4.3.4). This could be due to GFP being much larger in size than proinsulin which may inhibit the folding process of proinsulin. Given the unsuccessful expression of both pCMV6_*INS*_WT_Myc and hProCpepGFP, a new approach was sought.

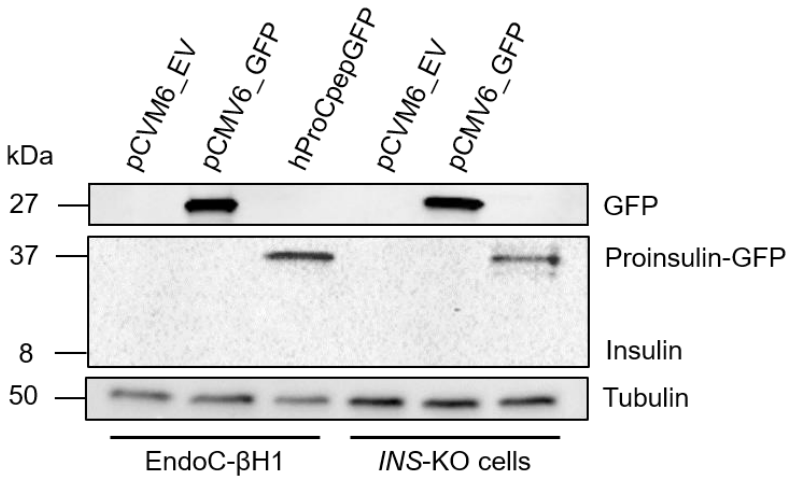


Figure 4.3.4 Transfection of hProCpepGFP shows incomplete processing of insulin via Western Blot
Western Blot analysis of transfection of EndoC-βH1 and *INS*-KO cells with pCMV6_GFP and hProCpepGFP. There was a singular band which was equivalent to proinsulin-GFP. Tubulin was used as a loading control. Representative blot of n=3

4.3.5 New cloning strategy to re-introduce the stop codon to the *INS* sequence and prevent expression of the Myc-FLAG tag in the pCVM6_*INS*_WT_Myc plasmid

As a further strategy to transfect *INS*-KO cells, the pCMV6_*INS*_WT_Myc plasmid was edited to introduce a stop codon into the *INS* sequence and thereby prevent expression of the Myc-FLAG tag. Successful integration of the stop codon and subsequent introduction of the p.C43G mutation via SDM was confirmed by Sanger sequencing (Section 4.2.1). Given that *INS*-KO cells are insulin protein free, the expression of any tag is not necessary and could corrupt proinsulin folding.

4.3.6 Transfected cells demonstrated increased *INS* expression as measured by RT-qPCR but insulin protein was unable to be detected via Western Blot

To assess whether *INS*-KO cells could be successfully transfected with the new constructs, both RNA and protein were collected for further analysis. Previous characterisation of the *INS*-KO demonstrated that the *INS* transcript was not significantly

reduced (Grotz, et al., 2019). The RT-qPCR analysis showed a significant increase of *INS* expression in the cells transfected with the wildtype (WT) alone and with the wildtype and mutant transfected together ($p=0.0413$ and $p=0.0054$, respectively). Transfection with the mutant alone also showed a significantly increased *INS* expression ($p=0.0121$). Insulin was unable to be detected via Western Blot, probably due to low levels of transfection efficiency (Figure 4.3.5). The significantly increased *INS* expression suggests that the constructs were able to be transfected into the *INS*-KO cells. Due to the inability to detect insulin protein using Western Blot, I decided to use immunofluorescence to determine if there were any differences in the localisation and distribution of insulin between the WT and mutant constructs within the successfully transfected *INS*-KO cells.

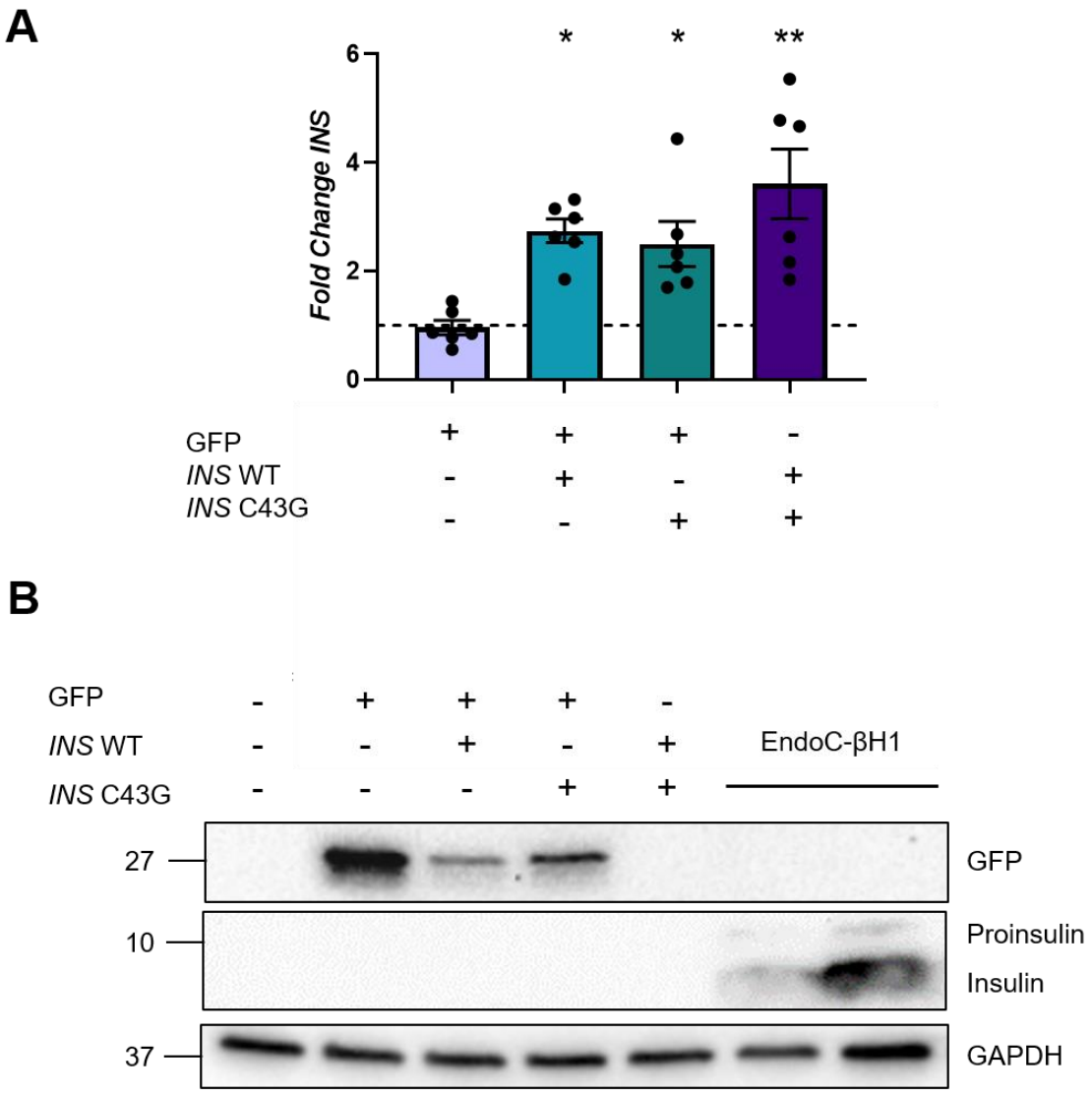


Figure 4.3.5 *INS* expression can be measured by RT-qPCR but not via Western Blot in transfected *INS*-KO cells.
(A) Evaluation of *INS* expression in transfected *INS*-KO cells via RT-qPCR. Expression levels are normalised to untransfected *INS*-KO cells (dotted line). n=6, mean±SEM. Data analysed using two-tailed unpaired t-test; *p<0.05, **p<0.01 in comparison to cells transfected with pCMV6_GFP which served as a transfection control.
(B) Evaluation of GFP and insulin expression via Western Blot. Representative blot of n=3. GFP=pCVM6_GFP; *INS* WT= pCMV6_*INS*_WT; *INS* C43G= pCMV6_*INS*_C43G

4.3.7 Immunofluorescence reveals successful transfection but a low level of co-transfection

The initial idea was to mimic the heterozygous cellular state of the patient by co-transfecting the wildtype and mutant constructs and assess cross-effects of p.C43G expression over WT and vice versa. Therefore, in order to both establish the transfection conditions and evaluate the level of co-transfection, I paired each construct with GFP. I transfected equal amounts of pCMV6_GFP and either pCMV6_*INS*_WT or pCVM6_*INS*_C43G into the *INS*-KO line. The pCMV6_GFP plasmid served as a transfection control and allowed for visualisation of co-transfected cells. Immunofluorescence revealed that although cells could be transfected successfully with both the WT and mutant construct separately, very few cells demonstrated co-transfection with both GFP and either the WT or mutant plasmid (Figure 4.3.6). It could be that cells that receive two plasmids are put under an increased amount of stress and undergo apoptosis. Although only a qualitative analysis, it was clear that there was a low transfection efficiency, with very few cells demonstrating any fluorescence for either insulin or GFP. My results indicated that the cellular system required major improvements in order to be used to study the heterozygous phenotype. Therefore, downstream experiments were performed with the transfection of either the WT or mutant alone.

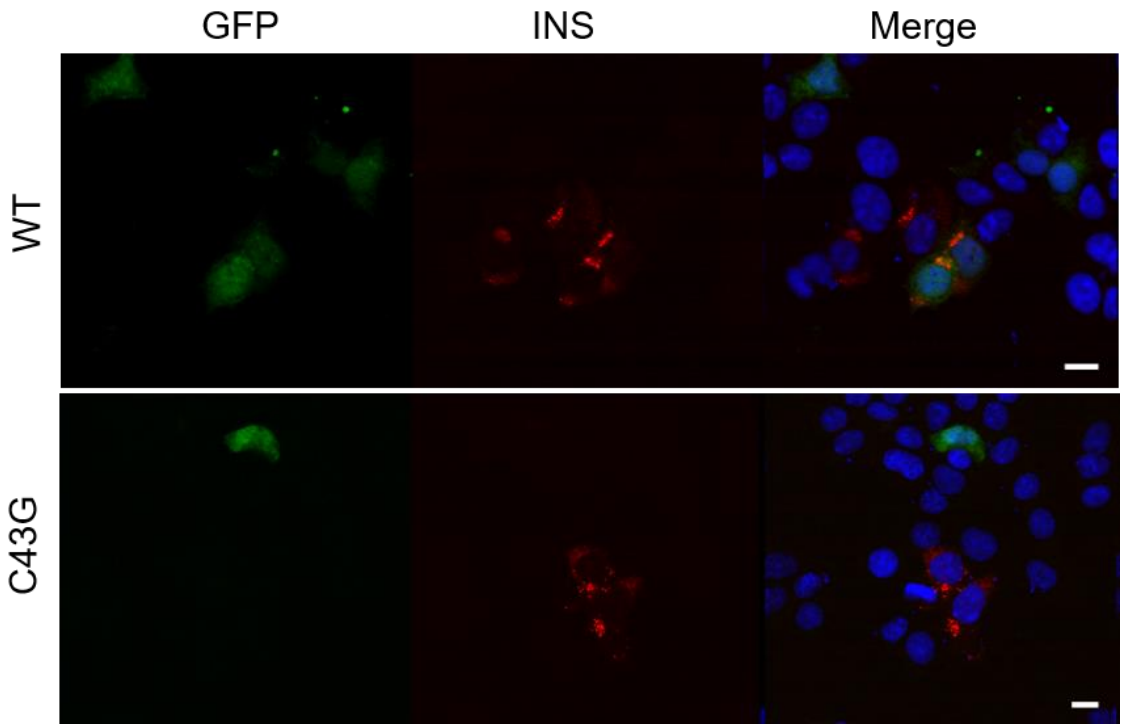


Figure 4.3.6 Immunofluorescence reveals very limited co-transfection and a low level of transfection
Fixed *INS*-KO cells were transfected with either pCMV6_*INS*_WT or pCMV6_*INS*_C43G and double immunostained with insulin (red) and GFP (green) and red dot2 (blue) applied as nuclear marker. Images were visualised using a BioRad Radiance 2100 confocal microscope and processed using ImageJ. Scale bar=10μm

4.3.8 Wildtype and mutant *INS* are both localised to the secretory vesicles

Previous work with transgenic mouse models, in addition to overexpression studies in HEK293T and rodent cell lines, have demonstrated that *INS* mutations can lead to misfolding of proinsulin and subsequent retention in the ER (Hodish, et al., 2010, Liu, et al., 2010, Rajan, et al., 2010, Sun, et al., 2020). I wanted to test whether this was the case within my cellular system. To do this, I double-stained my cells for both Chromogranin A (CgA), a marker of secretory vesicles, and insulin. I saw a similar staining pattern for CgA across my control EndoC-βH1 cells and for the *INS*-KO cell transfected with either the WT or mutant constructs, with co-localisation of insulin and CgA (Figure 4.3.7). This suggests that the mutant is able to be successfully packaged and targeted to the secretory

pathway. However, the mutant construct may still exert excess pressure on the cellular machinery that cannot be captured via immunofluorescence.

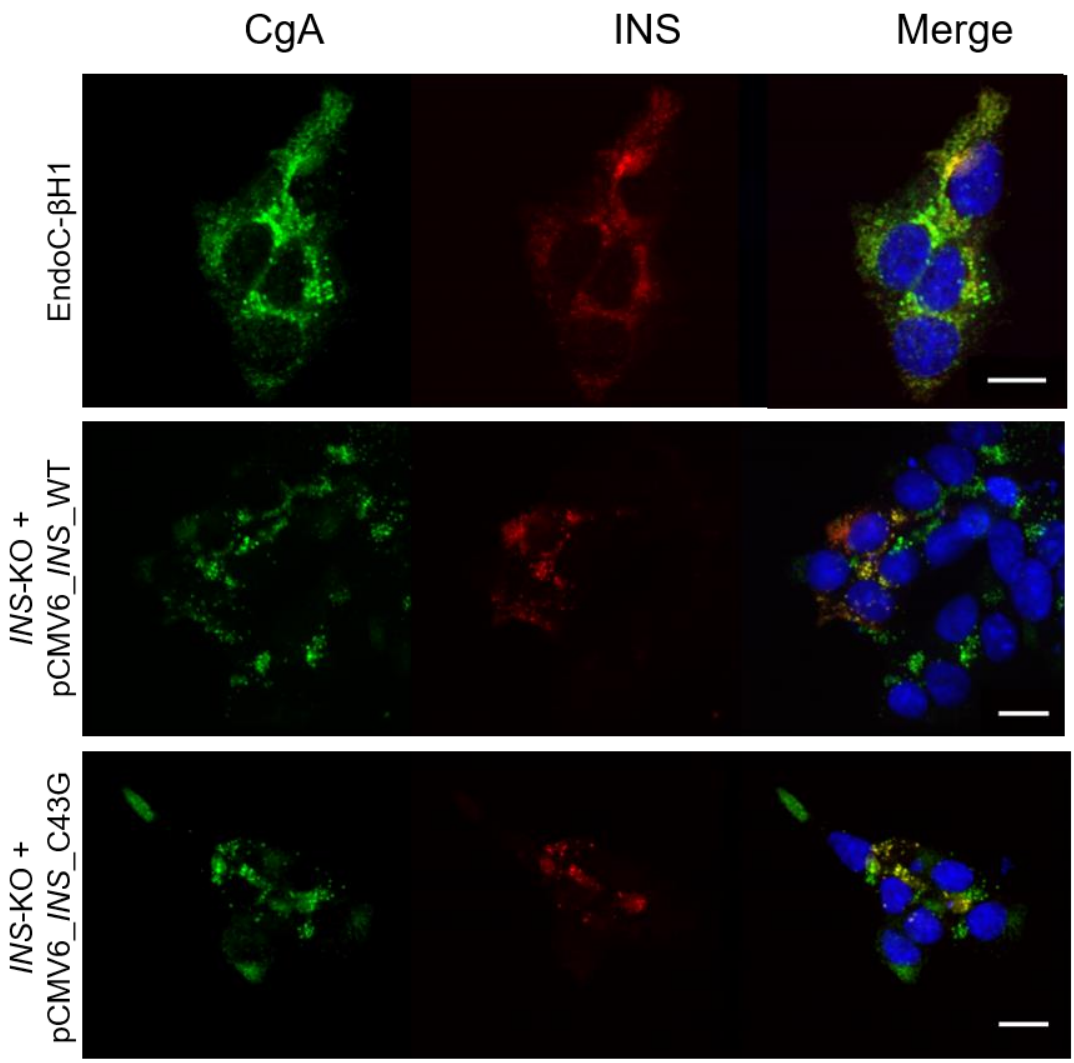


Figure 4.3.7 Wildtype and mutant *INS* are both localised to the secretory vesicles
Fixed EndoC-βH1 and *INS*-KO cells were double stained with Chromogranin A (green) and insulin (red), with red dot2 (blue) used as a nuclear marker. Cells were imaged with a BioRad Radiance 2100 confocal microscope. Scale bars indicate 10μm

4.3.9 Transfection of the wildtype or mutant construct does not induce ER stress when evaluated by RT-qPCR.

I looked at selected markers to assess whether the p.C43G mutation had an impact on ER stress, with ER stress inducer tunicamycin included as a positive control (Tsonkova, et al., 2018). There was no significant difference in the expression of either *DDIT3* or *XBP1* for any of the constructs (p=0.233-0.968), with levels of these markers upregulated as expected in response to tunicamycin treatment (p=0.0007 and p=0.0024, respectively) (Figure 4.3.8). It could be that there was no detectable differences in ER stress levels due to the low transfection efficiency.

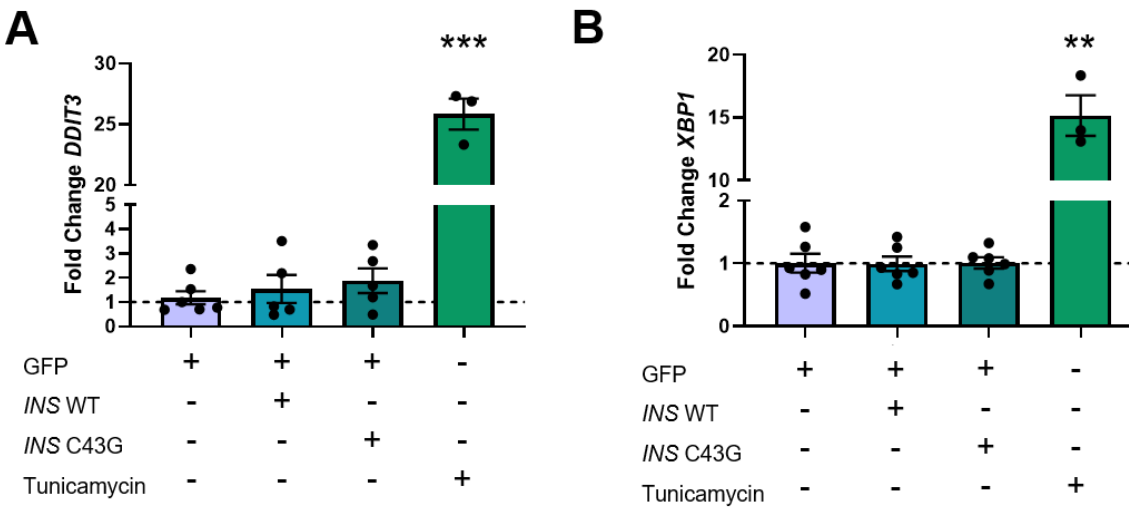


Figure 4.3.8 Transfection of the wildtype or mutant construct does not induce ER stress when evaluated by RT-qPCR.
Evaluation of the expression of ER stress markers (A) *DDIT3* and (B) *XBP1* in *INS*-KO cells transfected with different constructs. Expression levels are normalised to untransfected *INS*-KO cells (dotted line). n=3-6; GFP=pCVM6_GFP; *INS* WT= pCMV6_*INS*_WT; *INS* C43G= pCMV6_*INS*_C43G mean±SEM. Data analysed using two-tailed unpaired t-test. **p<0.01 ***p<0.001

4.3.10 Principal component analysis revealed a separation of samples treated with tunicamycin

There are several cellular mechanisms that co-ordinate together to alleviate ER stress. As RT-qPCR can only focus on a few selected markers, I wanted to ensure that I was not missing any potential targets and therefore utilised RNA-Seq to give a broader spectrum of analysis. Overall, 15 RNA-Seq libraries were sequenced to a mean read depth of 55.73 ± 3.75 million reads per sample. Principal component analysis (PCA) of the transcriptome of the samples showed no obvious groupings, although there was a separation of the samples treated with tunicamycin from the others (Figure 4.3.9). In case the samples treated with tunicamycin were skewing the results, these samples were removed and the PCA was performed again. As before, there was no obvious clustering of the samples, which may be demonstrative of the low transfection efficiency (Figure 4.3.9). Analysis of *INS* transcript levels revealed that there was a trend towards increased *INS* expression in cells transfected pCVM6_*INS*_WT or pCMV6_*INS*_C43G when compared with cells transfected with pCMV6_GFP (p=0.186 and p=0.115, respectively). There was also a trend towards increased *INS* levels when cells transfected pCVM6_*INS*_WT and pCMV6_*INS*_C43G were compared with *INS*-KO, however this did not reach statistical significance (p=0.243 and p=0.167, respectively) (Figure 4.3.10). The results help support the hypothesis from the RT-qPCR that the constructs were able to be transfected into the *INS*-KO cells.

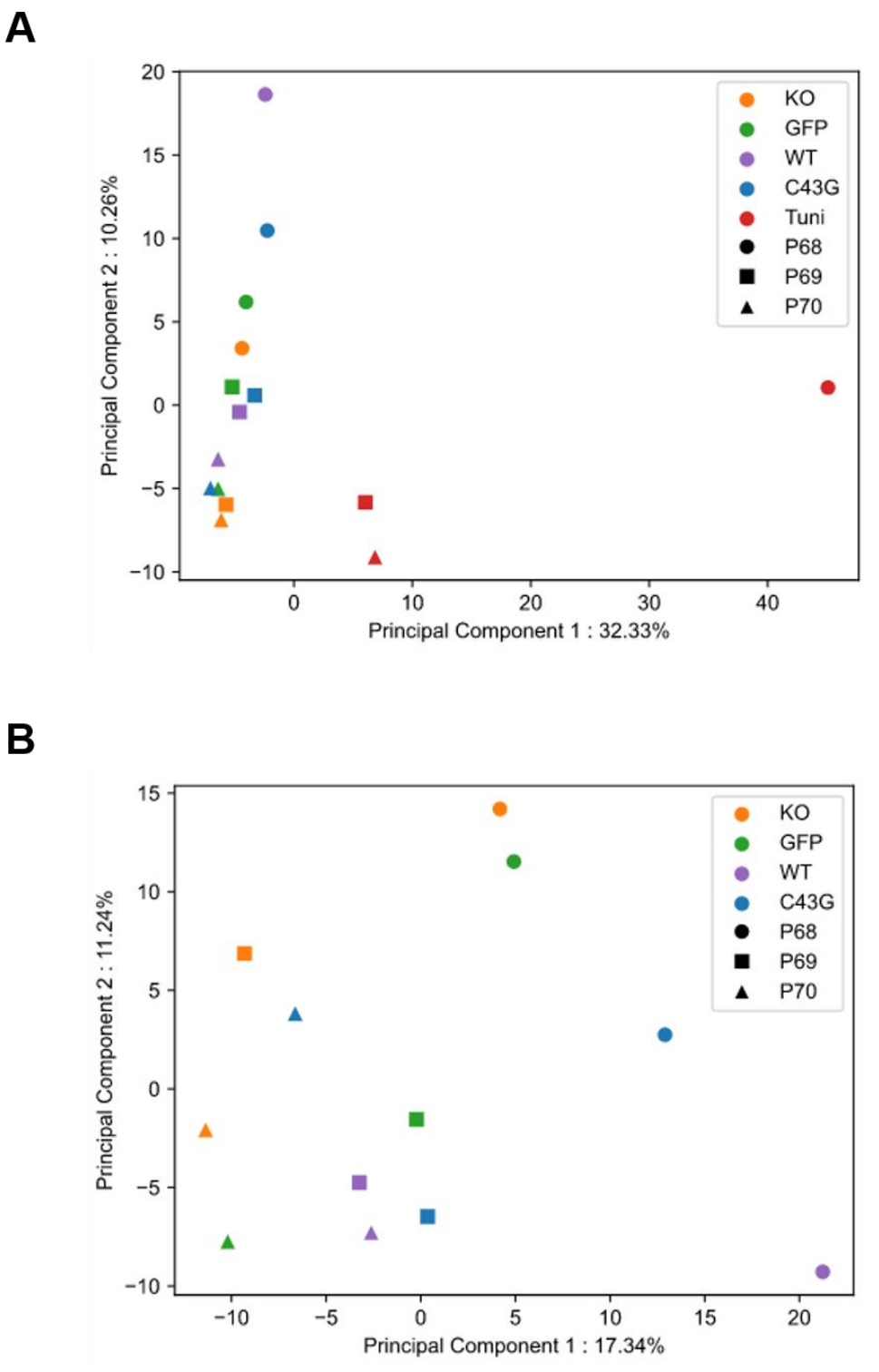


Figure 4.3.9 Principal component analysis revealed a separation of samples treated with tunicamycin

(A) The first two principal components (PC1, PC2) were calculated using normalised gene counts of the samples, with samples treated with tunicamycin separated from the other samples. (B) The first two principal components (PC1, PC2) once samples treated with tunicamycin were removed. n=3 for each condition; KO=*INS*-KO; GFP=pCVM6_GFP; WT=pCMV6_*INS*_WT; C43G=pCMV6_*INS*_C43G; P68-70 refers to the three different passage numbers of the cells

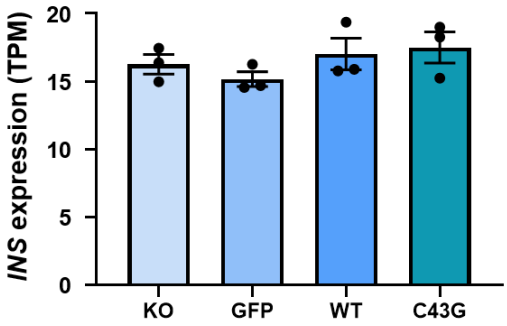


Figure 4.3.10 There was a trend towards increased expression of *INS* in cells transfected with either the wildtype or p.C43G construct
Expression of *INS* was compared between groups with a trend towards increased *INS* expression for cells transfected with the WT and p.C43G constructs. n=3 for all groups, plots display mean±SEM. TPM=transcripts per million, GFP=pCMV6_GFP, WT=pCMV6_*INS*_WT, C43G=pCVM6_*INS*_C43G. Data analysed using two-tailed unpaired t-test for each construct in comparisons to both pCMV6_GFP transfected cells and untransfected (*INS*-KO) cells.

4.3.10.1 Differential gene expression (DEG) analysis

The Bioconductor DESeq2 package (see Section 2.5.6 and Section 4.2.6) was used to identify differentially expressed genes between groups. The DEGs between groups were tested for enrichment in gene ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways using a hypergeometric test. In the first instance, I compared *INS*-KO cells transfected with pCMV6_GFP to untransfected *INS*-KO cell, as previous work has shown that transcriptional effects may be partly associated with the introduction of foreign DNA, rather than from the expression of a specific gene (Jacobsen, et al., 2009). However, my analysis revealed only 24 DEGs, with no enrichment in relevant physiological pathways, demonstrating that transfection of pCMV6_GFP did not appear to adversely affect the cells. To ensure that I could capture any effects on ER stress, I also included a positive control and compared my cells treated with tunicamycin to *INS*-KO cells. Analysis revealed that there was an enrichment of genes involved in the unfolded protein response (padj=1.14×10^{-12}), response to endoplasmic reticulum stress (padj=1.03×10^{-4}) and the ER-nucleus signalling pathway (padj=1.86×10^{-3}) (Figure

4.3.11). I then compared the effect of transfecting either pCMV6_*INS*_WT or pCMV6_*INS*_C43G to cells transfected with pCMV6_GFP. For cells transfected with pCMV6_*INS*_WT, analysis revealed 466 DEGs with enrichment of genes involved in the AMPK signalling pathway (padj=$6.5x10^{-5}$) and the insulin signalling pathway (padj=0.0087) (Figure 4.3.11). Comparing pCMV6_*INS*_C43G to cells transfected with pCMV6_GFP revealed 186 DEGs, of which there was again an enrichment of genes involved in the AMPK signalling pathway (padj=$1.27x10^{-4}$), as well as in the regulation of insulin secretion (padj=0.0014) (Figure 4.3.11). Previous work has suggested that activation of the AMPK pathway may alleviate ER stress, suggesting that my transfected cells may be eliciting a response to prevent ER stress (Simon-Szabo, et al., 2014). For both the wildtype and mutant construct, the enrichment of insulin-related pathways could illustrate there is some rescue of the loss of endogenous insulin expression. Comparing pCMV6_*INS*_C43G and pCMV6_*INS*_WT revealed no DEGs, reflecting the lack of clustering as seen in the PCA. This suggests that several effects of the transfected constructs were largely masked by the large proportion of untransfected cells. Overall, although I could see some enrichment of pathways involved in insulin secretion, I also saw enrichment of many unrelated terms suggesting that bulk analysis is not sufficiently sensitive to pick up all the effects that result from transfection. Further to this, RNA-Seq is only able to detect effects that are associated with up or down regulation of genes but there are other mechanisms that regulate ER stress, such as phosphorylation that cannot be capture using this method.

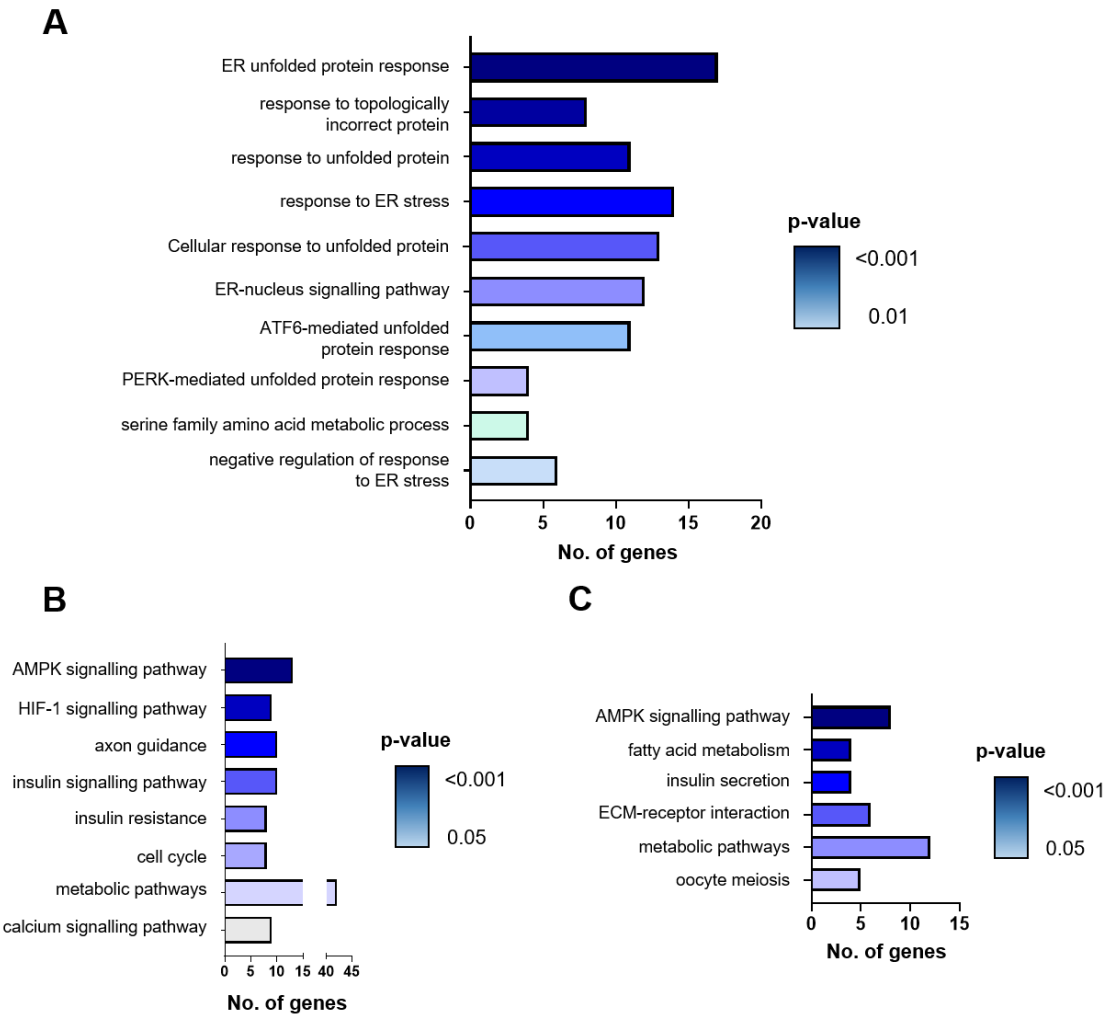


Figure 4.3.11 Pathway enrichment analysis of DEGs
Pathway enrichment analysis using DAVID 6.8 based on the list of DEGs for each condition and analysed for KEGG pathways. (A) Tunicamycin treated cells compared to *INS*-KO (untransfected) cells, (B) Cell transfected with pCMV6_*INS*_WT compared to cells transfected with pCMV6_GFP, (C) Cell transfected with pCMV6_*INS*_C43G compared to cells transfected with pCMV6_GFP. Selected pathways are displayed, ranked and colour coded by p-value.

4.4 Discussion

In Chapter 3, I was unable to demonstrate that the p.C43G mutation had any influence on ER stress when evaluated in an hiPSC system. The functional immaturity and polyhormonal nature of the differentiated iPSCs could mask effects that may be seen in a more functionally mature setting. Further to this, previous investigations into *INS* variants have been in mouse models and rodent cell lines, which have several differences from their human counterparts (Hay and Docherty, 2006, Steiner, et al., 2010, Jennings, et al., 2015). With these limitations in mind, I wanted to utilise a more physiologically relevant human model and hence chose the EndoC-βH1 cell line to examine the effect of *INS* mutations on ER stress. I avoided the possible confounding effects of endogenous *INS* expression by employing the *INS*-KO EndoC-βH1 line. In setting up a transfection system, I was met with several troubleshooting issues that needed to be addressed to confirm successful transfection of my constructs. In the first instance, the position of the Myc-FLAG tag at the C-terminal sequence of *INS* most likely caused the misfolding of the proinsulin molecule within the RER, preventing targeting to the secretory pathway (Pouli, et al., 1998, Wright, et al., 2013). Similarly, the hProCpepGFP plasmid contained a fluorescent marker larger than the proinsulin molecule which could have prevented proinsulin folding properly. This led to another approach, where the stop codon of *INS* was reintroduced to prevent expression of the Myc-FLAG tag. Although I was able to demonstrate a significantly increased *INS* expression via RT-qPCR, I was still unable to detect insulin at a protein level via Western Blot, probably due to the low transfection efficiency of the cells. Therefore, I decided to utilise immunofluorescence, where I was able to successfully detect cells transfected with either the WT or mutant construct. However, my transfection showed only a small proportion of co-transfected cells which prevented recapitulating the human genotype within my cellular system. I also showed

that the localisation and distribution of the WT and mutant constructs were similar in the cells, indicating that both were able to be successfully packaged and targeted to the secretory pathway. Previous work has demonstrated that misfolding of mutant proinsulin leads to formation of high molecular weight aggregates of WT/mutant proinsulin which prevent export of WT proinsulin from the ER (Cunningham, et al., 2017). My results could support the hypothesis that it is the interactions between WT and mutant proinsulin that cause impaired processing, rather than mutant-mutant interactions. It could also be that there are other deleterious collateral effects of the p.C43G mutation on the trafficking pathway that I was unable to capture. As with my hiPSCs system, I did not detect any difference in expression of ER stress markers via RT-qPCR or RNA-Seq, although I did see an upregulation of the AMPK pathway, suggesting that the cells may be producing a response to relieve ER stress. It could be that the cellular system undergoes ER stress from the process of transfection, as previous work has demonstrated transcriptional effects may be partly associated with the introduction of foreign DNA, rather than from the expression of a specific gene (Jacobsen, et al., 2009). More likely, the inability to detect changes in ER stress were due to very few cells being successfully transfected, which was also reflected in the lack of clustering seen from my RNA-Seq analysis. A future study could test electroporation as a method to increase efficiency. Another option would be to employ a lentiviral system, which would enable the study of constitutive mutant *INS* expression rather than transient expression. However, constitutive expression is more prone to compensation effects which might induce different phenotypic effects (Rossi, et al., 2015). Further experiments could also look at if there were any differences in caspase-3 activity, which can be used as a marker of apoptosis activation (Porter and Janicke, 1999). These results could then be used in combination with an optimised hiPSC

protocol, which would give the opportunity both to validate results within each system but also to examine different elements of the way in which *INS* mutants exert its effects.

This is the first study to utilise a mature human beta cell line for overexpressing *INS* mutations. With optimisation, primarily with increasing transfection efficiency, this system could be used to test a number of *INS* mutations incorporating different clinical phenotypes.

5

Characterisation of *HNF1A* knockout beta cell models

5.1 Introduction

Heterozygous mutations in *HNF1A* are the most frequent cause of MODY, yet the full mechanism that causes HNF1A-MODY is still not fully understood (Yamagata, et al., 1996, Stride, et al., 2005, Shields, et al., 2010). Mouse models have previously been utilised to unravel the underlying mechanisms of HNF1A-MODY, however mice with heterozygous mutations do not recapitulate the diabetic phenotype seen in humans (Pontoglio, et al., 1998, Boj, et al., 2010). *HNF1A* KO mice or those expressing dominant-negative *HNF1A* mutations can exhibit hyperglycaemia as a result of inadequate pancreas formation, impaired insulin secretion and beta cell dysfunction due to apoptosis and dysregulation of genes such as glucose transporter GLUT2 (*SLC2A2*) and those involved in mitochondrial metabolism (Wang, et al., 1998, Wobser, et al., 2002, Bonner, et al., 2010). *HNF1A* has also been shown to have a role in regulating both insulin transcription and genes involved in beta cell proliferation (Wang, et al., 1998, Yang, et al., 2002). However, differences between human and rodent islets, such as in development, architecture, transcriptional regulation and function, mean that there may be human-specific functions for *HNF1A* that could be missed by focussing on mouse models (Cabrera, et al., 2006, MacDonald, et al., 2011, McCulloch, et al., 2011, Benner, et al., 2014, Baron, et al., 2016). The use of hiPSC and gene editing technologies has enabled the development of human *in vitro* models of *HNF1A* mutations, with a recent study confirming *HNF1A*'s role in insulin secretion, with loss of the gene also leading to increased expression of alpha cell genes (Cardenas-Diaz, et al., 2019). In addition, pathway analysis of human islets expressing a disease-causing *HNF1A* variant demonstrated that *HNF1A* affected the regulatory processes required for glucose-stimulated insulin secretion, in addition to *HNF1A*'s involvement in core cellular pathways such as gene transcription and cell signalling (Haliyur, et al., 2019).

To investigate the impact of *HNF1A* KO on beta cell development and function, I utilised an *HNF1A* KO hiPSC line together with an isogenic control (NKX6.1-GFP) line and the parental (SB) hiPSC line. The cell lines were then characterised by cellular phenotyping and transcriptome profiling at three distinct stages of beta cell development. In addition, the data I generated from this cell line was compared to patient derived and CRISPR/Cas9 corrected cell lines.

5.2 Methods

5.2.1 Genome editing

Dr Shailesh Gupta at AstraZeneca (Gothenburg, Sweden) performed genome editing of the *HNF1A* KO line. The genome editing was performed on lines received from the StemBancc consortium (www.stembancc.org). In summary, two sgRNAs targeting exon 4 of *HNF1A* were used to generate an *HNF1A* KO iPSC line in an NKX6.1-GFP reporter hiPSC background (Table 5.2.1) (Gupta, et al., 2018). The sgRNAs were co-transfected with Cas9-GFP and, following transfection, single cell sorting was performed in a 384 well plate using GFP expression to isolate the successfully transfected cells. Clones were analysed by fragment analysis, with successfully edited clones undergoing amplicon sequencing analysis (Next Generation Sequencing, NGS). Four clones that had been successfully genome edited were shipped to the Wellcome Trust for further experimental analysis (Table 5.2.2).

Table 5.2.1 Single guide RNA (sgRNA) sequences and PCR primer pair targeting *HNF1A* exon 4

Gene (NM_001306179)	**Forward primer (5' to 3')**	**Reverse primer (3' to 5')**
HNF1A for sgRNA#1	ACCGGGCTTCTTCTTTGCGCCGGT	AAACACCGGCGCAAAGAAGAAGCC
HNF1A for sgRNA#2	ACCGAGACACGCACCTCCGTGACG	AAACCGTCACGGAGGTGCGTGTCT
HNF1A for PCR sequencing	CAGATCTGCCAGCCTCAAACC	ACCTTCCCACGTGTCCCTTG

Table 5.2.2 Summary of *HNF1A* clones received. Of the four clones received, three were predicted to have produced knockout of *HNF1A*. A1 = allele 1, A2 = allele 2

Cell Line	Clone	A1 indel	Frameshift	A2 indel	Frameshift	KO
HNF1a_10.3	A2	-9	N	N/A	N	N
HNF1a_21.3	D1	-11	Y	-8	Y	Y
HNF1a_21.3	E8	+1	Y	-8	Y	Y
HNF1a_21.3	F3	-8	Y	+1	Y	Y

A1=allele 1, A2=allele 2, indel refers to the size of deletion (-) or insertion (+) in base pairs, Y=yes, N=no, KO=knockout

5.2.2 Differentiation experiments

Section 2.7 describes a seven stage *in vitro* differentiation protocol applied to direct the hiPSC lines towards the pancreatic endocrine lineage.

5.2.3 Quantitative real-time PCR (RT-qPCR)

For quantification of gene expression, RT-qPCR was performed as outlined in Section 2.3.3. The specific Taqman probes used for gene expression analysis in this Chapter are detailed in Table 5.2.3. Data was analysed using the ddCT method, whereby relative gene expression of a test sample is compared to a control sample on the same plate. For comparison of two unmatched groups, a two-tailed unpaired t-test was used with Welch's correction used to compare one variable between two groups. For multiple groups, a one-way analysis of variance (ANOVA) followed by Sidak's comparison test was used.

Table 5.2.3 TaqMan® gene expression assays used for real-time PCR and their target amplification regions. All assays purchased from Thermo Fisher Scientific.

Target genes	Assay number	Target region
DPP4	Hs00897386_m1	Spans exon boundary
GCG	Hs01031536_m1	Spans exon boundary
HNF1A	Hs00167041_m1	Spans exon boundary
INS	Hs00355773_m1	Spans exon boundary
NEUROD1	Hs00159598_m1	Spans exon boundary
NEUROG3	Hs01875204_s1	Maps within a single exon
NKX6.1	Hs00232355_m1	Spans exon boundary
PAX4	Hs00173014_m1	Spans exon boundary
PDX1	Hs00236830_m1	Spans exon boundary
TBP	Hs00427620_m1	Spans exon boundary
SST	Hs00356144_m1	Spans exon boundary

5.2.4 Western Blotting Analysis

Full details of the protocol followed for Western Blotting can be found in Section 2.5.2. Primary antibodies used in this Chapter are listed in Table 5.2.4, with tubulin or GAPDH used as a loading control. Secondary antibodies are provided in Table 2.3.2. Bands were quantified using lane densitometry (Image Lab 5.0, BioRad) and adjusted for surrounding background signal. The protein of interest was normalised to the respective loading control (tubulin or GAPDH), with data analysed using GraphPad Prism 8.3.0 (GraphPad Software, San Diego, California). Normalised data was first log transformed before statistical analysis. For comparison of two unmatched groups, a two-tailed unpaired t-test was used with Welch's correction used to compare one variable between two groups. For multiple groups, a one-way analysis of variance (ANOVA) followed by Sidak's comparison test was used.

Table 5.2.4 Primary Antibodies used for Western Blotting. Antibodies were diluted in 5% milk/PBST.

Antibodies	Recommended dilution	Species	Manufacturer	Cat#
Anti-GAPDH	1:10,000	rabbit	Abcam	ab37168
Anti-HNF1A	1:1000	mouse	Santa Cruz	sc-393925
Anti-HNF1A	1:1000	rabbit	Abcam	ab174653
Anti-HNF1A	1:1000	rabbit	Cell Signaling Technology	89670
β-Tubulin (E-10)	1:2000	mouse	Santa Cruz	sc-365791

5.2.5 Transcriptome profiling

5.2.5.1 RNA extraction, sequencing and quantification

Cells were harvested using pre-warmed ACCUTASE™, pelleted, resuspended in Buffer RLT (Qiagen) and stored at -80°C. RNA was extracted as described in Section 2.3.1. Library preparation and sequencing is described in Section 2.4.1. The software package DESeq2 in R version 3.6.1 was used to analyse the data (Love, et al., 2014). The beta cell differentiation RNA-Seq dataset was analysed by looking for differences between *HNF1A* KO and NKX6.1-GFP WT cell lines at each of the three stages (DE, PE and BLC) separately, controlling for batch effects. Individual analysis methods are described in more detail in Section 2.4.

5.2.5.2 Gene ontology enrichment analysis

At each stage, differentially expressed genes between *HNF1A* KO and NKX6.1-GFP were assessed for enrichment in gene ontology (GO) terms using g:Profiler (https://biit.cs.ut.ee/gprofiler/gost). The software uses a custom input gene list to carry out an over-representation analysis to map genes to functional information to enable the detection of statistically significant enriched biological terms such as GO biological processes (Ensembl) and pathways (KEGG and Reactome) (Raudvere, et al., 2019). All

human genes annotated in the Ensembl database were used as background. A $p<0.01$ was used as a significance threshold. For gene lists from previously published datasets an enrichment analysis was performed using a hypergeometric test.

5.2.6 RNA interference

Gene silencing was performed in EndoC-βH1 cells using small interfering RNAs (siRNAs) in combination with lipid-based reverse transfection. ON-TARGETplus siRNAs were purchased from Dharmacon (GE Healthcare) as a pool of four oligonucleotides (SMART pool). Table 5.2.5 details the siRNAs used. siRNAs were diluted in Opti-MEM reduced serum-free medium (Gibco™, 31985062) to a final concentration of 20nM, mixed with 0.4% Lipofectamine RNAiMAX (Thermo Fisher Scientific, 13778150) and incubated for 15-20min at RT. Transfection complexes were added prior to addition of cells resuspended in culture media. Silencing efficiency was measured 96h later by RT-qPCR and/or WB analysis.

Table 5.2.5 siRNAs for RNA interference studies

Target Gene	Manufacturer	Cat #
Non-targeting pool	Dharmacon	D-001810-10
HNF1A	Dharmacon	L-008215-00

All siRNAs were purchased as ON-TARGETplus SMARTpool containing a mix of four siRNAs

5.2.7 Statistical analyses

Statistical analyses were performed in GraphPad Prism 8 (v8.3) with results from multiple experiments expressed as mean ±SEM. For two unmatched groups, the p-value was determined using a two-tailed unpaired t-test. Comparison of groups larger than two was done with a one-way ANOVA followed by Sidak's multiple comparisons test. A p-value of <0.05 was considered statistically significant, with n representing the number of biologically independent experiments.

5.3 Results

5.3.1 Generation of the *HNF1A* KO cell line

Generation of the *HNF1A* KO cell line was carried out at AstraZeneca (Gothenburg, Sweden). Of the 384 clones generated, 40 single cell clones harbouring indels were taken forward for fragment analysis. Of these, 10 were sent for next generation sequencing, with 5 clones then taken forward for expansion. One clone failed to be successfully expanded and the remaining four were subsequently shipped to the Wellcome Centre, University of Oxford for downstream experiments.

5.3.2 Characterisation of *HNF1A* KO hiPSC model

All *HNF1A* KO clones were re-genotyped upon arrival to ensure the correct identity of each cell line. The four *HNF1A* KO clones, alongside the parental SB line and NKX6.1-GFP reporter line were assessed for pluripotency and genomic stability as outlined in Section 3.4.4. All four *HNF1A* KO clones failed initial karyotyping analysis and were plated at low density for picking of single cell colonies. For each cell line, three sub-clones were sent for karyotyping analysis, with only one sub-clone of clone D1 successfully passing (Figure 5.3.1). This clone had deletions in exon 4, with one allele having a deletion of 11 base pairs and the other allele a deletion of 8 base pairs. Both the SB and NKX6.1-GFP reporter lines passed karyotyping. All three lines also expressed the pluripotency markers OCT4, SOX2, NANOG and SSEA-4 as assessed by flow cytometry, using an isotype (negative) control to account for non-specific background signal caused by primary antibodies (Figure 5.3.2).

To examine the role of *HNF1A* in beta cell development, one clone for each of SB, NKX6.1-GFP and *HNF1A* KO were taken forward for *in vitro* beta cell differentiation.

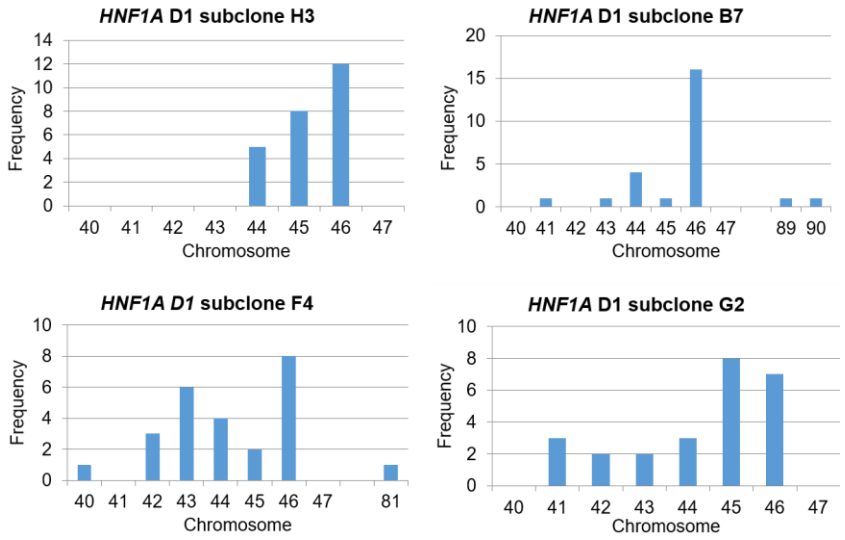


Figure 5.3.1 Karyotyping of *HNF1A* subclones
Subclones of *HNF1A* clone D1 (H3, B7, F4 and G2) were sent for karyotyping, with chromosome counting undertaken by Dr Daniela Moralli (Wellcome Centre for Human Genetics, Oxford). Only one subclone (H3) successfully passed with all other clones displaying an abnormal karyotype.

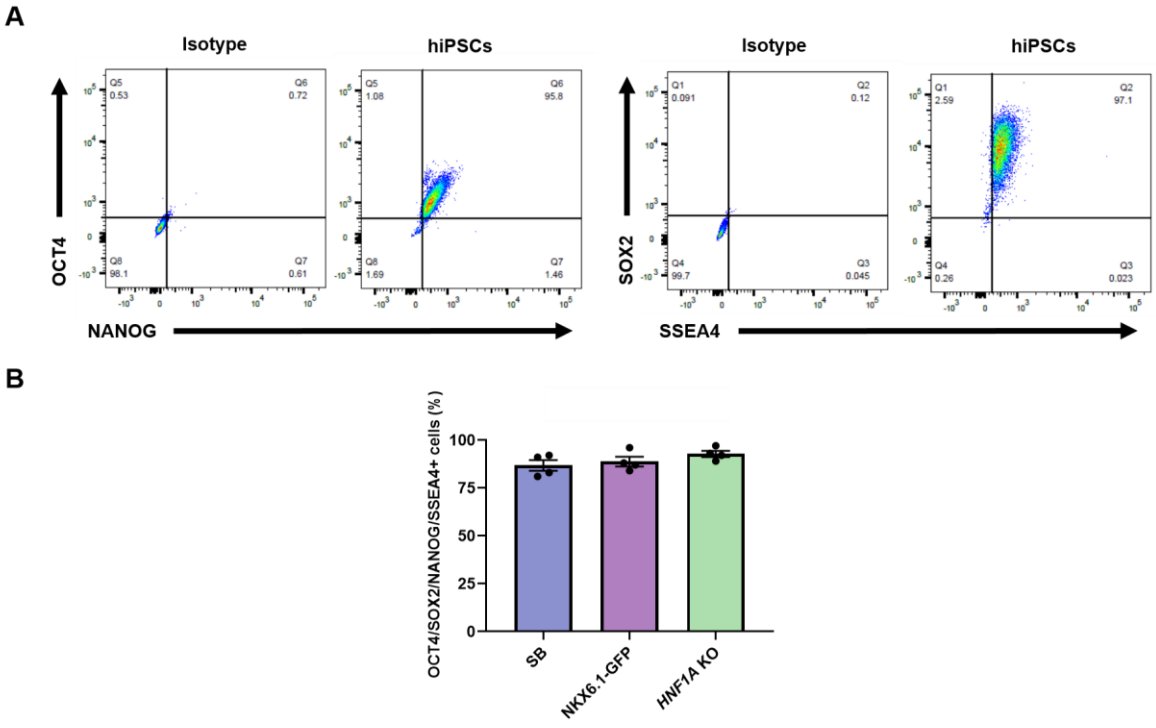


Figure 5.3.2 Characterisation of *HNF1A* KO hiPSC lines for stem cell properties
(A) Representative flow cytometry plots for edited hiPSC lines with isotype control, enabling the correct gating to be established. Cell lines were assessed for expression of common pluripotency markers including OCT4, NANOG, SSEA-4 and SOX2 as summarised in (B). SB n=4, NKX6.1-GFP n=4, *HNF1A* KO n=4. Plots display mean±SEM. Data analysed using one-way ANOVA.

5.3.3 Differentiation of *HNF1A* KO hiPSC models into pancreatic beta-like cells and cellular phenotyping at three stages of development

To model pancreatic endocrine development a directed step-wise protocol was undertaken (see Section 2.7). Flow cytometry analysis was performed to assess whether loss of *HNF1A* had any effect on differentiation efficiency. In addition, RNA and protein lysates were collected at stages 1, prior to *HNF1A* expression, stage 4, after the initiation of *HNF1A* expression and stage 7, the end point of differentiation.

5.3.3.1 Evaluation of developmental markers of beta-cell differentiation

As outlined in Section 1.4.5.1 and Section 2.7, *in vitro* endocrine development is characterised by expression of particular beta-cell development markers at specific stages. To assess whether *HNF1A* KO had any effect on differentiation efficiency, mRNA and/or protein expression was assessed to compare stage-specific developmental markers. In the first instance, data was compared between the SB, NKX6.1-GFP and *HNF1A* KO lines. As the SB line has been differentiated numerous times within our group previously, it was included to act as a differentiation control to the NKX6.1-GFP line, which had not been extensively differentiated within our laboratory. Additional comparisons were made between the NKX6.1-GFP and *HNF1A* KO lines alone, as the *HNF1A* KO line was derived from the NKX6.1-GFP line and, as there was no CRISPR sham for the *HNF1A* KO, the NKX6.1-GFP line acted as the nearest equivalent.

Stage 1 – Definitive endoderm

Cell lines were assessed at definitive endoderm as outline in Section 3.5.1. Assessment of the expression of CXCR4 and SOX17 using flow cytometry showed no significant differences between genotypes ($p=0.158$ and $p=0.321$, respectively) (Figure 5.3.3). Transcript levels of *SOX17* were also similar between cell lines ($p=0.217$) (Figure 5.3.3).

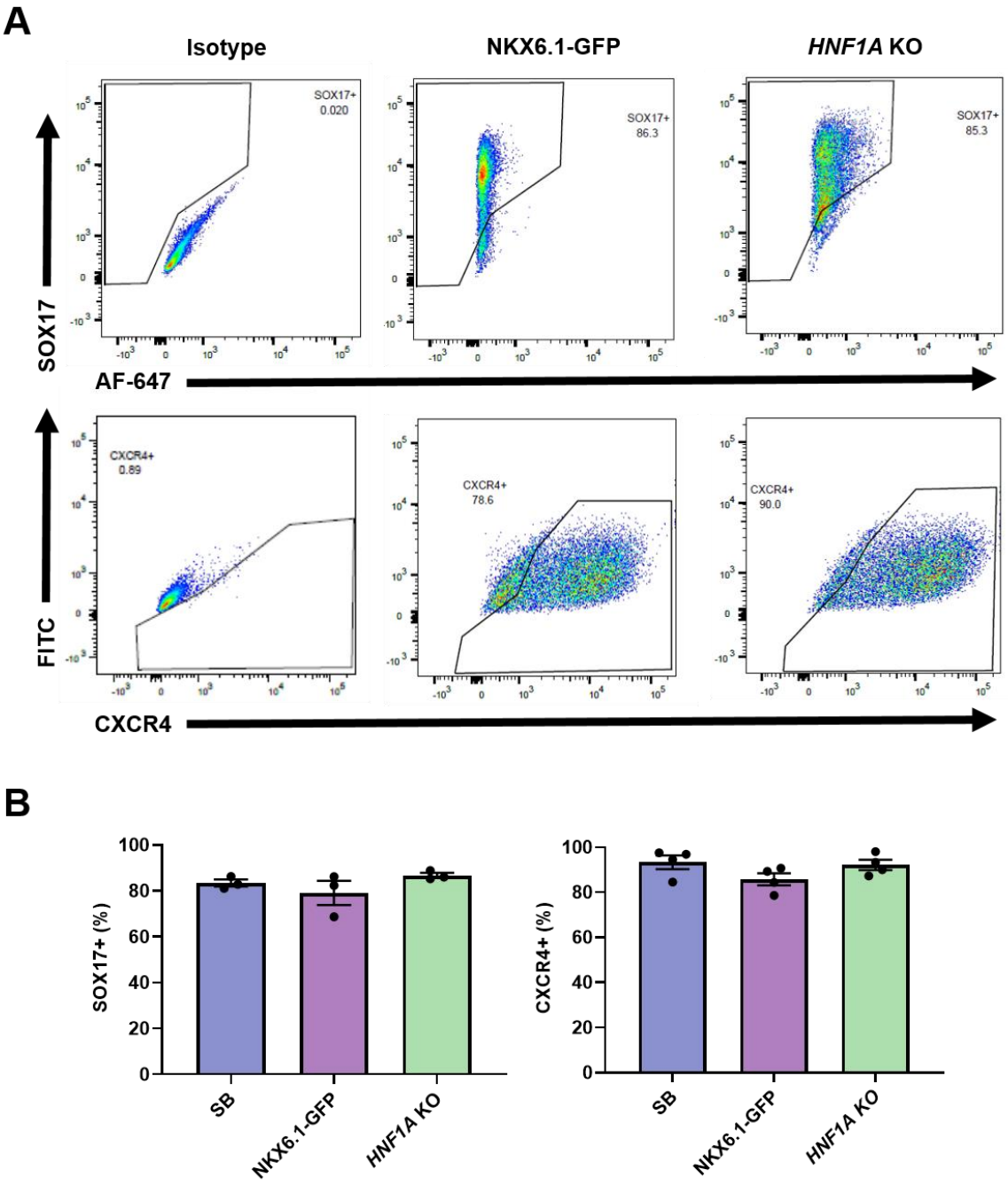


Figure 5.3.3. Assessment of markers for stage 1 (DE) of differentiation
Differentiated clones were evaluated for expression of SOX17 and the chemokine cell receptor CXCR4, with isotype control, enabling the correct gating to be established. (A) Representative flow cytometry plots for SOX17 and CXCR4 for the NKX6.1-GFP and *HNF1A* KO cell lines. (B) There were no significant differences in the expression of CXCR4 or SOX17 in the cell lines. SB n=3-4, NKX6.1-GFP n=3-4, *HNF1A* KO n=3-4. Plots display mean±SEM. Data analysed using one-way ANOVA.

Stage 4 – Pancreatic Endoderm

Transcript analysis showed no differences in the expression of either *NKX6.1* or *PDX1* between the cell lines (p=0.352 and p=0.222, respectively). Flow cytometry analysis demonstrated a robust expression of PDX1 in the majority of clones with no significant difference between cell lines (p=0.727). Co-expression of PDX1/NKX6.1 was low with no difference between cell lines (p=0.0706) (Figure 5.3.4). Previous work has shown that high levels of NKX6.1 expression is critical for the establishment of functional endocrine cells, particularly in the formation of insulin positive cells (Rezania, et al., 2013). An absence of *NKX6.1* induction has been shown to result primarily in the formation of polyhormonal cells (Petersen, et al., 2017). These studies suggest that my clones, at stage 7, may produce mostly polyhormonal cells and have low insulin expression. One *HNF1A* KO clone exhibited an almost complete absence of co-expression, in addition to negligible expression of NKX6.1 and PDX1 alone, indicating that this clone could be an outlier. Removal of this clone from the analysis revealed that the SB clones expressed significantly less PDX1+/NKX6.1+ cells than both the NKX6.1-GFP and *HNF1A* KO lines (p=0.0163 and p=0.0045, respectively) (Figure 5.3.5). This indicates that the SB clones established fewer pancreatic progenitors at this stage. Previous work with the SB line within our lab has shown a higher proportion of PDX1+/NKX6.1+ cells (personal communication, Dr Nicole Krentz, Gloyn Group), suggesting that differentiation for this line was less efficient within this experiment.

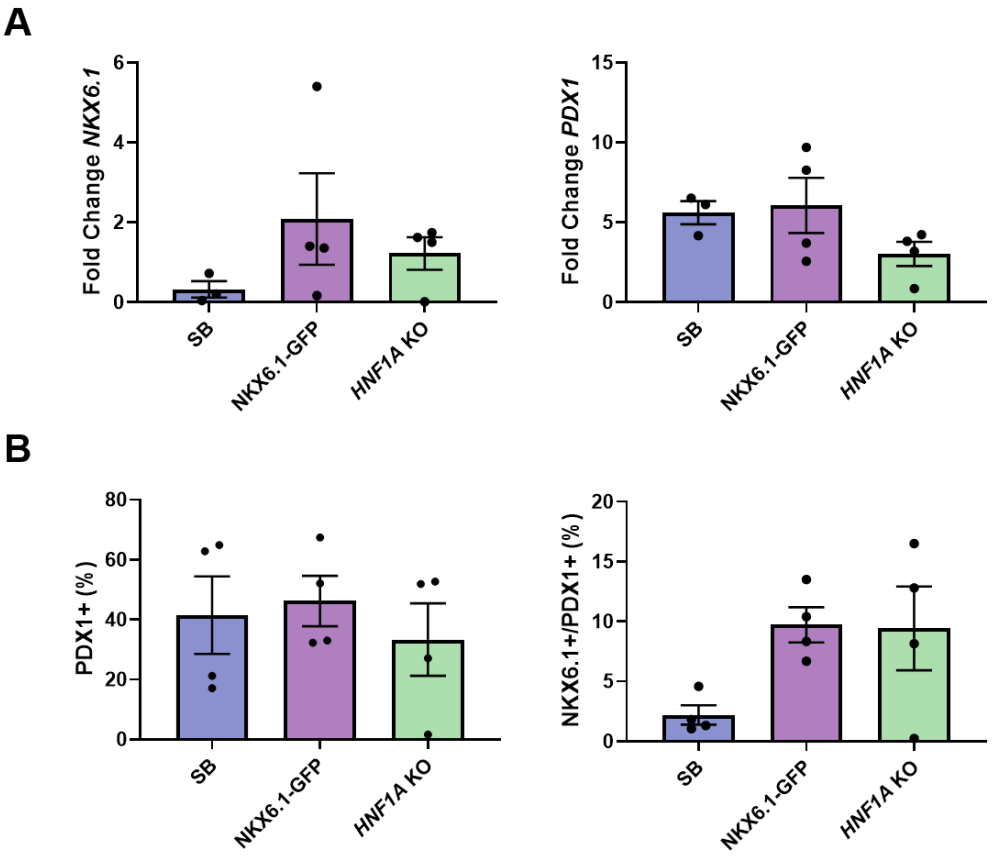


Figure 5.3.4 Assessment of the expression of key developmental markers at stage 4 of differentiation
Differentiated clones were evaluated for expression of *PDX1* and *NKX6*.1 by (A) RT-qPCR and for (B) PDX1 and NKX6.1 by FACS. SB n=4, NKX6.1-GFP n=4, *HNF1A* KO n=4. Plots display mean±SEM. Data analysed using one-way ANOVA.

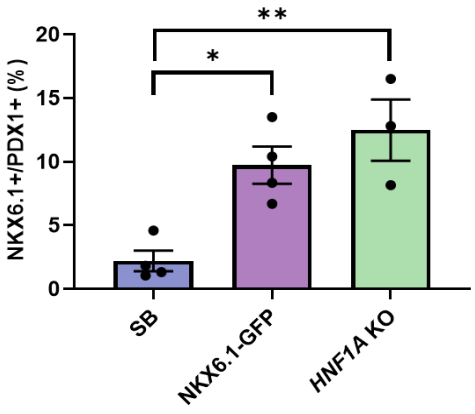


Figure 5.3.5 Co-expression of NKX6.1 and PDX1 at stage 4 of differentiation
Differentiated clones were evaluated for expression of PDX1 and NKX6.1. There was a significant decrease in the proportion of cells expressing both NKX6.1 and PDX1 in the SB line but only when one potential *HNF1A* KO outlier sample was removed. SB n=4, NKX6.1-GFP n=4, *HNF1A* KO n=3. Plots display mean±SEM. Data analysed using one-way ANOVA. *p<0.05, **p<0.01

Stage 7 – Beta-like cells

As outlined in Section 3.5.1, cells were assessed for C-peptide as a proxy for insulin expression, with glucagon also evaluated. There were some samples that had suboptimal cell numbers for flow cytometry analysis and gating was difficult to correctly establish due to a shift in cellular population between the isotype control and certain samples (Figure 5.3.6). Therefore, results must be interpreted with some caution. Analysis of the cell lines showed that the *HNF1A* KO had significantly reduced expression of C-peptide compared to the SB line ($p=0.047$), with this finding replicated if *HNF1A* KO was compared to NKX6.1-GFP alone ($p=0.0442$). There was no significant difference between the cell lines for glucagon expression ($p=0.128$) but the *HNF1A* KO cell line had significantly reduced glucagon expression when compared to NKX6.1-GFP alone ($p=0.0044$) (Figure 5.3.7).

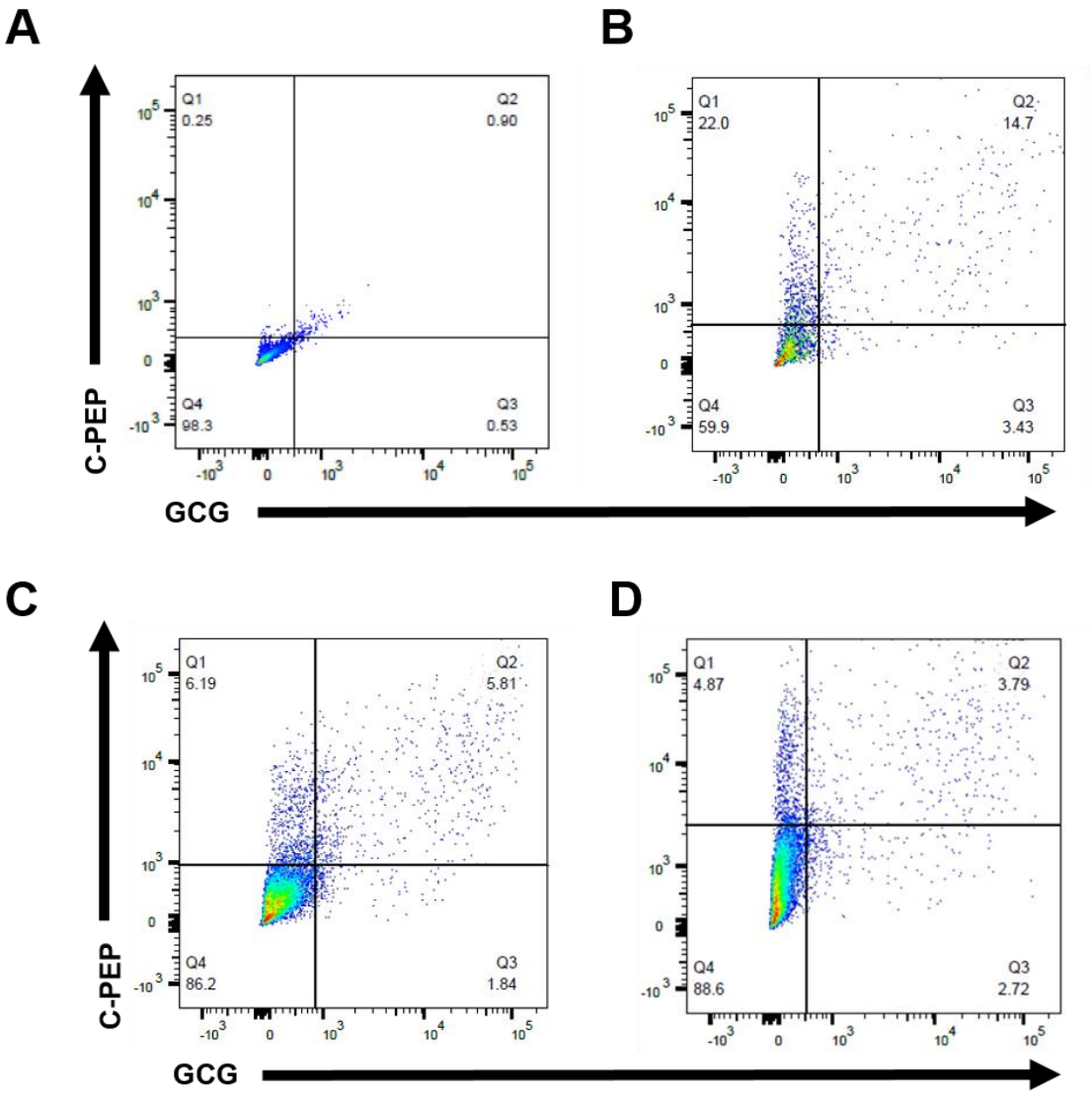


Figure 5.3.6. Variation in cell number and expression profile of different flow cytometry samples

Flow cytometry assessment of C-peptide and glucagon was impaired by variation in sample expression profiles and cell numbers (A) Isotype control used for gating (B) Low cell number of test sample making it difficult to accurately capture expression profile. Fluorescent shift from isotype to test sample impairs ability to confidently gate sample with (C) and (D) showing differences in expression profiles of two different differentiations of the same cell line (NKX6.1-GFP)

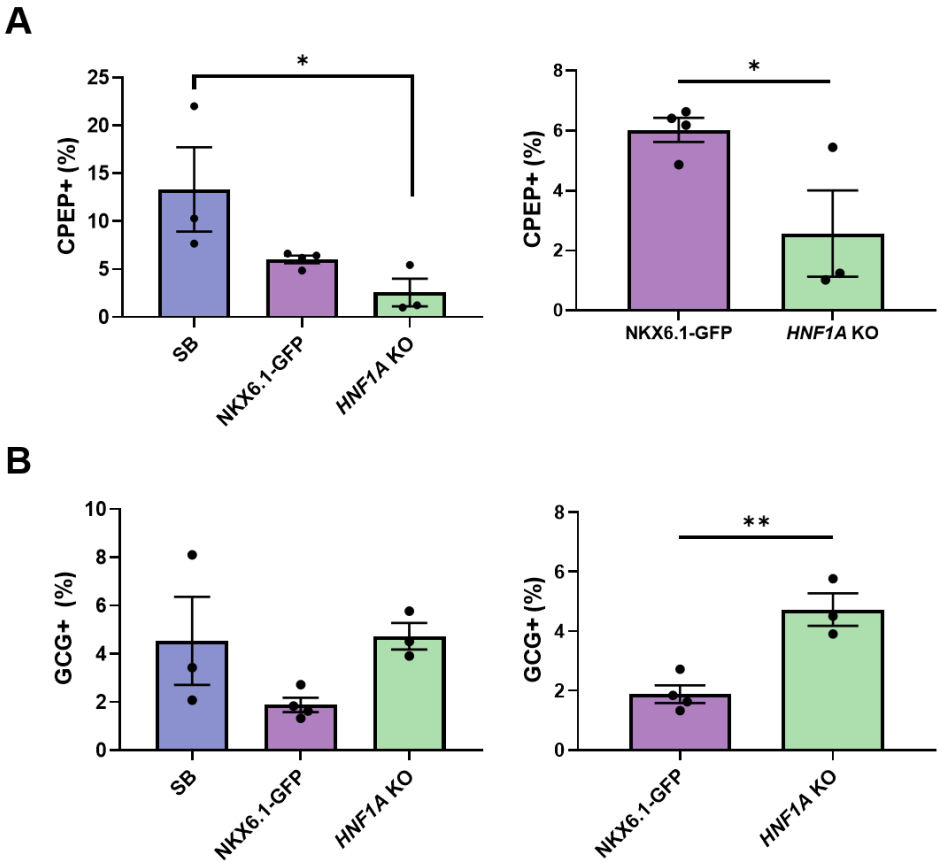


Figure 5.3.7 Analysis of C-peptide and glucagon at stage 7 of differentiation
Flow cytometry assessment at stage 7 of differentiation (A) Comparison of C-peptide between all groups (*left*) with subsequent comparison of C-peptide between the NKX6.1-GFP and *HNF1A* KO lines (*right*) (B) Comparison of glucagon between all groups (*left*) with subsequent comparison of glucagon between the NKX6.1-GFP and *HNF1A* KO lines (*right*). C-PEP=C-peptide, GCG=glucagon. SB n=4, NKX6.1-GFP n=4, *HNF1A* KO n=3. Plots display mean±SEM. Analysis between three groups (SB, NKX6.1-GFP, *HNF1A* KO) performed using one-way ANOVA, comparison between NKX6.1-GFP and *HNF1A* KO lines analysed using two-tailed unpaired t-test *p<0.05, **p<0.01

Transcript levels of *INS* and *GCG* were also measured via RT-qPCR. Transcript analysis of all three cell lines for *GCG* and *INS* expression showed no significant differences between cell lines (p=0.186 and p=0.1179, respectively). However, when the NKX6.1-GFP line was compared to *HNF1A* KO alone, the *HNF1A* KO line displayed a significantly reduced expression of *GCG* (p=0.0134) and *INS* (p=0.0027) (Figure 5.3.8).

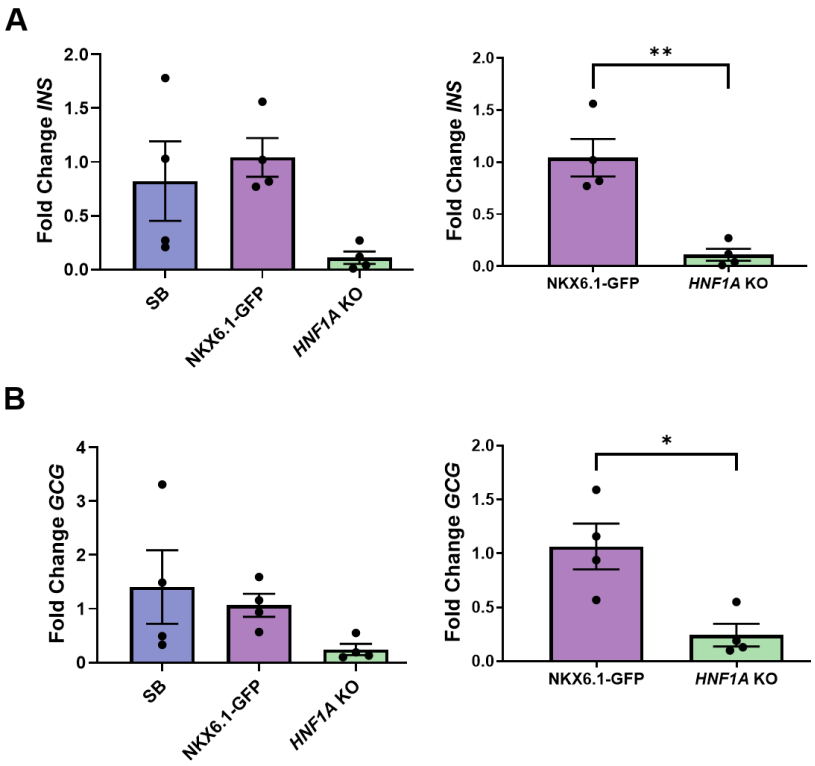


Figure 5.3.8 Significant differences in the expression of *INS* and *GCG* at stage 7 between the NKX6.1-GFP and *HNF1A* KO cell lines
Expression levels of (A) *INS* and (B) *GCG* were measured by RT-qPCR at stage 7 of differentiation in the three groups (*left*) and in the NKX6.1-GFP and *HNF1A* KO cell lines (*right*). SB n=4, NKX6.1-GFP n=4, *HNF1A* KO n=4. Plots display mean±SEM. For three groups, data analysed using one-way ANOVA, between two groups the data was analysed using a two-tailed unpaired t-test. *p<0.05, **p<0.01

5.3.3.2 Evaluation of *HNF1A* mRNA and HNF1A protein expression in edited *HNF1A* KO hiPSC lines

Transcript levels of *HNF1A* were measured at stage 4 and 7. At stage 4, all three cell lines exhibited expression of *HNF1A*, with the *HNF1A* KO line expressing significantly less *HNF1A* to the SB line (p=0.0333). The levels of *HNF1A* between NKX6.1-GFP and the *HNF1A* KO when compared in isolation were similar (p=0.132). At stage 7, expression of *HNF1A* was similar between cell lines (p=0.190) but there was a significant decrease of *HNF1A* expression in the *HNF1A* KO cell line when compared to NKX6.1-GFP alone (p=0.0372) (Figure 5.3.9).

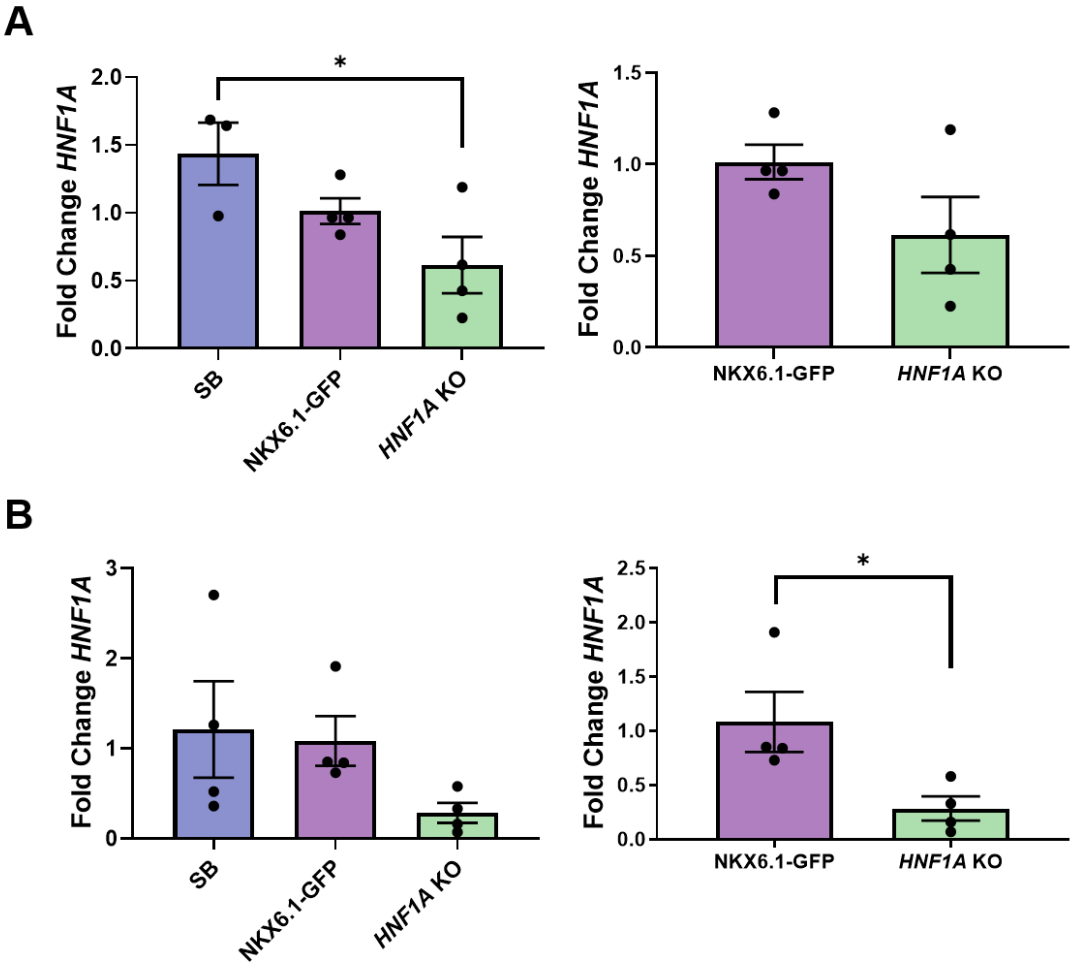


Figure 5.3.9 Expression of *HNF1A* evaluated with RT-qPCR at stage 4 and 7 of differentiation

Expression of *HNF1A* was evaluated using RT-qPCR at (A) stage 4 for all three groups (*left*) and between the NKX6.1-GFP and *HNF1A* KO lines (*right*), and (B) stage 7 for all three groups (*left*) and between the NKX6.1-GFP and *HNF1A* KO lines (*right*). SB n=4, NKX6.1-GFP n=4, *HNF1A* KO n=3. Plots display mean±SEM. Analysis between three groups performed using one-way ANOVA, comparison between NKX6.1-GFP and *HNF1A* KO lines analysed using two-tailed unpaired t-test. *p<0.05, **p<0.01

The *HNF1A* Taqman® probe spans exon 2-3, prior to the expected CRISPR/Cas9 cut site and therefore this region may still be amplified, with previous studies showing that incomplete NMD can occur after introduction of a frameshift mutation (Reber, et al., 2018). Protein levels of HNF1A were unable to be reliably detected at stage 4 and 7, despite trying three different antibodies (Table 5.2.4) that had been used previously within our lab for HNF1A detection, probably due to low levels of HNF1A within the samples (Figure 5.3.10).

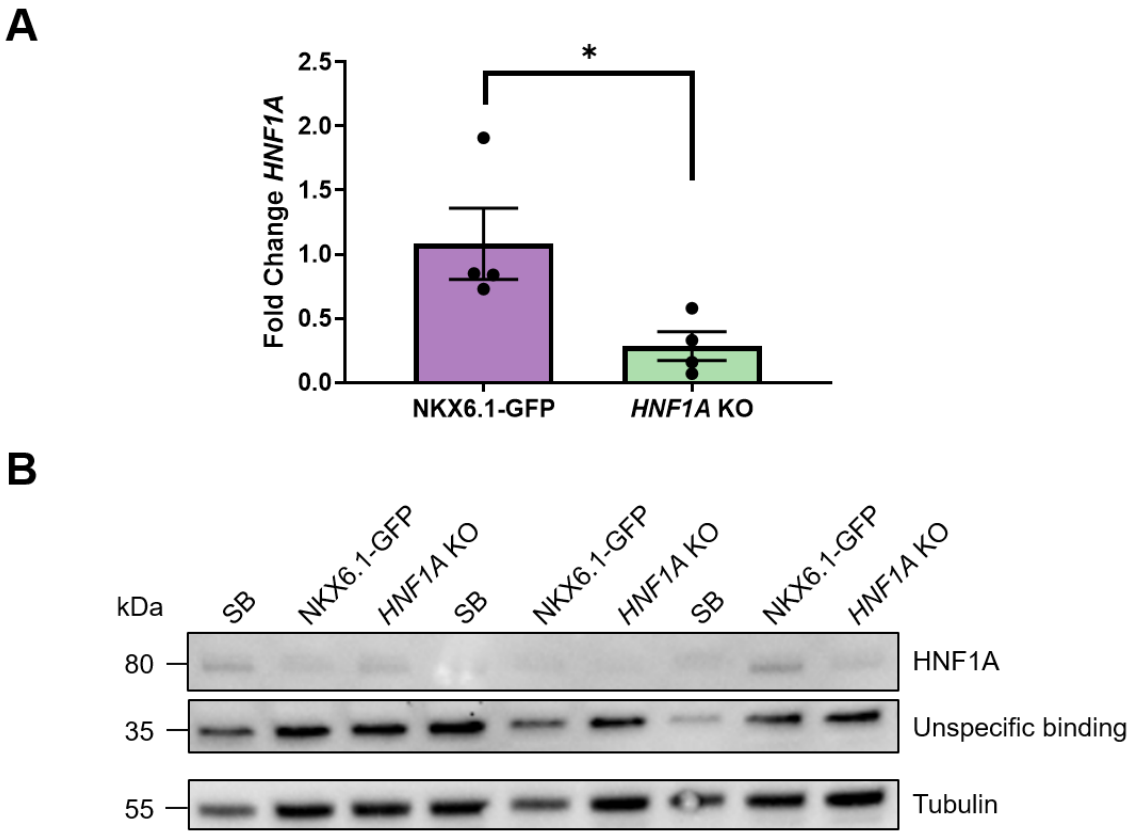


Figure 5.3.10 Significant difference in the expression of *HNF1A* at stage 7 between the NKX6.1-GFP and *HNF1A* KO cell lines but difficulty in detection of HNF1A protein using Western Blot
Expression levels of *HNF1A* were measured by (A) RT-qPCR and (B) Western Blot at stage 7 of differentiation in the NKX6.1-GFP and *HNF1A* KO cell lines. NKX6.1-GFP n=4, *HNF1A* KO n=4. The RT-qPCR plot displays mean±SEM and was analysed using two-tailed unpaired t-test. *p<0.05

5.3.4 Transcriptome profiling of *HNF1A* KO and WT hiPSC lines differentiated along the endocrine lineage

Having shown that loss of *HNF1A* led to decreased expression of insulin and glucagon at stage 7, I then wanted to examine how loss of *HNF1A* may affect genes involved in beta cell development and function. Therefore, I performed transcriptome analysis on SB, NKX6.1-GFP and *HNF1A* KO lines at definitive endoderm (stage 1), pancreatic endoderm (stage 4) and beta-like cell stages (stage 7).

5.3.4.1 *HNF1A* transcript levels were down-regulated in differentiated *HNF1A* KO lines

Overall, 36 RNA-Seq libraries were sequenced to a mean read depth of 16.2±1.4 million reads per sample. Whilst the RT-qPCR results revealed a difference in *HNF1A* expression between SB and *HNF1A* KO at stage 4, transcriptomic analysis revealed no such difference but did reveal a significant reduction in *HNF1A* expression at stage 7 in the *HNF1A* KO cell line when compared to the NKX6.1-GFP line (LFC=-1.58, padj=6.29x10^{-4}) (Figure 5.3.11).

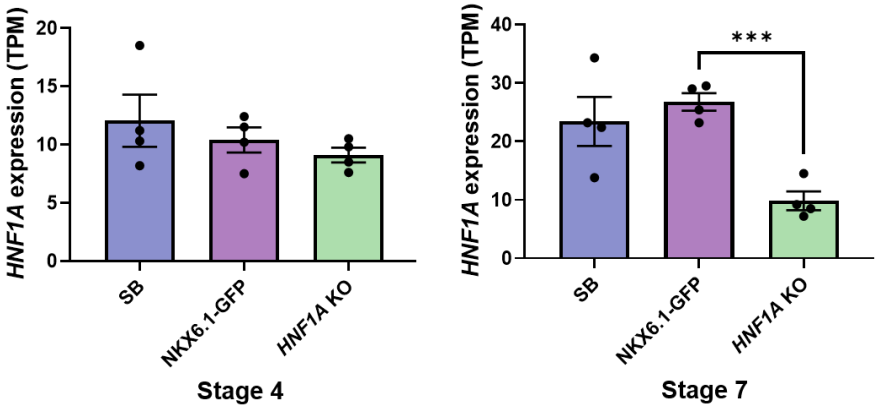


Figure 5.3.11 *HNF1A* transcript levels were down-regulated in differentiated *HNF1A* KO lines at stage 7
Transcriptomic analysis at stage 4 revealed no differences in *HNF1A* expression between lines. At stage 7, there was a significant reduction in *HNF1A* expression in the *HNF1A* KO cell line when compared to the NKX6.1-GFP line, as demonstrated by its appearance as a differentially expressed gene (LFC=-1.58, padj=6.29x10^{-4}). TPM=transcripts per million. Data displayed as mean±SEM, analysed using one-way ANOVA. ***p<0.001

5.3.4.2 Principal component analysis revealed samples grouped by stage and genotype

Principal component analysis (PCA) of the transcriptome of the samples showed that they grouped by their developmental stage (Figure 5.3.12). PCA was then performed on each individual stage. Even at stage 1 (DE), prior to *HNF1A* expression, samples grouped by genotype, with separation between the SB, NKX6.1-GFP and *HNF1A* KO groups. This continued at stage 4 (PE) and stage 7 (BLC), with one *HNF1A* KO sample diverging from the other *HNF1A* KO clones. This sample also had negligible expression of PDX1/NKX6.1 at stage 4 and insufficient sample to be confidently assessed at stage 7 (BLC) via flow cytometry. Due to the divergence of this sample from those of the same background, the sample was classed as an outlier and removed from subsequent analysis (Figure 5.3.12).

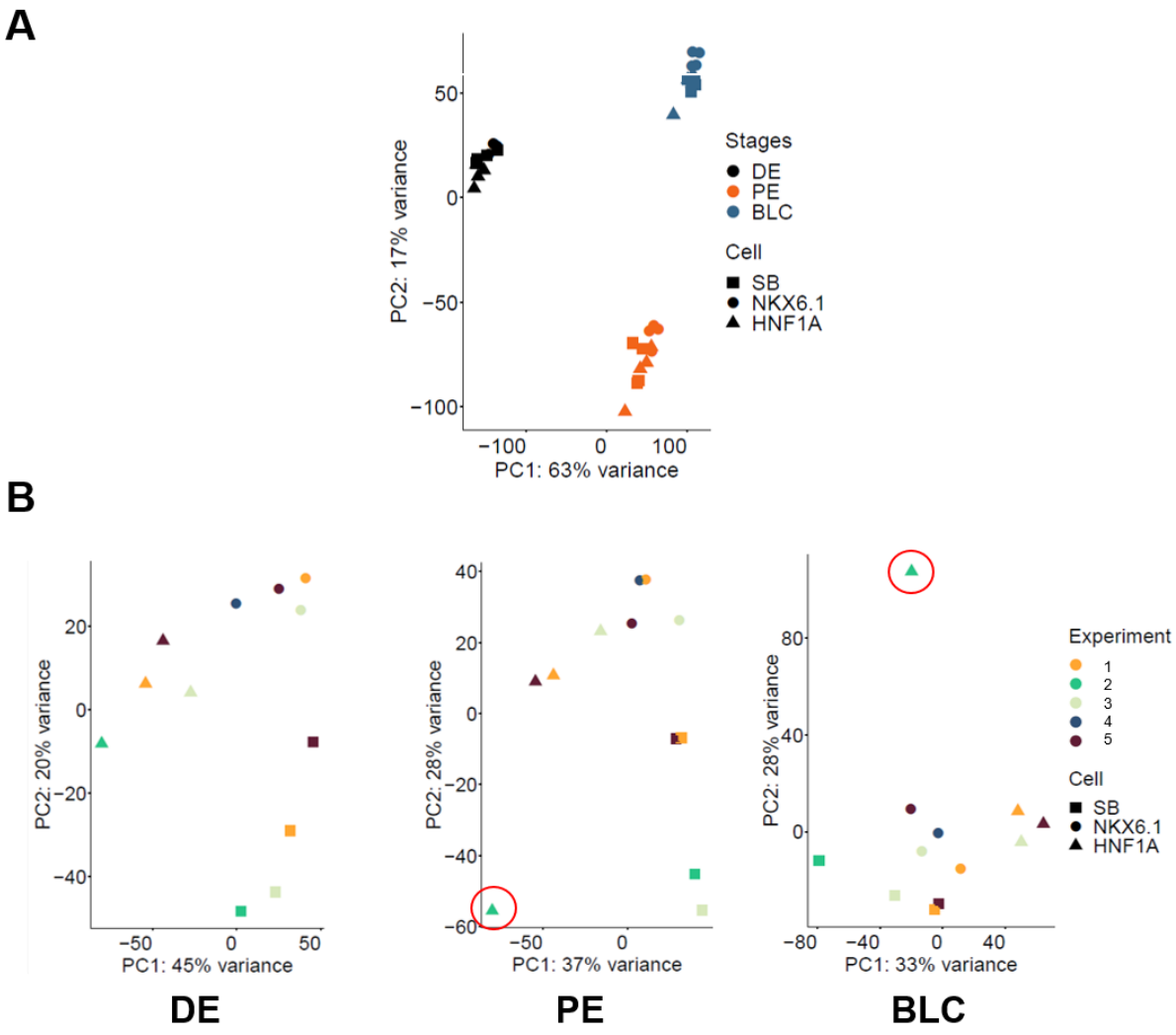


Figure 5.3.12 Principal component analysis of the transcriptomes of differentiated SB, NKX6.1-GFP and *HNF1A* KO clones

(A) The first two principal components (PC1, PC2) were calculated using normalised gene counts of the clones for the three selected stages of *in vitro* beta differentiation, with samples clustering by stage. (B) The first two principal components (PC1, PC2) for the three selected stages of differentiation. The outlier *HNF1A* KO is highlighted at the PE and BLC stages by the red circle. SB n=4, NKX6.1-GFP n=4, *HNF1A* KO n=4, DE, definitive endoderm; PE, pancreatic endoderm; BLC, beta-like cells.

5.3.4.3 Differential gene expression (DEG) analysis

The Bioconductor DESeq2 package (see Section 2.4.2.1) was used to identify differentially expressed genes at the three development stages, with those at stage 4 (PE) and stage 7 (BLC) of particular importance as *HNF1A* is not expressed at the DE stage. A total of 2,722 differentially expressed genes (DEGs) were identified in the three stages between *HNF1A* KO and NKX6.1-GFP lines. A selection of the enriched biological terms and pathways that were among the DEGs for the *HNF1A* KO clones at each of the three stages are listed in Table 5.3.1. At both the PE and BLC stage, there was an enrichment of genes involved in "regulation of gene expression in endocrine committed (NEUROG3+) progenitor cells" and at stage 7 for genes involved in "insulin secretion", "regulation of gene expression in beta cells" and "regulation of beta cell function".

Table 5.3.1 Subset of enriched biological terms and pathways among DEGs at stages DE, PE and BLC in the *HNF1A* KO cell lines

Stage		# enriched terms	Examples of enriched terms	padj
DE	GO	68	Extracellular matrix organisation	$2.8x10^{-7}$
			Negative regulation of cell growth	0.0012
			Negative regulation of cell proliferation	0.0012
	KEGG	13	PI3K-Akt signalling pathway	0.0035
			Calcium signalling pathway	0.0020
	REACTOME	18	FOXO-mediated transcription of cell cycle genes	0.0010
			Extracellular matrix organisation	0.0020
PE	GO	25	Glycolytic process	$1.2x10^{-5}$
			Potassium ion transport	$6.7x10^{-4}$
	KEGG	25	Glycolysis / Gluconeogenesis	$3.2x10^{-4}$
			Pentose phosphate pathway	0.0063
	REACTOME	22	FOXO-mediated transcription	$3.06x10^{-4}$
			Regulation of gene expression in endocrine-committed (NEUROG3+) progenitor cells	$3.92x10^{-4}$
			Potassium channels	0.0030
BLC	GO	102	Extracellular matrix organization	$3.1x10^{-12}$
			Ion transmembrane transport	$1.1x10^{-5}$
			Glucose homeostasis	$3.7x10^{-4}$
			Regulation of insulin secretion	$4.4x10^{-4}$
	KEGG	20	Cell adhesion molecules	$3.6x10^{-8}$
			Maturity onset diabetes of the young	$8.8x10^{-6}$
			Insulin secretion	0.0039
	REACTOME	34	Regulation of gene expression in beta cells	$9.11x10^{-5}$
			Regulation of gene expression in endocrine-committed (NEUROG3+) progenitor cells	$8.05x10^{-5}$
			Regulation of beta cell development	$9.11x10^{-5}$

5.3.4.4 Pathway analysis demonstrated that MODY and insulin secretion genes were enriched among DEGS in the *HNF1A* KO lines at the beta-like cell stage

To evaluate whether loss of HNF1A had an impact on beta cell function, DEGs between *HNF1A* KO and NKX6.1-GFP clones were tested for enrichment in gene ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways using a hypergeometric test. The analysis revealed an enrichment of MODY-causing genes among genes downregulated in the *HNF1A* KO cell line at stage 7 (padj=0.00552) such as *HNF4A* (LFC=-1.44, padj=$1.08x10^{-05}$) and *PAX4* (LFC=-2.31, padj=$3.59x10^{-06}$). In addition there was an enrichment of insulin secretion genes among genes upregulated in the *HNF1A* KO line (padj=0.0267) such as *NNAT* and *KCNA5*. However, there were also notable markers of insulin secretion such as *INS* (LFC=-2.65, padj=$3.20x10^{-40}$), *GCG* (LFC=-1.67, padj=$3.16x10^{-09}$), *NEUROD1* (LFC=-1.59, padj=$2.89x10^{-09}$), *RFX6* (LFC=-2.26, padj=$3.13x10^{-11}$) and *DPP4* (LFC=-1.79, padj=$1.1x10^{-08}$), which were downregulated in the *HNF1A* KO cells, although this did not reach statistical significance (padj=0.076). Overall, my data demonstrates a role for *HNF1A* in regulating genes involved in insulin secretory machinery.

5.3.4.5 *HNF1A* KO DEGs at stage 7 are significantly enriched in a previously published HNF1A-MODY dataset

To further investigate my findings, I compared my results to recently published data from an RNA-Seq performed on the isolated islets of an individual with an *HNF1A* variant (Haliyur, et al., 2019). There was a significant enrichment of my *HNF1A* KO DEGs at stage 7 within the up and downregulated alpha cell genes from the patient data set (p=0.00737 and p=$1.72x10^{-6}$, respectively) (Figure 5.3.13). There was also an enrichment of my upregulated DEGs in the upregulated beta cell genes and my downregulated DEGs

in the downregulated beta cell genes of the patient (p=0.00283 and p=0.000383, respectively) (Figure 5.3.13). Although the islet data only comes from one individual, it is reassuring to see that many of the genes captured in a patient are also present in my iPSC model.

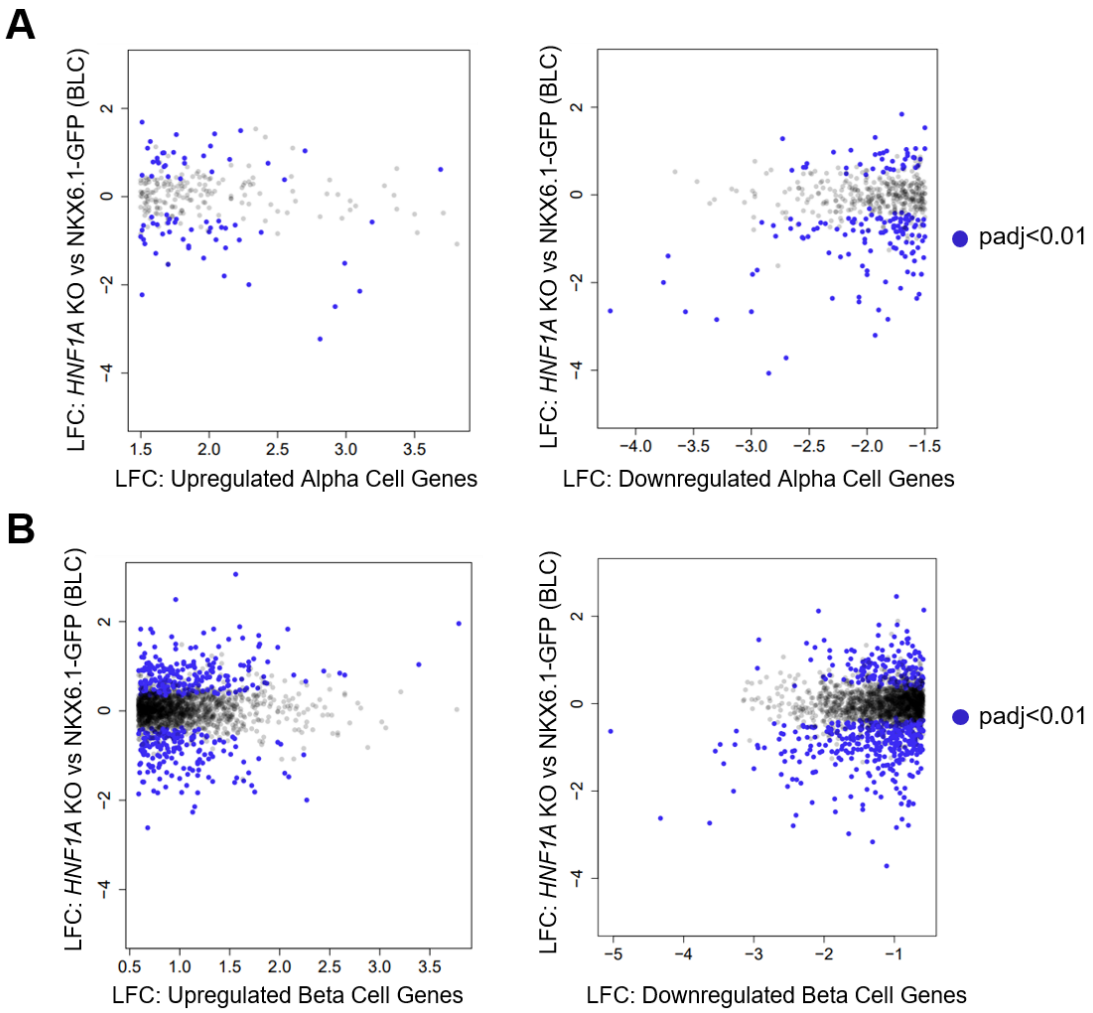


Figure 5.3.13 *HNF1A* KO DEGs at stage 7 are significantly enriched in a previously published HNF1A-MODY patient dataset
There was a significant enrichment of my *HNF1A* KO DEGs at stage 7 within the (A) up and downregulated alpha cell genes from the patient data set. (B) There was an enrichment of my *HNF1A* KO upregulated DEGs in the upregulated beta cell genes and my *HNF1A* KO downregulated DEGs in the downregulated beta cell genes of the same patient dataset (Haliyur, et al., 2019). Data analysed using a hypergeometric enrichment test. LFC=log fold change, BLC=beta-like cells. Genes with a padj<0.01 for both datasets are coloured blue

I also compared my data to an iPSC model where both an *HNF1A* heterozygous and homozygous KO were generated and differentiated along the endocrine lineage (Cardenas-Diaz, et al., 2019). There was no enrichment between my *HNF1A* KO data and the heterozygous KO. However, there was a significant enrichment of my downregulated *HNF1A* KO DEGs within the DEGs for the homozygous KO (p=0.0383) (Figure 5.3.14).

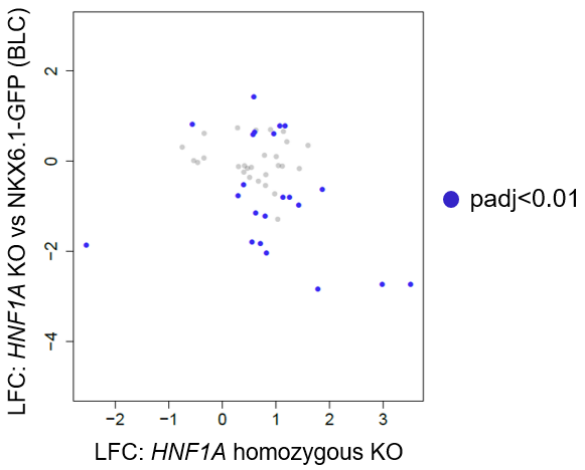


Figure 5.3.14 Downregulated *HNF1A* KO DEGs at stage 7 (BLC) are significantly enriched within the DEGs of a previously published *HNF1A* homozygous KO dataset
There was a significant enrichment of my downregulated *HNF1A* KO DEGs within the DEGs for the previously published *HNF1A* homozygous KO (Cardenas-Diaz, et al., 2019). Data analysed using a hypergeometric enrichment test. LFC=log fold change, BLC=beta-like cells. Genes with a padj<0.01 for both datasets are coloured blue

Similarly, I examined my data with that of our collaborators at Novo Nordisk (Dr Christian Honoré and Rikke Rejnholdt Jensen) who used CRISPR/Cas9 to correct, and then differentiated, hiPSC lines obtained from patients with HNF1A-MODY due to the Pro291fsincC mutation. I found that there was a significant enrichment at stage 7 of both my up- (p=0.00022) and downregulated (p=0.013) DEGs in the respective populations of the patient-derived iPSC line (Figure 5.3.15).

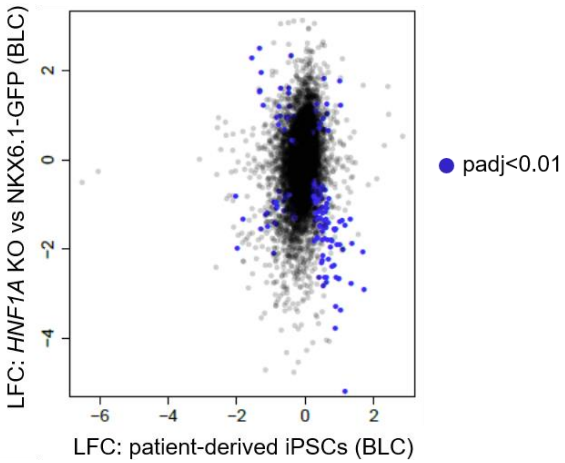


Figure 5.3.15 Significant enrichment of *HNF1A* KO DEGs at stage 7 (BLC) within the DEGs of patient-derived HNF1A-MODY iPSCs at stage 7 (BLC)
There was a significant enrichment at stage 7 of both my up- and downregulated *HNF1A* KO DEGs in the same respective populations of the DEGs for stage 7 of the patient-derived HNF1A-MODY iPSC line. Data analysed using a hypergeometric enrichment test. LFC=log fold change, BLC=beta-like cells. Genes with a padj<0.01 for both datasets are coloured blue

The overlap between the datasets demonstrates that common gene targets of *HNF1A* can be identified between different experimental platforms. However, one dissimilarity is that of *GCG*, which was significantly upregulated in Cardenas-Diaz et al.'s data but downregulated in both our collaborators' and my own data. Within the *HNF1A* patient islet data, it was found that the donor islet alpha cells had impaired glucagon secretion in response to alpha cell stimuli such as low glucose, although the patient did have a higher proportion of alpha cells (Haliyur, et al., 2019). Similarly, patients with *HNF1A* mutations have been shown to have an inappropriate suppression of glucagon in response to an oral glucose tolerance test (Ostoft, et al., 2014). The differences in data may be due to subtle differences in differentiation protocol and differentiation efficiencies, which could affect the ratio of beta to alpha cells as well as specific feedback loops affecting glucagon.

5.3.5 Analysis demonstrates a significant overlap between *HNF1A* KO DEGs and *HNF1A* targets identified through an independent *HNF1A* knock-down experiment in the EndoC-βH1 cell line

I was also able to compare my results to an RNA-Seq dataset from an independent *HNF1A* knockdown experiment performed on the EndoC-βH1 line by Dr Antje Grotz (Gloyn Lab, OCDEM, University of Oxford). I saw a significant enrichment of my upregulated *HNF1A* KO DEGs and the upregulated DEGs in the EndoC-βH1 *HNF1A* KD ($p=0.00164$), as well as a significant enrichment of my downregulated *HNF1A* KO DEGs in the downregulated EndoC-βH1 *HNF1A* KD DEGs ($p=1.65x10^{-11}$) (Figure 5.3.16). However, certain DEGs from the *HNF1A* KO iPSCs, such as *NEUROD1* and *INS*, were not captured within the DEGs of the EndoC-βH1 *HNF1A* KD. This suggests that *HNF1A* may exert its effects on insulin secretion genes during the mid-stages of pancreatic development rather than the latter stages.

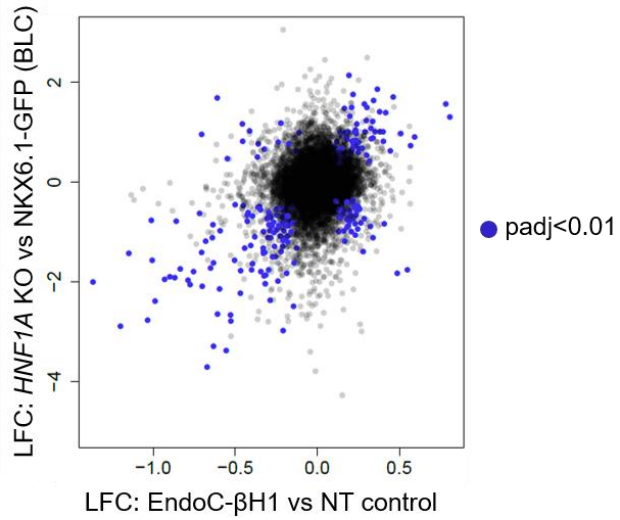


Figure 5.3.16 Significant enrichment of *HNF1A* KO DEGs at stage 7 (BLC) in the DEGs of the EndoC-βH1 line
There was a significant enrichment at stage 7 (BLC) of both my up- and downregulated *HNF1A* KO DEGs in the respective populations of the DEGs for the *HNF1A* KD in the EndoC-βH1 line. Data analysed using a hypergeometric enrichment test. LFC=log fold change, BLC=beta-like cells, NT=non-targeting. Genes with a padj<0.01 for both datasets are coloured blue

To further validate specific genes from the RNA-Seq, I carried out another independent *HNF1A* knockdown experiment in the EndoC-βH1, firstly assessing knockdown efficiency at both an mRNA and protein level (Figure 5.3.17). The subset of genes chosen were all downregulated in my *HNF1A* KO iPSC line, with TBP used as a control as it was not significant in my differential expression analysis at BLC (LFC=0.11, padj=0.611). Of the genes tested, *HNF1A, INS*, *DPP4* and *PAX4* were significantly downregulated in the independent EndoC-βH1 line (Figure 5.3.17). The results taken together, suggests that inactivation of *HNF1A* impacts endocrine cell development and that *HNF1A* has a functional role in both early and late stages of pancreatic development.

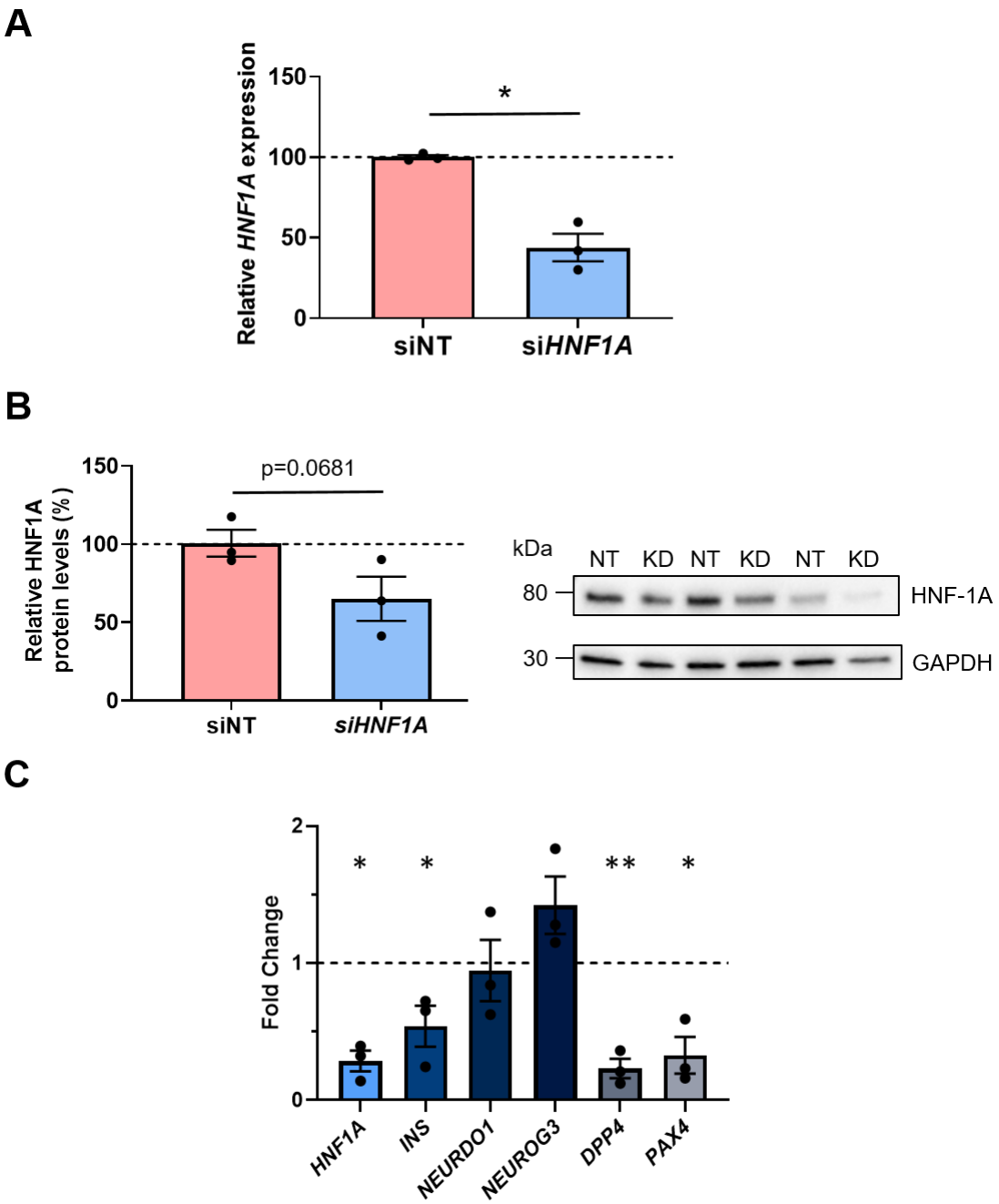


Figure 5.3.17 Knock-down (KD) of *HNF1A* in EndoC-βH1 cells
EndoC-βH1 cells were incubated with either non-targeting (NT) or si*HNF1A* oligos for 72hrs, with KD efficiency evaluated using (A) qRT-PCR and (B) western blotting to assess *HNF1A* mRNA expression and HNF1A protein levels, respectively, before testing of a subset of genes from the DEGs of the *HNF1A* KO iPSC line via RT-qPCR (C). Data are normalised to non-targeting (NT) transfected cells (dashed line) and are representative of 3 independent experiments. Plots display mean±SEM, *p<0.05, **p<0.01 for two-tailed unpaired Student's t-test with Welch's correction on transformed (log) data.

5.4 Discussion

This Chapter describes the characterisation of an *HNF1A* KO hiPSC line and its differentiation toward the pancreatic endocrine lineage. Previous studies of *HNF1A* have been hampered by mouse models not fully recapitulating the human phenotype and thus the recent emergence of hiPSC models has lent itself to modelling monogenic diseases, particularly with the success of modelling other haploinsufficient disease genes such as *GATA6* (Pontoglio, et al., 1998, Shi, et al., 2017). Within my *HNF1A* KO experiment, the expression of several essential developmental markers were evaluated in differentiated cells at three different distinct stages of *in vitro* beta cell differentiation, with *HNF1A* KO cells having decreased levels of glucagon and insulin at stage 7. Analysis of the lines showed that the *HNF1A* KO line had a significantly decreased expression of *HNF1A* at stage 7, leading to differences in expression of genes involved in beta cell development, MODY and insulin secretion.

There are several limitations of the hiPSC model described here that must be take into consideration. The differentiation efficiency can vary greatly between different hiPSC lines and from differentiation to differentiation. There can also be off-target effects that occur through the genome editing process that might affect the ability of individual clones to differentiate. Additionally, the cells do not reach functional maturity and so aspects of the human phenotype are unable to be fully recapitulated. Similarly, EndoC-βH1 are derived from human foetal pancreatic buds and may not possess full functional maturity, but do respond to glucose stimulus (Ravassard, et al., 2011). Furthermore, these cells have limited possibilities in terms of genetic modification due to their low clonal efficiency, although knockout lines are now possible (Grotz, et al., 2019). However, I performed a knockdown using siRNAs, rather than knockout with CRISPR/Cas9, in my EndoC-βH1 line and therefore there could be different compensation effects within the

two systems. Nevertheless, I was able to demonstrate that there was a significant overlap between my *HNF1A* KO hiPSC line and the independent EndoC-βH1 line, as well as showing an enrichment between my data and other datasets investigating *HNF1A* mutations. The overlap in results between the datasets validates the use of hiPSCs and genomic engineering as a system to model monogenic variants *in vitro*. However, the differences also show that utilising more than one model system is essential to be able to draw firmer conclusions on which targets of *HNF1A* should be focussed on for functional studies and to identify at which stages in development *HNF1A* may be exerting its effects. A future study could utilise an *HNF1A* KO EndoC-βH1 line in parallel to performing knockdown of *HNF1A* targets that arose from the RNA-Seq, such as *TRIM61* and *NNAT*, to investigate the subsequent effects on insulin secretion in response to different glucose levels. The *HNF1A* KO EndoC-βH1 line could also be used to confirm the targets identified from the knockdown model and to set-up a rescue cellular system. In addition, the recent advancements in differentiation protocols has allowed for a more mature phenotype of beta-like cells and could provide the opportunity to look at insulin secretion in an hiPSC model (Nair, et al., 2019, Velazco-Cruz, et al., 2019, Hogrebe, et al., 2020, Maxwell, et al., 2020).

Whilst RNA-Seq provides a view on the gene expression of the bulk cellular population, it can mask regulatory relationships between genes within individual cellular populations. Therefore, using single-cell RNA-Seq may facilitate uncovering cell specific pathways and genes that could be influenced by loss of *HNF1A* expression. As the *HNF1A* KO line was created from the NKX6.1-GFP reporter line, FACS could also enable the isolation of a more mature endocrine cell population for downstream analysis. An additional technique that could be incorporated in future experiments would be to visualise co-localisation of the expression of *HNF1A* and its target genes using RNAscope®

Technology, an *in situ* hybridization (ISH) assay for detection of target RNA within intact cells.

Overall, my *HNF1A* KO iPSC model recapitulates several features of the HNF1A-MODY patient phenotype such as disruption of insulin secretion and beta cell development genes. Although functional studies need to be performed to further unravel the mechanisms behind the effects of *HNF1A*, my data provides a valuable resource for studying cellular and molecular phenotypes caused by HNF1A deficiency.

6

Discussion

The overall aim of my thesis was to utilise human beta cell models to assess the impact of diabetes causing mutations in *HNF1A* and *INS* in beta cell development (*HNF1A*) and function (*HNF1A* and *INS*). Previous work has implicated mutations in both *HNF1A* and *INS* as a cause of monogenic diabetes but the mechanism behind this has yet to be completely determined (Stride, et al., 2005, Stoy, et al., 2007, Balboa, et al., 2018, Cardenas-Diaz, et al., 2019). Studies focussing on *HNF1A* have shown that mutations in the gene can cause impaired insulin secretion and beta cell dysfunction *in vitro* and *in vivo* but how *HNF1A* contributes to disease pathogenesis remains unclear (Cardenas-Diaz, et al., 2019, Haliyur, et al., 2019). Diabetes causing mutations within *INS* have been shown within mouse and rodent cell lines to cause misfolding of proinsulin, triggering accumulation of proinsulin within the ER, leading to upregulation of the unfolded protein response and beta cell dysfunction but the complete mechanism is yet to be fully elucidated (Wang, et al., 1999, Liu, et al., 2010, Cunningham, et al., 2017).

Chapter 3 described the introduction by genome-editing of the diabetes-causing p.C43G *INS* mutation in a human iPSC model with subsequent differentiation and assessment of markers of differentiation efficiency and ER stress. In parallel, I developed a pipeline for introduction of the same *INS* variant into a reporter iPSC line, enabling the isolation of cells representing a more mature beta cell phenotype. The results from my differentiation experiments showed no effect of the p.C43G cell line on levels of expression of *INS* or ER stress markers. Most likely this was due to the functional immaturity of the cells and inefficient differentiation. Previous work on an *INS* variant with a similar mechanism of action that did demonstrate an effect on ER stress utilised methods which produce more functionally mature cells, those of 3D-suspension culture and transplantation into immunocompromised mice (Rezania, et al., 2014, Balboa, et al., 2018, Xu, et al., 2019). Furthermore, this study utilised single-cell RNA-Seq to isolate the beta cell population,

which I could also incorporate into a future study. At the time of writing my thesis, a manuscript has been published reporting a new INSULIN-H2B-cherry reporter line, enabling FACS separation of insulin positive cells without impairment of insulin processing. Although the fluorescent protein is not localised with insulin, it can be viewed both via immunofluorescence and live cell imaging (Blochinger, et al., 2020). This technique would allow me to visualise where misfolded protein aggregates.

Another potential avenue for follow-up would be to undertake a small compound screen, utilising a library of FDA approved small molecules, to identify those which alleviate ER stress. As *INS* variants predominantly cause misfolded protein in the ER, I could employ cellular readouts of ER transmembrane proteins that sense unfolded protein accumulation in the ER: *PERK*, *ATF6*, and *IRE1* (Gardner, et al., 2013). Similarly, to assess the ability of the cell to respond to severe ER stress, I could measure genes that are upregulated in apoptotic conditions such as *DDIT3* (Li, et al., 2014). A previous high-throughput screen managed to demonstrate that ER stress can be alleviated through small molecule compounds such as azoromide, whereby ER protein folding is improved alongside induced ER chaperone capacity. This compound also resulted in improved fasting glucose in an obese rodent model (Fu, et al., 2015). These studies and others have utilised screens to investigate compounds that could alleviate ER stress and thus protect against beta cell death and help in the treatment of diabetes, which demonstrate both the feasibility and benefit of implementing this for future research. (Tran, et al., 2014, Duan, et al., 2017).

Following on from my hiPSC work, I wanted to explore the effect of *INS* variants in a more functionally mature cell model. To achieve this, I utilised a recently generated *INS*-KO EndoC-βH1 line, a human beta cell line, which avoided the confounding effects of endogenous insulin expression, as well as providing a more physiologically relevant cell line than those derived from rodents. This was the first study to use a human beta cell line

to investigate the effects of overexpressing an *INS* variant. To begin with, I had to establish a successful transfection system, which proved to be a challenge. The separate incorporation of two different fluorescent markers (one C-terminal, the other fused within C-peptide) within the *INS* sequence prevented the proper folding and packaging of the proinsulin molecule. Following this, expression of the C-terminal fluorescent marker was prevented by re-introduction of the stop codon within the *INS* sequence. Expression of both the mutant and wildtype constructs could be successfully detected via immunofluorescence but it was not possible to recapitulate the heterozygous genotype of patients as co-transfection could not be reliably achieved. Both the WT and p.C43G were successfully packaged and localised to the Chromogranin A positive vesicles, demonstrating that neither the WT nor mutant alone impaired trafficking of proinsulin. My findings could support the hypothesis that it is the interactions between WT and mutant proinsulin that cause impaired processing, rather than mutant-mutant interactions. Furthermore, it was evident that the transfection efficiency was suboptimal, with insulin unable to be detected via Western Blot. This meant that any effects as a result of the expression of the mutant construct were likely masked by the untransfected cells within the same culture. This was also reflected in the RNA-Seq analysis which did not demonstrate any enrichment of genes involved in pathways such as the unfolded protein response, as was seen with cells transfected with the known ER-stressor tunicamycin. Future approaches could include using a stable expression system that would allow for the enrichment of transfected or transduced cells through, for example, antibiotic selection. Electron microscopy could also be used to investigate the impact on the cellular ultrastructure, which has been shown previously with *INS* variants, such as with the Akita and Munich *Ins2*^{C95S} mutant mouse (Izumi, et al., 2003, Zuber, et al., 2004, Herbach, et al., 2007). If a fluorescent marker could be incorporated into the insulin sequence without

disruption of proinsulin folding, then this would allow for FACS of successfully transfected cells to be taken forward for downstream experiments such as RNA-Seq where comparisons could be made between transfected and untransfected cells. If a fluorescent marker could not be incorporated, then further optimisation of the transfection system could allow for several *INS* variants to be transfected to investigate a spectrum of clinical phenotypes.

In Chapter 5, I focussed on another gene involved in monogenic diabetes, *HNF1A*. Previous work has shown that mouse models of HNF1A-diabetes do not fully recapitulate the human phenotype, meaning that there may be human-specific functional effects of *HNF1A* that could be overlooked by focussing on mouse models (Pontoglio, et al., 1998, Boj, et al., 2010). Therefore, in the first instance, I utilised hiPSCs which had been genome-edited by collaborators to knockout the *HNF1A* gene. The cell lines I used were all compound heterozygous for different deletions of the gene. One of these cell lines was directed down the endocrine pancreatic lineage, as a method to interrogate the potential developmental effects of loss of *HNF1A* expression. I demonstrated that loss of *HNF1A* led to decreased expression of insulin and glucagon at stage 7 of differentiation. I then examined the differentially expressed genes between the *HNF1A* KO and WT line. I showed that the *HNF1A* KO line had an enrichment of differentially expressed genes involved in regulation of beta cell function, beta cell development and insulin secretion, supporting previous work on *HNF1A* (Pontoglio, et al., 1998, Servitja, et al., 2009). I also found that MODY genes were enriched among differentially expressed genes in the *HNF1A* KO lines at the beta-like cell stage. Following this, I compared my data to a previously published study and found a significant overlap between the downregulated genes from my *HNF1A* KO DEGs and the study's homozygous *HNF1A* KO data (Cardenas-Diaz, et al., 2019). I also evaluated my results with those from our lab and

collaborators at Novo Nordisk, who differentiated cells from a patient with an HNF1A-MODY causing variant. The datasets contained a significant number of common gene targets of *HNF1A*, demonstrating that differentiating cells with artificially introduced mutations can capture many of the same gene signatures that are seen when using patient-derived cells, which demonstrates that artificially introduced mutations are a practical alternative to patient-derived cells, which are limited in their availability. One notable gene that was different between data sets was that of glucagon, which was upregulated in Cardenas-Diaz et al.'s data but downregulated in the patient iPSC and my own data. A study on human islets from an individual with HNF1A-MODY showed that their alpha cells had impaired glucagon secretion but that they also had a higher proportion of alpha cells (Haliyur, et al., 2019). Within the clinic, patients with variations within *HNF1A* have also been shown to have impaired glucagon response to an oral glucose tolerance test (Ostoft, et al., 2014). The discrepancies between datasets could be due to influences of genetic background outside of *HNF1A* variations, and differences in differentiation protocol and efficiency. I also sought to validate my findings by comparing my differentiated cells to the more functionally mature EndoC-βH1 line. My analysis revealed a significant overlap between the data demonstrating that although differentiated iPSCs cells do not reach functional maturity, many of the same gene signatures can be captured, meaning that iPSCs still represent a reliable system for investigating patient-derived mutations. Another method that could be incorporated in future experiments would be to visualise co-localisation of the expression of *HNF1A* and its target genes using RNAscope® Technology, an *in situ* hybridization (ISH) assay for detection of target RNA within intact cells. A further approach could utilise an *HNF1A* KO EndoC-βH1 line to see if any downstream targets of *HNF1A* were sufficient to rescue the phenotypes observed, such as the decreased expression of *DPP4*. These gene targets could also have

KO lines generated, in both iPSCs and EndoC-βH1 which would further understanding of the specific role of each of these genes and their pathway(s). Future experiments could transplant differentiated iPSCs into immunocompromised mice, to enable maturation of beta cells that are glucose responsive. Mouse beta cells could then be ablated with streptozotocin to assess functional differences between patient and corrected cell lines.

Human iPSC approaches are still in the process of being optimised and thus there are various limitations that must be taken into consideration when evaluating results.

Differentiation efficiency can vary greatly between hiPSC lines of different genetic backgrounds, making it difficult to distinguish the effects of genetic variants in isolation, as cellular phenotypes may be markedly influenced by genetic differences between individuals (Rouhani, et al., 2014, Kyttala, et al., 2016). These differences have been somewhat addressed by the generation of isogenic cell lines, whereby correction of a patient-derived iPSCs line or introduction of a patient-specific variant in a control line can be performed using genome engineering. Even with this improvement, there can still be variation between differentiations. This could be a result of undetected off-target effects that alter the ability of particular clones to differentiate. Furthermore, these differences in differentiation efficiency may prevent detection of smaller sized effects of genetic variants that may be influencing differentiation. This could potentially cause technical variation to be wrongly interpreted as a biological effect. Recently, there have been numerous techniques established to help identify off-target effects associated with CRISPR/Cas9 editing, such as *in vitro* and cell-based assays (Kim, et al., 2015, Cameron, et al., 2017, Yan, et al., 2017). These allow off-target effects to be more readily assessed and thus can be taken into account when performing differentiation experiments. Another method that limits the generation of off-target effects is base editing, whereby a catalytically inactive Cas9 is used alongside cytidine deaminases to introduce site-

specific modifications of DNA without the need for a double-strand break (Komor, et al., 2016).

One of the main shortcomings with *in vitro* beta cell differentiation is the generation of heterogeneous polyhormonal endocrine cell populations (Pagliuca, et al., 2014, Rezania, et al., 2014, Perez-Alcantara, et al., 2018). Bulk analysis, such as transcriptome profiling, may therefore miss cell-specific effects. At stage 7 of my differentiations for Chapter 3 (*INS*) and Chapter 5 (*HNF1A*), the cells co-expressed C-peptide and glucagon, indicating a polyhormonal cell population. However, in Chapter 3, using a fluorescent marker of NKX6.1 expression allowed me to purify my mature cell population, giving a significant enrichment of insulin positive cells. A similar approach could be used with my *HNF1A* differentiation, alongside single-cell RNA-Seq, which would enable the investigation of specific cellular populations and overcome potential confounding effects resulting from heterogeneous cell populations (Ramond et al. 2017; Ramond et al. 2018; Petersen et al. 2017; Balboa et al. 2018). However, focussing on one cell type can be at the expense of gaining insight into cell-cell communication and the effects of such interaction, which is pertinent to beta cell function in the islet.

A potential future therapeutic avenue, particularly in the treatment of monogenic diabetes, could be the derivation of hiPSCs from diabetes patients, with CRISPR/Cas9 editing used to correct genetic mutations. These engineered hiPSCs can then be differentiated into beta-like cells and transplanted back into the individual, correcting diabetes. An ongoing Phase I/II clinical trial, albeit for type 1 diabetes, by the company ViaCyte (https://clinicaltrials.gov/ct2/show/NCT02239354) is investigating the safety and efficacy of transplanting pancreas progenitors derived from stem cells subcutaneously. Although using stem-cell derived beta cells therapeutically is in its infancy, this trial provides the potential to usher in the development of curative diabetes therapies.

Presently, *in vitro* differentiation protocols generate beta-like cells that are functionally immature, with their transcriptome more closely resembling foetal rather than adult beta cells (Bruin, et al., 2014, Hrvatin, et al., 2014). This means that the function of genes that regulate insulin secretion and metabolism cannot be modelled effectively in an hiPSC-based system. In this regard, the effects of variations within *INS* and *HNF1A* were assessed in the mature human beta cell line EndoC-βH1.

One strategy for investigating more mature beta cell function is transplanting beta-like cells under the kidney capsule of immunocompromised mice, which can even alleviate diabetic phenotypes (Kroon, et al., 2008, Pagliuca, et al., 2014, Rezania, et al., 2014, Ma, et al., 2018). This illustrates that there are additional factors in the *in vivo* environment that promote functional maturation and that *in vitro* differentiations may not be able to fully recapture diabetes pathogenesis. These drawbacks must be considered when translating findings to the clinic. Research groups are continually refining and improving *in vitro* differentiation protocols to overcome the shortcomings described that will certainly improve experimental reproducibility and functionality in years to come.

My work demonstrates, for the first time, modelling of the p.C43G mutation in two human beta cell models and the first time any *INS* variant has been investigated in the EndoC-βH1 line. Additionally, my *HNF1A* KO iPSC and *HNF1A* KD EndoC-βH1 model recapitulated several features of the HNF1A-MODY patient phenotype, such as disruption of insulin secretion and beta cell development genes. The results described in this thesis illustrate the value of utilising both iPSCs and EndoC-βH1 for disease modelling with each addressing different aspects of diabetes pathogenesis. The optimisation of these systems will facilitate the modelling of more complex and common forms of diabetes.

Bibliography

Abel, E.D., Peroni, O., Kim, J.K., Kim, Y.B., Boss, O., Hadro, E., et al. (2001) Adipose-selective targeting of the GLUT4 gene impairs insulin action in muscle and liver, *Nature* **409**: 729-733.

Akpinar, P., Kuwajima, S., Krutzfeldt, J., and Stoffel, M. (2005) Tmem27: a cleaved and shed plasma membrane protein that stimulates pancreatic beta cell proliferation, *Cell Metab* **2**: 385-397.

Alberti, K.G., and Zimmet, P.Z. (1998) Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation, *Diabet Med* **15**: 539-553.

Allen, H.L., Flanagan, S.E., Shaw-Smith, C., De Franco, E., Akerman, I., Caswell, R., et al. (2011) GATA6 haploinsufficiency causes pancreatic agenesis in humans, *Nat Genet* **44**: 20-22.

Almgren, P., Lehtovirta, M., Isomaa, B., Sarelin, L., Taskinen, M.R., Lyssenko, V., et al. (2011) Heritability and familiality of type 2 diabetes and related quantitative traits in the Botnia Study, *Diabetologia* **54**: 2811-2819.

Althari, S., Najmi, L.A., Bennett, A.J., Aukrust, I., Rundle, J.K., Colclough, K., et al. (2020) Unsupervised Clustering of Missense Variants in HNF1A Using Multidimensional Functional Data Aids Clinical Interpretation, *Am J Hum Genet* **107**: 670-682.

Amr, S., Heisey, C., Zhang, M., Xia, X.J., Shows, K.H., Ajlouni, K., et al. (2007) A homozygous mutation in a novel zinc-finger protein, ERIS, is responsible for Wolfram syndrome 2, *Am J Hum Genet* **81**: 673-683.

Andersson, L.E., Valtat, B., Bagge, A., Sharoyko, V.V., Nicholls, D.G., Ravassard, P., et al. (2015) Characterization of stimulus-secretion coupling in the human pancreatic EndoC-betaH1 beta cell line, *PLoS One* **10**: e0120879.

Andralojc, K.M., Mercalli, A., Nowak, K.W., Albarello, L., Calcagno, R., Luzi, L., et al. (2009) Ghrelin-producing epsilon cells in the developing and adult human pancreas, *Diabetologia* **52**: 486-493.

Aramata, S., Han, S.I., and Kataoka, K. (2007) Roles and regulation of transcription factor MafA in islet beta-cells, *Endocr J* **54**: 659-666.

Ashcroft, F.M., Harrison, D.E., and Ashcroft, S.J. (1984) Glucose induces closure of single potassium channels in isolated rat pancreatic beta-cells, *Nature* **312**: 446-448.

Ashcroft, F.M., and Rorsman, P. (1989) Electrophysiology of the pancreatic beta-cell, *Prog Biophys Mol Biol* **54**: 87-143.

Babenko, A.P., Polak, M., Cave, H., Busiah, K., Czernichow, P., Scharfmann, R., et al. (2006) Activating mutations in the ABCC8 gene in neonatal diabetes mellitus, *N Engl J Med* **355**: 456-466.

Baekkeskov, S., Aanstoot, H.J., Christgau, S., Reetz, A., Solimena, M., Cascalho, M., et al. (1990) Identification of the 64K autoantigen in insulin-dependent diabetes as the GABA-synthesizing enzyme glutamic acid decarboxylase, *Nature* **347**: 151-156.

Balboa, D., Saarimaki-Vire, J., Borshagovski, D., Survila, M., Lindholm, P., Galli, E., et al. (2018) Insulin mutations impair beta-cell development in a patient-derived iPSC model of neonatal diabetes, *Elife* **7**.

Ballian, N., and Brunicardi, F.C. (2007) Islet vasculature as a regulator of endocrine pancreas function, *World J Surg* **31**: 705-714.

Baron, M., Veres, A., Wolock, S.L., Faust, A.L., Gaujoux, R., Vetere, A., et al. (2016) A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure, *Cell Syst* **3**: 346-360 e344.

Barrett, T.G., Bundey, S.E., and Macleod, A.F. (1995) Neurodegeneration and diabetes: UK nationwide study of Wolfram (DIDMOAD) syndrome, *Lancet* **346**: 1458-1463.

Bell, G.I., Horita, S., and Karam, J.H. (1984) A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus, *Diabetes* **33**: 176-183.

Bell, G.I., Swain, W.F., Pictet, R., Cordell, B., Goodman, H.M., and Rutter, W.J. (1979) Nucleotide sequence of a cDNA clone encoding human preproinsulin, *Nature* **282**: 525-527.

Bellanne-Chantelot, C., Chauveau, D., Gautier, J.F., Dubois-Laforgue, D., Clauin, S., Beaufils, S., et al. (2004) Clinical spectrum associated with hepatocyte nuclear factor-1beta mutations, *Ann Intern Med* **140**: 510-517.
Benner, C., van der Meulen, T., Caceres, E., Tigyi, K., Donaldson, C.J., and Huising, M.O. (2014) The transcriptional landscape of mouse beta cells compared to human beta cells reveals notable species differences in long non-coding RNA and protein-coding gene expression, *BMC Genomics* **15**: 620.
Bennett, S.T., Wilson, A.J., Esposito, L., Bouzekri, N., Undlien, D.E., Cucca, F., et al. (1997) Insulin VNTR allele-specific effect in type 1 diabetes depends on identity of untransmitted paternal allele. The IMDIAB Group, *Nat Genet* **17**: 350-352.
Billings, L.K., and Florez, J.C. (2010) The genetics of type 2 diabetes: what have we learned from GWAS?, *Ann N Y Acad Sci* **1212**: 59-77.
Blochinger, A.K., Siehler, J., Wissmiller, K., Shahryari, A., Burtscher, I., and Lickert, H. (2020) Generation of an INSULIN-H2B-Cherry reporter human iPSC line, *Stem Cell Res* **45**: 101797.
Blum, B., Hrvatin, S., Schuetz, C., Bonal, C., Rezania, A., and Melton, D.A. (2012) Functional beta-cell maturation is marked by an increased glucose threshold and by expression of urocortin 3, *Nat Biotechnol* **30**: 261-264.
Boj, S.F., Petrov, D., and Ferrer, J. (2010) Epistasis of transcriptomes reveals synergism between transcriptional activators Hnf1alpha and Hnf4alpha, *PLoS Genet* **6**: e1000970.
Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S.D. (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin, *Microbiology (Reading)* **151**: 2551-2561.
Bonadonna, R.C., Saccomani, M.P., Cobelli, C., and DeFronzo, R.A. (1993) Effect of insulin on system A amino acid transport in human skeletal muscle, *J Clin Invest* **91**: 514-521.
Bonnefond, A., Philippe, J., Durand, E., Dechaume, A., Huyvaert, M., Montagne, L., et al. (2012) Whole-exome sequencing and high throughput genotyping identified KCNJ11 as the thirteenth MODY gene, *PLoS One* **7**: e37423.
Bonner, C., Bacon, S., Concannon, C.G., Rizvi, S.R., Baquie, M., Farrelly, A.M., et al. (2010) INS-1 cells undergoing caspase-dependent apoptosis enhance the regenerative capacity of neighboring cells, *Diabetes* **59**: 2799-2808.
Bouatia-Naji, N., Rocheleau, G., Van Lommel, L., Lemaire, K., Schuit, F., Cavalcanti-Proenca, C., et al. (2008) A polymorphism within the G6PC2 gene is associated with fasting plasma glucose levels, *Science* **320**: 1085-1088.
Bowman, P., Flanagan, S.E., Edghill, E.L., Damhuis, A., Shepherd, M.H., Paisey, R., et al. (2012) Heterozygous ABCC8 mutations are a cause of MODY, *Diabetologia* **55**: 123-127.
Bowman, P., Sulen, A., Barbetti, F., Beltrand, J., Svalastoga, P., Codner, E., et al. (2018) Effectiveness and safety of long-term treatment with sulfonylureas in patients with neonatal diabetes due to KCNJ11 mutations: an international cohort study, *Lancet Diabetes Endocrinol* **6**: 637-646.
Bramswig, N.C., Everett, L.J., Schug, J., Dorrell, C., Liu, C., Luo, Y., et al. (2013) Epigenomic plasticity enables human pancreatic alpha to beta cell reprogramming, *J Clin Invest* **123**: 1275-1284.
Braun, M., Ramracheya, R., Bengtsson, M., Zhang, Q., Karanauskaite, J., Partridge, C., et al. (2008) Voltage-gated ion channels in human pancreatic beta-cells: electrophysiological characterization and role in insulin secretion, *Diabetes* **57**: 1618-1628.
Braverman-Gross, C., Nudel, N., Ronen, D., Beer, N.L., McCarthy, M.I., and Benvenisty, N. (2018) Derivation and molecular characterization of pancreatic differentiated MODY1-iPSCs, *Stem Cell Res* **31**: 16-26.
Breschi, M.C., Seghieri, G., Bartolomei, G., Gironi, A., Baldi, S., and Ferrannini, E. (1993) Relation of birthweight to maternal plasma glucose and insulin concentrations during normal pregnancy, *Diabetologia* **36**: 1315-1321.
Brinkman, E.K., Chen, T., Amendola, M., and van Steensel, B. (2014) Easy quantitative assessment of genome editing by sequence trace decomposition, *Nucleic Acids Res* **42**: e168.

Brissova, M., Fowler, M.J., Nicholson, W.E., Chu, A., Hirshberg, B., Harlan, D.M., and Powers, A.C. (2005) Assessment of human pancreatic islet architecture and composition by laser scanning confocal microscopy, *J Histochem Cytochem* **53**: 1087-1097.
Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., et al. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes, *Science* **321**: 960-964.
Bruin, J.E., Erener, S., Vela, J., Hu, X., Johnson, J.D., Kurata, H.T., et al. (2014) Characterization of polyhormonal insulin-producing cells derived in vitro from human embryonic stem cells, *Stem Cell Res* **12**: 194-208.
Butler, A.E., Janson, J., Soeller, W.C., and Butler, P.C. (2003) Increased beta-cell apoptosis prevents adaptive increase in beta-cell mass in mouse model of type 2 diabetes: evidence for role of islet amyloid formation rather than direct action of amyloid, *Diabetes* **52**: 2304-2314.
Cabrera, O., Berman, D.M., Kenyon, N.S., Ricordi, C., Berggren, P.O., and Caicedo, A. (2006) The unique cytoarchitecture of human pancreatic islets has implications for islet cell function, *Proc Natl Acad Sci U S A* **103**: 2334-2339.
Calatayud, C., Carola, G., Fernandez-Carasa, I., Valtorta, M., Jimenez-Delgado, S., Diaz, M., et al. (2019) CRISPR/Cas9-mediated generation of a tyrosine hydroxylase reporter iPSC line for live imaging and isolation of dopaminergic neurons, *Sci Rep* **9**: 6811.
Cameron, P., Fuller, C.K., Donohoue, P.D., Jones, B.N., Thompson, M.S., Carter, M.M., et al. (2017) Mapping the genomic landscape of CRISPR-Cas9 cleavage, *Nat Methods* **14**: 600-606.
Cardenas-Diaz, F.L., Osorio-Quintero, C., Diaz-Miranda, M.A., Kishore, S., Leavens, K., Jobaliya, C., et al. (2019) Modeling Monogenic Diabetes using Human ESCs Reveals Developmental and Metabolic Deficiencies Caused by Mutations in HNF1A, *Cell Stem Cell* **25**: 273-289 e275.
Carpenter, M.C., and Wilcox, D.E. (2014) Thermodynamics of formation of the insulin hexamer: metal-stabilized proton-coupled assembly of quaternary structure, *Biochemistry* **53**: 1296-1301.
Chan, S.J., Seino, S., Gruppuso, P.A., Schwartz, R., and Steiner, D.F. (1987) A mutation in the B chain coding region is associated with impaired proinsulin conversion in a family with hyperproinsulinemia, *Proc Natl Acad Sci U S A* **84**: 2194-2197.
Chandra, V., Albagli-Curiel, O., Hastoy, B., Piccand, J., Randriamampita, C., Vaillant, E., et al. (2014) RFX6 regulates insulin secretion by modulating Ca2+ homeostasis in human beta cells, *Cell Rep* **9**: 2206-2218.
Chang, S.G., Choi, K.D., Jang, S.H., and Shin, H.C. (2003) Role of disulfide bonds in the structure and activity of human insulin, *Mol Cells* **16**: 323-330.
Chang, T.J., Chiu, Y.F., Sheu, W.H., Shih, K.C., Hwu, C.M., Quertermous, T., et al. (2015) Genetic polymorphisms of PCSK2 are associated with glucose homeostasis and progression to type 2 diabetes in a Chinese population, *Sci Rep* **5**: 14380.
Chimienti, F., Devergnas, S., Favier, A., and Seve, M. (2004) Identification and cloning of a beta-cell-specific zinc transporter, ZnT-8, localized into insulin secretory granules, *Diabetes* **53**: 2330-2337.
Cho, Y.S., Chen, C.H., Hu, C., Long, J., Ong, R.T., Sim, X., et al. (2011) Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians, *Nat Genet* **44**: 67-72.
Choeiri, C., Hewitt, K., Durkin, J., Simard, C.J., Renaud, J.M., and Messier, C. (2005) Longitudinal evaluation of memory performance and peripheral neuropathy in the Ins2C96Y Akita mice, *Behav Brain Res* **157**: 31-38.
Clocquet, A.R., Egan, J.M., Stoffers, D.A., Muller, D.C., Wideman, L., Chin, G.A., et al. (2000) Impaired insulin secretion and increased insulin sensitivity in familial maturity-onset diabetes of the young 4 (insulin promoter factor 1 gene), *Diabetes* **49**: 1856-1864.
Cnop, M., Foufelle, F., and Velloso, L.A. (2012) Endoplasmic reticulum stress, obesity and diabetes, *Trends Mol Med* **18**: 59-68.
Colclough, K., Bellanne-Chantelot, C., Saint-Martin, C., Flanagan, S.E., and Ellard, S. (2013) Mutations in the genes encoding the transcription factors hepatocyte nuclear factor 1 alpha and 4 alpha in maturity-onset diabetes of the young and hyperinsulinemic hypoglycemia, *Hum Mutat* **34**: 669-685.

Concannon, P., Chen, W.M., Julier, C., Morahan, G., Akolkar, B., Erlich, H.A., et al. (2009) Genome-wide scan for linkage to type 1 diabetes in 2,496 multiplex families from the Type 1 Diabetes Genetics Consortium, *Diabetes* **58**: 1018-1022.
Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., et al. (2013) Multiplex genome engineering using CRISPR/Cas systems, *Science* **339**: 819-823.
Consortium, S.T.D., Estrada, K., Aukrust, I., Bjorkhaug, L., Burtt, N.P., Mercader, J.M., et al. (2014) Association of a low-frequency variant in HNF1A with type 2 diabetes in a Latino population, *JAMA* **311**: 2305-2314.
Cooper, J.D., Howson, J.M., Smyth, D., Walker, N.M., Stevens, H., Yang, J.H., et al. (2012) Confirmation of novel type 1 diabetes risk loci in families, *Diabetologia* **55**: 996-1000.
Coore, H.G., and Randle, P.J. (1964) Regulation of insulin secretion studied with pieces of rabbit pancreas incubated in vitro, *Biochem J* **93**: 66-78.
Credle, J.J., Finer-Moore, J.S., Papa, F.R., Stroud, R.M., and Walter, P. (2005) On the mechanism of sensing unfolded protein in the endoplasmic reticulum, *Proc Natl Acad Sci U S A* **102**: 18773-18784.
Cudworth, A.G., and Woodrow, J.C. (1975) Evidence for HL-A-linked genes in "juvenile" diabetes mellitus, *Br Med J* **3**: 133-135.
Cunningham, C.N., He, K., Arunagiri, A., Paton, A.W., Paton, J.C., Arvan, P., and Tsai, B. (2017) Chaperone-Driven Degradation of a Misfolded Proinsulin Mutant in Parallel With Restoration of Wild-Type Insulin Secretion, *Diabetes* **66**: 741-753.
Cunningham, C.N., Williams, J.M., Knupp, J., Arunagiri, A., Arvan, P., and Tsai, B. (2019) Cells Deploy a Two-Pronged Strategy to Rectify Misfolded Proinsulin Aggregates, *Mol Cell* **75**: 442-456 e444.
D'Amour, K.A., Bang, A.G., Eliazer, S., Kelly, O.G., Agulnick, A.D., Smart, N.G., et al. (2006) Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells, *Nat Biotechnol* **24**: 1392-1401.
D'Angelo, A., Bluteau, O., Garcia-Gonzalez, M.A., Gresh, L., Doyen, A., Garbay, S., et al. (2010) Hepatocyte nuclear factor 1alpha and beta control terminal differentiation and cell fate commitment in the gut epithelium, *Development* **137**: 1573-1582.
Daniel, S., Noda, M., Straub, S.G., and Sharp, G.W. (1999) Identification of the docked granule pool responsible for the first phase of glucose-stimulated insulin secretion, *Diabetes* **48**: 1686-1690.
Davidson, H.W., and Hutton, J.C. (1987) The insulin-secretory-granule carboxypeptidase H. Purification and demonstration of involvement in proinsulin processing, *Biochem J* **245**: 575-582.
Davidson, H.W., Peshavaria, M., and Hutton, J.C. (1987) Proteolytic conversion of proinsulin into insulin. Identification of a Ca2+-dependent acidic endopeptidase in isolated insulin-secretory granules, *Biochem J* **246**: 279-286.
Davies, M.J., D'Alessio, D.A., Fradkin, J., Kernan, W.N., Mathieu, C., Mingrone, G., et al. (2018) Management of Hyperglycemia in Type 2 Diabetes, 2018. A Consensus Report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD), *Diabetes Care* **41**: 2669-2701.
De Franco, E., Flanagan, S.E., Houghton, J.A., Lango Allen, H., Mackay, D.J., Temple, I.K., et al. (2015) The effect of early, comprehensive genomic testing on clinical care in neonatal diabetes: an international cohort study, *Lancet* **386**: 957-963.
De Franco, E., Flanagan, S.E., Yagi, T., Abreu, D., Mahadevan, J., Johnson, M.B., et al. (2017) Dominant ER Stress-Inducing WFS1 Mutations Underlie a Genetic Syndrome of Neonatal/Infancy-Onset Diabetes, Congenital Sensorineural Deafness, and Congenital Cataracts, *Diabetes* **66**: 2044-2053.
De Franco, E., Lytrivi, M., Ibrahim, H., Montaser, H., Wakeling, M.N., Fantuzzi, F., et al. (2020) YIPF5 mutations cause neonatal diabetes and microcephaly through endoplasmic reticulum stress, *J Clin Invest* **130**: 6338-6353.
De Franco, E., Shaw-Smith, C., Flanagan, S.E., Edghill, E.L., Wolf, J., Otte, V., et al. (2013) Biallelic PDX1 (insulin promoter factor 1) mutations causing neonatal diabetes without exocrine pancreatic insufficiency, *Diabet Med* **30**: e197-200.

De Franco, E., Shaw-Smith, C., Flanagan, S.E., Shepherd, M.H., International, N.D.M.C., Hattersley, A.T., and Ellard, S. (2013) GATA6 mutations cause a broad phenotypic spectrum of diabetes from pancreatic agenesis to adult-onset diabetes without exocrine insufficiency, *Diabetes* **62**: 993-997.
De Vos, A., Heimberg, H., Quartier, E., Huypens, P., Bouwens, L., Pipeleers, D., and Schuit, F. (1995) Human and rat beta cells differ in glucose transporter but not in glucokinase gene expression, *J Clin Invest* **96**: 2489-2495.
Defronzo, R.A. (1979) Glucose intolerance and aging: evidence for tissue insensitivity to insulin, *Diabetes* **28**: 1095-1101.
Delepine, M., Nicolino, M., Barrett, T., Golamaully, M., Lathrop, G.M., and Julier, C. (2000) EIF2AK3, encoding translation initiation factor 2-alpha kinase 3, is mutated in patients with Wolcott-Rallison syndrome, *Nat Genet* **25**: 406-409.
Demirbilek, H., Hatipoglu, N., Gul, U., Tatli, Z.U., Ellard, S., Flanagan, S.E., et al. (2018) Permanent neonatal diabetes mellitus and neurological abnormalities due to a novel homozygous missense mutation in NEUROD1, *Pediatr Diabetes* **19**: 898-904.
Deriy, L.V., Gomez, E.A., Jacobson, D.A., Wang, X., Hopson, J.A., Liu, X.Y., et al. (2009) The granular chloride channel ClC-3 is permissive for insulin secretion, *Cell Metab* **10**: 316-323.
Dimas, A.S., Lagou, V., Barker, A., Knowles, J.W., Magi, R., Hivert, M.F., et al. (2014) Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity, *Diabetes* **63**: 2158-2171.
Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013) STAR: ultrafast universal RNA-seq aligner, *Bioinformatics* **29**: 15-21.
Dodson, G., and Steiner, D. (1998) The role of assembly in insulin's biosynthesis, *Curr Opin Struct Biol* **8**: 189-194.
Dolensek, J., Rupnik, M.S., and Stozer, A. (2015) Structural similarities and differences between the human and the mouse pancreas, *Islets* **7**: e1024405.
Drucker, D.J., Philippe, J., Mojsov, S., Chick, W.L., and Habener, J.F. (1987) Glucagon-like peptide I stimulates insulin gene expression and increases cyclic AMP levels in a rat islet cell line, *Proc Natl Acad Sci U S A* **84**: 3434-3438.
Duan, H., Li, Y., Arora, D., Xu, D., Lim, H.Y., and Wang, W. (2017) Discovery of a Benzamide Derivative That Protects Pancreatic beta-Cells against Endoplasmic Reticulum Stress, *J Med Chem* **60**: 6191-6204.
Dukes, I.D., Sreenan, S., Roe, M.W., Levisetti, M., Zhou, Y.P., Ostrega, D., et al. (1998) Defective pancreatic beta-cell glycolytic signaling in hepatocyte nuclear factor-1alpha-deficient mice, *J Biol Chem* **273**: 24457-24464.
Dwivedi, O.P., Lehtovirta, M., Hastoy, B., Chandra, V., Krentz, N.A.J., Kleiner, S., et al. (2019) Loss of ZnT8 function protects against diabetes by enhanced insulin secretion, *Nat Genet* **51**: 1596-1606.
Edghill, E.L., Bingham, C., Ellard, S., and Hattersley, A.T. (2006) Mutations in hepatocyte nuclear factor-1beta and their related phenotypes, *J Med Genet* **43**: 84-90.
Edghill, E.L., Dix, R.J., Flanagan, S.E., Bingley, P.J., Hattersley, A.T., Ellard, S., and Gillespie, K.M. (2006) HLA genotyping supports a nonautoimmune etiology in patients diagnosed with diabetes under the age of 6 months, *Diabetes* **55**: 1895-1898.
Edghill, E.L., Flanagan, S.E., Patch, A.M., Boustred, C., Parrish, A., Shields, B., et al. (2008) Insulin mutation screening in 1,044 patients with diabetes: mutations in the INS gene are a common cause of neonatal diabetes but a rare cause of diabetes diagnosed in childhood or adulthood, *Diabetes* **57**: 1034-1042.
Ellard, S., Bellanne-Chantelot, C., Hattersley, A.T., and European Molecular Genetics Quality Network, M.g. (2008) Best practice guidelines for the molecular genetic diagnosis of maturity-onset diabetes of the young, *Diabetologia* **51**: 546-553.
Ellard, S., and Colclough, K. (2006) Mutations in the genes encoding the transcription factors hepatocyte nuclear factor 1 alpha (HNF1A) and 4 alpha (HNF4A) in maturity-onset diabetes of the young, *Hum Mutat* **27**: 854-869.

Emdin, S.O., Dodson, G.G., Cutfield, J.M., and Cutfield, S.M. (1980) Role of zinc in insulin biosynthesis. Some possible zinc-insulin interactions in the pancreatic B-cell, *Diabetologia* **19**: 174-182.
Evangelou, M., Smyth, D.J., Fortune, M.D., Burren, O.S., Walker, N.M., Guo, H., et al. (2014) A method for gene-based pathway analysis using genomewide association study summary statistics reveals nine new type 1 diabetes associations, *Genet Epidemiol* **38**: 661-670.
Evans, M.J., and Kaufman, M.H. (1981) Establishment in culture of pluripotential cells from mouse embryos, *Nature* **292**: 154-156.
Faguer, S., Decramer, S., Chassaing, N., Bellanne-Chantelot, C., Calvas, P., Beaufils, S., et al. (2011) Diagnosis, management, and prognosis of HNF1B nephropathy in adulthood, *Kidney Int* **80**: 768-776.
Fajans, S.S., Bell, G.I., Paz, V.P., Below, J.E., Cox, N.J., Martin, C., et al. (2010) Obesity and hyperinsulinemia in a family with pancreatic agenesis and MODY caused by the IPF1 mutation Pro63fsX60, *Transl Res* **156**: 7-14.
Fang, J., Liu, M., Zhang, X., Sakamoto, T., Taatjes, D.J., Jena, B.P., et al. (2015) COPII-Dependent ER Export: A Critical Component of Insulin Biogenesis and beta-Cell ER Homeostasis, *Mol Endocrinol* **29**: 1156-1169.
Farmer, G., Russell, G., Hamilton-Nicol, D.R., Ogenbede, H.O., Ross, I.S., Pearson, D.W., et al. (1988) The influence of maternal glucose metabolism on fetal growth, development and morbidity in 917 singleton pregnancies in nondiabetic women, *Diabetologia* **31**: 134-141.
Fex, M., and Mulder, H. (2008) Lipases in the pancreatic beta-cell: implications for insulin secretion, *Biochem Soc Trans* **36**: 885-890.
Fiedorowicz, K., Rozwadowska, N., Zimna, A., Malcher, A., Tutak, K., Szczerbal, I., et al. (2020) Tissue-specific promoter-based reporter system for monitoring cell differentiation from iPSCs to cardiomyocytes, *Sci Rep* **10**: 1895.
Flanagan, S.E., Edghill, E.L., Gloyn, A.L., Ellard, S., and Hattersley, A.T. (2006) Mutations in KCNJ11, which encodes Kir6.2, are a common cause of diabetes diagnosed in the first 6 months of life, with the phenotype determined by genotype, *Diabetologia* **49**: 1190-1197.
Flanagan, S.E., Patch, A.M., Mackay, D.J., Edghill, E.L., Gloyn, A.L., Robinson, D., et al. (2007) Mutations in ATP-sensitive K+ channel genes cause transient neonatal diabetes and permanent diabetes in childhood or adulthood, *Diabetes* **56**: 1930-1937.
Flannick, J., Mercader, J.M., Fuchsberger, C., Udler, M.S., Mahajan, A., Wessel, J., et al. (2019) Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls, *Nature* **570**: 71-76.
Flannick, J., Thorleifsson, G., Beer, N.L., Jacobs, S.B., Grarup, N., Burtt, N.P., et al. (2014) Loss-of-function mutations in SLC30A8 protect against type 2 diabetes, *Nat Genet* **46**: 357-363.
Fonseca, S.G., Fukuma, M., Lipson, K.L., Nguyen, L.X., Allen, J.R., Oka, Y., and Urano, F. (2005) WFS1 is a novel component of the unfolded protein response and maintains homeostasis of the endoplasmic reticulum in pancreatic beta-cells, *J Biol Chem* **280**: 39609-39615.
Fonseca, S.G., Ishigaki, S., Oslowski, C.M., Lu, S., Lipson, K.L., Ghosh, R., et al. (2010) Wolfram syndrome 1 gene negatively regulates ER stress signaling in rodent and human cells, *J Clin Invest* **120**: 744-755.
Frayling, T.M., Bulman, M.P., Appleton, M., Hattersley, A.T., and Ellard, S. (1997) A rapid screening method for hepatocyte nuclear factor 1 alpha frameshift mutations; prevalence in maturity-onset diabetes of the young and late-onset non-insulin dependent diabetes, *Hum Genet* **101**: 351-354.
Friedrichsen, M., Birk, J.B., Richter, E.A., Ribel-Madsen, R., Pehmoller, C., Hansen, B.F., et al. (2013) Akt2 influences glycogen synthase activity in human skeletal muscle through regulation of NH(2)-terminal (sites 2 + 2a) phosphorylation, *Am J Physiol Endocrinol Metab* **304**: E631-639.
Froguel, P., Vaxillaire, M., Sun, F., Velho, G., Zouali, H., Butel, M.O., et al. (1992) Close linkage of glucokinase locus on chromosome 7p to early-onset non-insulin-dependent diabetes mellitus, *Nature* **356**: 162-164.
Fu, D., Cong, X., Ma, Y., Cai, H., Cai, M., Li, D., et al. (2013) Genetic polymorphism of glucokinase on the risk of type 2 diabetes and impaired glucose regulation: evidence based on 298,468 subjects, *PLoS One* **8**: e55727.

Fu, S., Yalcin, A., Lee, G.Y., Li, P., Fan, J., Arruda, A.P., et al. (2015) Phenotypic assays identify azoramide as a small-molecule modulator of the unfolded protein response with antidiabetic activity, *Sci Transl Med* **7**: 292ra298.
Fuchsberger, C., Flannick, J., Teslovich, T.M., Mahajan, A., Agarwala, V., Gaulton, K.J., et al. (2016) The genetic architecture of type 2 diabetes, *Nature* **536**: 41-47.
Furuta, M., Carroll, R., Martin, S., Swift, H.H., Ravazzola, M., Orci, L., and Steiner, D.F. (1998) Incomplete processing of proinsulin to insulin accompanied by elevation of Des-31,32 proinsulin intermediates in islets of mice lacking active PC2, *J Biol Chem* **273**: 3431-3437.
Gardner, B.M., Pincus, D., Gotthardt, K., Gallagher, C.M., and Walter, P. (2013) Endoplasmic reticulum stress sensing in the unfolded protein response, *Cold Spring Harb Perspect Biol* **5**: a013169.
Garin, I., Edghill, E.L., Akerman, I., Rubio-Cabezas, O., Rica, I., Locke, J.M., et al. (2010) Recessive mutations in the INS gene result in neonatal diabetes through reduced insulin biosynthesis, *Proc Natl Acad Sci U S A* **107**: 3105-3110.
Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., et al. (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA, *Nature* **468**: 67-71.
Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria, *Proc Natl Acad Sci U S A* **109**: E2579-2586.
Gaulton, K.J., Ferreira, T., Lee, Y., Raimondo, A., Magi, R., Reschen, M.E., et al. (2015) Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci, *Nat Genet* **47**: 1415-1425.
Gaulton, K.J., Nammo, T., Pasquali, L., Simon, J.M., Giresi, P.G., Fogarty, M.P., et al. (2010) A map of open chromatin in human pancreatic islets, *Nat Genet* **42**: 255-259.
Gepts, W. (1965) Pathologic anatomy of the pancreas in juvenile diabetes mellitus, *Diabetes* **14**: 619-633.
Gerich, J.E. (1993) Control of glycaemia, *Baillieres Clin Endocrinol Metab* **7**: 551-586.
Glaser, B. (2008) Insulin mutations in diabetes: the clinical spectrum, *Diabetes* **57**: 799-800.
Gloyn, A.L., Pearson, E.R., Antcliff, J.F., Proks, P., Bruining, G.J., Slingerland, A.S., et al. (2004) Activating mutations in the gene encoding the ATP-sensitive potassium-channel subunit Kir6.2 and permanent neonatal diabetes, *N Engl J Med* **350**: 1838-1849.
Gloyn, A.L., Weedon, M.N., Owen, K.R., Turner, M.J., Knight, B.A., Hitman, G., et al. (2003) Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes, *Diabetes* **52**: 568-572.
Glucksmann, M.A., Lehto, M., Tayber, O., Scotti, S., Berkemeier, L., Pulido, J.C., et al. (1997) Novel mutations and a mutational hotspot in the MODY3 gene, *Diabetes* **46**: 1081-1086.
Gonsorcikova, L., Pruhova, S., Cinek, O., Ek, J., Pelikanova, T., Jorgensen, T., et al. (2008) Autosomal inheritance of diabetes in two families characterized by obesity and a novel H241Q mutation in NEUROD1, *Pediatr Diabetes* **9**: 367-372.
Goodge, K.A., and Hutton, J.C. (2000) Translational regulation of proinsulin biosynthesis and proinsulin conversion in the pancreatic beta-cell, *Semin Cell Dev Biol* **11**: 235-242.
Goodison, S., Kenna, S., and Ashcroft, S.J. (1992) Control of insulin gene expression by glucose, *Biochem J* **285 (Pt 2)**: 563-568.
Gormsen, L.C., Sondergaard, E., Christensen, N.L., Brosen, K., Jessen, N., and Nielsen, S. (2019) Metformin increases endogenous glucose production in non-diabetic individuals and individuals with recent-onset type 2 diabetes, *Diabetologia* **62**: 1251-1256.
Grasso, S., Palumbo, G., Rugolo, S., Cianci, A., Tumino, G., and Reitano, G. (1980) Human fetal insulin secretion in response to maternal glucose and leucine administration, *Pediatr Res* **14**: 782-783.
Greeley, S.A., Naylor, R.N., Philipson, L.H., and Bell, G.I. (2011) Neonatal diabetes: an expanding list of genes allows for improved diagnosis and treatment, *Curr Diab Rep* **11**: 519-532.
Greeley, S.A.W., Letourneau, L.R., and Philipson, L.H. (2018) Precision medicine in KCNJ11 permanent neonatal diabetes, *Lancet Diabetes Endocrinol* **6**: 594-595.

Gribble, F.M., and Ashcroft, F.M. (1999) Differential sensitivity of beta-cell and extrapancreatic K(ATP) channels to gliclazide, *Diabetologia* **42**: 845-848.
Grobarczyk, B., Franco, B., Hanon, K., and Malgrange, B. (2015) Generation of Isogenic Human iPS Cell Line Precisely Corrected by Genome Editing Using the CRISPR/Cas9 System, *Stem Cell Rev Rep* **11**: 774-787.
Gross, D.J., Halban, P.A., Kahn, C.R., Weir, G.C., and Villa-Komaroff, L. (1989) Partial diversion of a mutant proinsulin (B10 aspartic acid) from the regulated to the constitutive secretory pathway in transfected AtT-20 cells, *Proc Natl Acad Sci U S A* **86**: 4107-4111.
Grotz, A.K., Abaitua, F., Navarro-Guerrero, E., Hastoy, B., Ebner, D., and Gloyn, A.L. (2019) A CRISPR/Cas9 genome editing pipeline in the EndoC-betaH1 cell line to study genes implicated in beta cell function, *Wellcome Open Res* **4**: 150.
Guest, P.C., Rhodes, C.J., and Hutton, J.C. (1989) Regulation of the biosynthesis of insulin-secretory-granule proteins. Co-ordinate translational control is exerted on some, but not all, granule matrix constituents, *Biochem J* **257**: 431-437.
Gunn, T., Bortolussi, R., Little, J.M., Andermann, F., Fraser, F.C., and Belmonte, M.M. (1976) Juvenile diabetes mellitus, optic atrophy, sensory nerve deafness, and diabetes insipidus--a syndrome, *J Pediatr* **89**: 565-570.
Gupta, S.K., Wesolowska-Andersen, A., Ringgaard, A.K., Jaiswal, H., Song, L., Hastoy, B., et al. (2018) NKX6.1 induced pluripotent stem cell reporter lines for isolation and analysis of functionally relevant neuronal and pancreas populations, *Stem Cell Res* **29**: 220-231.
Gurgul-Convey, E., Kaminski, M.T., and Lenzen, S. (2015) Physiological characterization of the human EndoC-betaH1 beta-cell line, *Biochem Biophys Res Commun* **464**: 13-19.
Haataja, L., Manickam, N., Soliman, A., Tsai, B., Liu, M., and Arvan, P. (2016) Disulfide Mispairing During Proinsulin Folding in the Endoplasmic Reticulum, *Diabetes* **65**: 1050-1060.
Hagenfeldt-Johansson, K.A., Herrera, P.L., Wang, H., Gjinovci, A., Ishihara, H., and Wollheim, C.B. (2001) Beta-cell-targeted expression of a dominant-negative hepatocyte nuclear factor-1 alpha induces a maturity-onset diabetes of the young (MODY)3-like phenotype in transgenic mice, *Endocrinology* **142**: 5311-5320.
Haliyur, R., Tong, X., Sanyoura, M., Shrestha, S., Lindner, J., Saunders, D.C., et al. (2019) Human islets expressing HNF1A variant have defective beta cell transcriptional regulatory networks, *J Clin Invest* **129**: 246-251.
Hansson, M., Tonning, A., Frandsen, U., Petri, A., Rajagopal, J., Englund, M.C., et al. (2004) Artifactual insulin release from differentiated embryonic stem cells, *Diabetes* **53**: 2603-2609.
Harding, H.P., Zeng, H., Zhang, Y., Jungries, R., Chung, P., Plesken, H., et al. (2001) Diabetes mellitus and exocrine pancreatic dysfunction in perk-/- mice reveals a role for translational control in secretory cell survival, *Mol Cell* **7**: 1153-1163.
Harding, H.P., Zhang, Y., Bertolotti, A., Zeng, H., and Ron, D. (2000) Perk is essential for translational regulation and cell survival during the unfolded protein response, *Mol Cell* **5**: 897-904.
Harries, L.W., Brown, J.E., and Gloyn, A.L. (2009) Species-specific differences in the expression of the HNF1A, HNF1B and HNF4A genes, *PLoS One* **4**: e7855.
Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., et al. (2012) GENCODE: the reference human genome annotation for The ENCODE Project, *Genome Res* **22**: 1760-1774.
Hart, N.J., and Powers, A.C. (2019) Use of human islets to understand islet biology and diabetes: progress, challenges and suggestions, *Diabetologia* **62**: 212-222.
Hart, T., Tong, A.H.Y., Chan, K., Van Leeuwen, J., Seetharaman, A., Aregger, M., et al. (2017) Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens, *G3 (Bethesda)* **7**: 2719-2727.
Hartley, T., Siva, M., Lai, E., Teodoro, T., Zhang, L., and Volchuk, A. (2010) Endoplasmic reticulum stress response in an INS-1 pancreatic beta-cell line with inducible expression of a folding-deficient proinsulin, *BMC Cell Biol* **11**: 59.
Hastoy, B., Godazgar, M., Clark, A., Nylander, V., Spiliotis, I., van de Bunt, M., et al. (2018) Electrophysiological properties of human beta-cell lines EndoC-betaH1 and -betaH2 conform with human beta-cells, *Sci Rep* **8**: 16994.

Hattersley, A., Bruining, J., Shield, J., Njolstad, P., Donaghue, K., International Society for, P., and Adolescent, D. (2006) ISPAD Clinical Practice Consensus Guidelines 2006-2007. The diagnosis and management of monogenic diabetes in children, *Pediatr Diabetes* **7**: 352-360.
Hattersley, A.T., Beards, F., Ballantyne, E., Appleton, M., Harvey, R., and Ellard, S. (1998) Mutations in the glucokinase gene of the fetus result in reduced birth weight, *Nat Genet* **19**: 268-270.
Hattersley, A.T., and Tooke, J.E. (1999) The fetal insulin hypothesis: an alternative explanation of the association of low birthweight with diabetes and vascular disease, *Lancet* **353**: 1789-1792.
Hattersley, A.T., Turner, R.C., Permutt, M.A., Patel, P., Tanizawa, Y., Chiu, K.C., et al. (1992) Linkage of type 2 diabetes to the glucokinase gene, *Lancet* **339**: 1307-1310.
Hauge-Evans, A.C., King, A.J., Carmignac, D., Richardson, C.C., Robinson, I.C., Low, M.J., et al. (2009) Somatostatin secreted by islet delta-cells fulfills multiple roles as a paracrine regulator of islet function, *Diabetes* **58**: 403-411.
Hay, C.W., and Docherty, K. (2006) Comparative analysis of insulin gene promoters: implications for diabetes research, *Diabetes* **55**: 3201-3213.
He, X.J., Qin, F.Y., Hu, C.L., Zhu, M., Tian, C.Q., and Li, L. (2015) Is gestational diabetes mellitus an independent risk factor for macrosomia: a meta-analysis?, *Arch Gynecol Obstet* **291**: 729-735.
Hedeskov, C.J. (1980) Mechanism of glucose-induced insulin secretion, *Physiol Rev* **60**: 442-509.
Hegele, R.A., Cao, H., Harris, S.B., Hanley, A.J., and Zinman, B. (1999) The hepatic nuclear factor-1alpha G319S variant is associated with early-onset type 2 diabetes in Canadian Oji-Cree, *J Clin Endocrinol Metab* **84**: 1077-1082.
Heimberg, H., De Vos, A., Pipeleers, D., Thorens, B., and Schuit, F. (1995) Differences in glucose transporter gene expression between rat pancreatic alpha- and beta-cells are correlated to differences in glucose transport but not in glucose utilization, *J Biol Chem* **270**: 8971-8975.
Hendel, A., Bak, R.O., Clark, J.T., Kennedy, A.B., Ryan, D.E., Roy, S., et al. (2015) Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells, *Nat Biotechnol* **33**: 985-989.
Henquin, J.C. (2018) Influence of organ donor attributes and preparation characteristics on the dynamics of insulin secretion in isolated human islets, *Physiol Rep* **6**.
Henquin, J.C. (1980) The potassium permeability of pancreatic islet cells: mechanisms of control and influence on insulin release, *Horm Metab Res Suppl* **Suppl 10**: 66-73.
Henquin, J.C. (2000) Triggering and amplifying pathways of regulation of insulin secretion by glucose, *Diabetes* **49**: 1751-1760.
Henquin, J.C., Dufrane, D., and Nenquin, M. (2006) Nutrient control of insulin secretion in isolated normal human islets, *Diabetes* **55**: 3470-3477.
Henquin, J.C., Gembal, M., Detimary, P., Gao, Z.Y., Warnotte, C., and Gilon, P. (1994) Multisite control of insulin release by glucose, *Diabete Metab* **20**: 132-137.
Henquin, J.C., Nenquin, M., Stiernet, P., and Ahren, B. (2006) In vivo and in vitro glucose-induced biphasic insulin secretion in the mouse: pattern and role of cytoplasmic Ca2+ and amplification signals in beta-cells, *Diabetes* **55**: 441-451.
Herbach, N., Rathkolb, B., Kemter, E., Pichl, L., Klaften, M., de Angelis, M.H., et al. (2007) Dominant-negative effects of a novel mutated Ins2 allele causes early-onset diabetes and severe beta-cell loss in Munich Ins2C95S mutant mice, *Diabetes* **56**: 1268-1276.
Hill, C.P., Dauter, Z., Dodson, E.J., Dodson, G.G., and Dunn, M.F. (1991) X-ray structure of an unusual Ca2+ site and the roles of Zn2+ and Ca2+ in the assembly, stability, and storage of the insulin hexamer, *Biochemistry* **30**: 917-924.
Hirashima, Y., Tsuruzoe, K., Kodama, S., Igata, M., Toyonaga, T., Ueki, K., et al. (2003) Insulin down-regulates insulin receptor substrate-2 expression through the phosphatidylinositol 3-kinase/Akt pathway, *J Endocrinol* **179**: 253-266.
Hockemeyer, D., Wang, H., Kiani, S., Lai, C.S., Gao, Q., Cassady, J.P., et al. (2011) Genetic engineering of human pluripotent cells using TALE nucleases, *Nat Biotechnol* **29**: 731-734.
Hodish, I., Liu, M., Rajpal, G., Larkin, D., Holz, R.W., Adams, A., et al. (2010) Misfolded proinsulin affects bystander proinsulin in neonatal diabetes, *J Biol Chem* **285**: 685-694.

Hodson, D.J., and Rorsman, P. (2020) A Variation on the Theme: SGLT2 Inhibition and Glucagon Secretion in Human Islets, *Diabetes* **69**: 864-866.
Hogrebe, N.J., Augsornworawat, P., Maxwell, K.G., Velazco-Cruz, L., and Millman, J.R. (2020) Targeting the cytoskeleton to direct pancreatic differentiation of human pluripotent stem cells, *Nat Biotechnol* **38**: 460-470.
Hohendorff, J., Szopa, M., Skupien, J., Kapusta, M., Zapala, B., Platek, T., et al. (2017) A single dose of dapagliflozin, an SGLT-2 inhibitor, induces higher glycosuria in GCK- and HNF1A-MODY than in type 2 diabetes mellitus, *Endocrine* **57**: 272-279.
Holmkvist, J., Cervin, C., Lyssenko, V., Winckler, W., Anevski, D., Cilio, C., et al. (2006) Common variants in HNF-1 alpha and risk of type 2 diabetes, *Diabetologia* **49**: 2882-2891.
Horikawa, Y., Enya, M., Mabe, H., Fukushima, K., Takubo, N., Ohashi, M., et al. (2018) NEUROD1-deficient diabetes (MODY6): Identification of the first cases in Japanese and the clinical features, *Pediatr Diabetes* **19**: 236-242.
Horikawa, Y., Iwasaki, N., Hara, M., Furuta, H., Hinokio, Y., Cockburn, B.N., et al. (1997) Mutation in hepatocyte nuclear factor-1 beta gene (TCF2) associated with MODY, *Nat Genet* **17**: 384-385.
Hrvatin, S., O'Donnell, C.W., Deng, F., Millman, J.R., Pagliuca, F.W., DiIorio, P., et al. (2014) Differentiated human stem cells resemble fetal, not adult, beta cells, *Proc Natl Acad Sci U S A* **111**: 3038-3043.
Hu, F.B. (2011) Globalization of diabetes: the role of diet, lifestyle, and genes, *Diabetes Care* **34**: 1249-1257.
Huang, D.W., Sherman, B.T., Tan, Q., Collins, J.R., Alvord, W.G., Roayaei, J., et al. (2007) The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists, *Genome Biol* **8**: R183.
Huh, Y.H., Jeon, S.H., and Yoo, S.H. (2003) Chromogranin B-induced secretory granule biogenesis: comparison with the similar role of chromogranin A, *J Biol Chem* **278**: 40581-40589.
Hutton, J.C., Davidson, H.W., and Peshavaria, M. (1987) Proteolytic processing of chromogranin A in purified insulin granules. Formation of a 20 kDa N-terminal fragment (betagranin) by the concerted action of a Ca2+-dependent endopeptidase and carboxypeptidase H (EC 3.4.17.10), *Biochem J* **244**: 457-464.
Hvidberg, A., Nielsen, M.T., Hilsted, J., Orskov, C., and Holst, J.J. (1994) Effect of glucagon-like peptide-1 (proglucagon 78-107amide) on hepatic glucose production in healthy man, *Metabolism* **43**: 104-108.
Ilonen, J., Hammais, A., Laine, A.P., Lempainen, J., Vaarala, O., Veijola, R., et al. (2013) Patterns of beta-cell autoantibody appearance and genetic associations during the first years of life, *Diabetes* **62**: 3636-3640.
Inoguchi, T., Li, P., Umeda, F., Yu, H.Y., Kakimoto, M., Imamura, M., et al. (2000) High glucose level and free fatty acid stimulate reactive oxygen species production through protein kinase C--dependent activation of NAD(P)H oxidase in cultured vascular cells, *Diabetes* **49**: 1939-1945.
Inoue, H., Tanizawa, Y., Wasson, J., Behn, P., Kalidas, K., Bernal-Mizrachi, E., et al. (1998) A gene encoding a transmembrane protein is mutated in patients with diabetes mellitus and optic atrophy (Wolfram syndrome), *Nat Genet* **20**: 143-148.
Izumi, T., Yokota-Hashimoto, H., Zhao, S., Wang, J., Halban, P.A., and Takeuchi, T. (2003) Dominant negative pathogenesis by mutant proinsulin in the Akita diabetic mouse, *Diabetes* **52**: 409-416.
Jacobsen, L., Calvin, S., and Lobenhofer, E. (2009) Transcriptional effects of transfection: the potential for misinterpretation of gene expression data generated from transiently transfected cells, *Biotechniques* **47**: 617-624.
James, D.E., Lederman, L., and Pilch, P.F. (1987) Purification of insulin-dependent exocytic vesicles containing the glucose transporter, *J Biol Chem* **262**: 11817-11824.
Jennings, R.E., Berry, A.A., Kirkwood-Wilson, R., Roberts, N.A., Hearn, T., Salisbury, R.J., et al. (2013) Development of the human pancreas from foregut to endocrine commitment, *Diabetes* **62**: 3514-3522.
Jennings, R.E., Berry, A.A., Strutt, J.P., Gerrard, D.T., and Hanley, N.A. (2015) Human pancreas development, *Development* **142**: 3126-3137.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity, *Science* **337**: 816-821.
Johansson, K.A., Dursun, U., Jordan, N., Gu, G., Beermann, F., Gradwohl, G., and Grapin-Botton, A. (2007) Temporal control of neurogenin3 activity in pancreas progenitors reveals competence windows for the generation of different endocrine cell types, *Dev Cell* **12**: 457-465.
Johnson, M.B., De Franco, E., Greeley, S.A.W., Letourneau, L.R., Gillespie, K.M., International, D.S.P.C., et al. (2019) Trisomy 21 Is a Cause of Permanent Neonatal Diabetes That Is Autoimmune but Not HLA Associated, *Diabetes* **68**: 1528-1535.
Johnson, M.B., De Franco, E., Lango Allen, H., Al Senani, A., Elbarbary, N., Siklar, Z., et al. (2017) Recessively Inherited LRBA Mutations Cause Autoimmunity Presenting as Neonatal Diabetes, *Diabetes* **66**: 2316-2322.
Jones, A.G., and Hattersley, A.T. (2013) The clinical utility of C-peptide measurement in the care of patients with diabetes, *Diabet Med* **30**: 803-817.
Jonnakuty, C., and Gragnoli, C. (2007) Karyotype of the human insulinoma CM cell line--beta cell model in vitro?, *J Cell Physiol* **213**: 661-662.
Jonsson, A., Isomaa, B., Tuomi, T., Eriksson, J.G., Groop, L., and Lyssenko, V. (2012) Effect of a common variant of the PCSK2 gene on reduced insulin secretion, *Diabetologia* **55**: 3245-3251.
Juszczak, A., Pryse, R., Schuman, A., and Owen, K.R. (2016) When to consider a diagnosis of MODY at the presentation of diabetes: aetiology matters for correct management, *Br J Gen Pract* **66**: e457-459.
Kahn, S.E., Haffner, S.M., Heise, M.A., Herman, W.H., Holman, R.R., Jones, N.P., et al. (2006) Glycemic durability of rosiglitazone, metformin, or glyburide monotherapy, *N Engl J Med* **355**: 2427-2443.
Kahn, S.E., Prigeon, R.L., McCulloch, D.K., Boyko, E.J., Bergman, R.N., Schwartz, M.W., et al. (1993) Quantification of the relationship between insulin sensitivity and beta-cell function in human subjects. Evidence for a hyperbolic function, *Diabetes* **42**: 1663-1672.
Kajiwara, M., Aoi, T., Okita, K., Takahashi, R., Inoue, H., Takayama, N., et al. (2012) Donor-dependent variations in hepatic differentiation from human-induced pluripotent stem cells, *Proc Natl Acad Sci U S A* **109**: 12538-12543.
Kaprio, J., Tuomilehto, J., Koskenvuo, M., Romanov, K., Reunanen, A., Eriksson, J., et al. (1992) Concordance for type 1 (insulin-dependent) and type 2 (non-insulin-dependent) diabetes mellitus in a population-based cohort of twins in Finland, *Diabetologia* **35**: 1060-1067.
Kaufmann, J.E., Irminger, J.C., and Halban, P.A. (1995) Sequence requirements for proinsulin processing at the B-chain/C-peptide junction, *Biochem J* **310 (Pt 3)**: 869-874.
Kavvoura, F.K., and Owen, K.R. (2012) Maturity onset diabetes of the young: clinical characteristics, diagnosis and management, *Pediatr Endocrinol Rev* **10**: 234-242.
Kayo, T., and Koizumi, A. (1998) Mapping of murine diabetogenic gene mody on chromosome 7 at D7Mit258 and its involvement in pancreatic islet and beta cell development during the perinatal period, *J Clin Invest* **101**: 2112-2118.
Kayton, N.S., Poffenberger, G., Henske, J., Dai, C., Thompson, C., Aramandla, R., et al. (2015) Human islet preparations distributed for research exhibit a variety of insulin-secretory profiles, *Am J Physiol Endocrinol Metab* **308**: E592-602.
Kemmler, W., and Steiner, D.F. (1970) Conversion of proinsulin to insulin in a subcellular fraction from rat islets, *Biochem Biophys Res Commun* **41**: 1223-1230.
Kersten, S. (2001) Mechanisms of nutritional and hormonal regulation of lipogenesis, *EMBO Rep* **2**: 282-286.
Kim, D., Bae, S., Park, J., Kim, E., Kim, S., Yu, H.R., et al. (2015) Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells, *Nat Methods* **12**: 237-243, 231 p following 243.
Kim, T., Tao-Cheng, J.H., Eiden, L.E., and Loh, Y.P. (2001) Chromogranin A, an "on/off" switch controlling dense-core secretory granule biogenesis, *Cell* **106**: 499-509.
Klupa, T., Kowalska, I., Wyka, K., Skupien, J., Patch, A.M., Flanagan, S.E., et al. (2009) Mutations in the ABCC8 (SUR1 subunit of the K(ATP) channel) gene are associated with a variable clinical phenotype, *Clin Endocrinol (Oxf)* **71**: 358-362.

Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage, *Nature* **533**: 420-424.
Kristinsson, S.Y., Thorolfsdottir, E.T., Talseth, B., Steingrimsson, E., Thorsson, A.V., Helgason, T., et al. (2001) MODY in Iceland is associated with mutations in HNF-1alpha and a novel mutation in NeuroD1, *Diabetologia* **44**: 2098-2103.
Kroon, E., Martinson, L.A., Kadoya, K., Bang, A.G., Kelly, O.G., Eliazer, S., et al. (2008) Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells in vivo, *Nat Biotechnol* **26**: 443-452.
Kyttala, A., Moraghebi, R., Valensisi, C., Kettunen, J., Andrus, C., Pasumarthy, K.K., et al. (2016) Genetic Variability Overrides the Impact of Parental Cell Type and Determines iPSC Differentiation Potential, *Stem Cell Reports* **6**: 200-212.
Lawlor, N., Marquez, E.J., Orchard, P., Narisu, N., Shamim, M.S., Thibodeau, A., et al. (2019) Multiomic Profiling Identifies cis-Regulatory Networks Underlying Human Pancreatic beta Cell Identity and Function, *Cell Rep* **26**: 788-801 e786.
Lean, M.E.J., Leslie, W.S., Barnes, A.C., Brosnahan, N., Thom, G., McCombie, L., et al. (2019) Durability of a primary care-led weight-management intervention for remission of type 2 diabetes: 2-year results of the DiRECT open-label, cluster-randomised trial, *Lancet Diabetes Endocrinol* **7**: 344-355.
Lee, Y.H., Sauer, B., and Gonzalez, F.J. (1998) Laron dwarfism and non-insulin-dependent diabetes mellitus in the Hnf-1alpha knockout mouse, *Mol Cell Biol* **18**: 3059-3068.
Lefebvre, A.M., Laville, M., Vega, N., Riou, J.P., van Gaal, L., Auwerx, J., and Vidal, H. (1998) Depot-specific differences in adipose tissue gene expression in lean and obese subjects, *Diabetes* **47**: 98-103.
Leroux, L., Desbois, P., Lamotte, L., Duvillie, B., Cordonnier, N., Jackerott, M., et al. (2001) Compensatory responses in mice carrying a null mutation for Ins1 or Ins2, *Diabetes* **50 Suppl 1**: S150-153.
Li, Y., Guo, Y., Tang, J., Jiang, J., and Chen, Z. (2014) New insights into the roles of CHOP-induced apoptosis in ER stress, *Acta Biochim Biophys Sin (Shanghai)* **46**: 629-640.
Liao, Y., Smyth, G.K., and Shi, W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features, *Bioinformatics* **30**: 923-930.
Lipkind, G., and Steiner, D.F. (1999) Predicted structural alterations in proinsulin during its interactions with prohormone convertases, *Biochemistry* **38**: 890-896.
Liu, L., Nagashima, K., Yasuda, T., Liu, Y., Hu, H.R., He, G., et al. (2013) Mutations in KCNJ11 are associated with the development of autosomal dominant, early-onset type 2 diabetes, *Diabetologia* **56**: 2609-2618.
Liu, M., Haataja, L., Wright, J., Wickramasinghe, N.P., Hua, Q.X., Phillips, N.F., et al. (2010) Mutant INS-gene induced diabetes of youth: proinsulin cysteine residues impose dominant-negative inhibition on wild-type proinsulin transport, *PLoS One* **5**: e13333.
Liu, M., Hodish, I., Rhodes, C.J., and Arvan, P. (2007) Proinsulin maturation, misfolding, and proteotoxicity, *Proc Natl Acad Sci U S A* **104**: 15841-15846.
Liu, M., Li, Y., Cavener, D., and Arvan, P. (2005) Proinsulin disulfide maturation and misfolding in the endoplasmic reticulum, *J Biol Chem* **280**: 13209-13212.
Liu, M., Ramos-Castaneda, J., and Arvan, P. (2003) Role of the connecting peptide in insulin biosynthesis, *J Biol Chem* **278**: 14798-14805.
Liu, M., Sun, J., Cui, J., Chen, W., Guo, H., Barbetti, F., and Arvan, P. (2015) INS-gene mutations: from genetics and beta cell biology to clinical disease, *Mol Aspects Med* **42**: 3-18.
Liu, X., Qin, J., Chang, M., Wang, S., Li, Y., Pei, X., and Wang, Y. (2020) Enhanced differentiation of human pluripotent stem cells into pancreatic endocrine cells in 3D culture by inhibition of focal adhesion kinase, *Stem Cell Res Ther* **11**: 488.
Love, M.I., Huber, W., and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, *Genome Biol* **15**: 550.
Luco, R.F., Maestro, M.A., del Pozo, N., Philbrick, W.M., de la Ossa, P.P., and Ferrer, J. (2006) A conditional model reveals that induction of hepatocyte nuclear factor-1alpha in Hnf1alpha-null mutant beta-cells can activate silenced genes postnatally, whereas overexpression is deleterious, *Diabetes* **55**: 2202-2211.

Luni, C., Marth, J.D., and Doyle, F.J., 3rd (2012) Computational modeling of glucose transport in pancreatic beta-cells identifies metabolic thresholds and therapeutic targets in diabetes, *PLoS One* **7**: e53130.
Lyon, J., Manning Fox, J.E., Spigelman, A.F., Kim, R., Smith, N., O'Gorman, D., et al. (2016) Research-Focused Isolation of Human Islets From Donors With and Without Diabetes at the Alberta Diabetes Institute IsletCore, *Endocrinology* **157**: 560-569.
Ma, S., Viola, R., Sui, L., Cherubini, V., Barbetti, F., and Egli, D. (2018) beta Cell Replacement after Gene Editing of a Neonatal Diabetes-Causing Mutation at the Insulin Locus, *Stem Cell Reports* **11**: 1407-1415.
MacDonald, M.J., Longacre, M.J., Stoker, S.W., Kendrick, M., Thonpho, A., Brown, L.J., et al. (2011) Differences between human and rodent pancreatic islets: low pyruvate carboxylase, atp citrate lyase, and pyruvate carboxylation and high glucose-stimulated acetoacetate in human pancreatic islets, *J Biol Chem* **286**: 18383-18396.
Madiraju, A.K., Erion, D.M., Rahimi, Y., Zhang, X.M., Braddock, D.T., Albright, R.A., et al. (2014) Metformin suppresses gluconeogenesis by inhibiting mitochondrial glycerophosphate dehydrogenase, *Nature* **510**: 542-546.
Mahajan, A., Taliun, D., Thurner, M., Robertson, N.R., Torres, J.M., Rayner, N.W., et al. (2018) Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps, *Nat Genet* **50**: 1505-1513.
Malaisse, W.J., Best, L., Kawazu, S., Malaisse-Lagae, F., and Sener, A. (1983) The stimulus-secretion coupling of glucose-induced insulin release: fuel metabolism in islets deprived of exogenous nutrient, *Arch Biochem Biophys* **224**: 102-110.
Malecki, M.T., Jhala, U.S., Antonellis, A., Fields, L., Doria, A., Orban, T., et al. (1999) Mutations in NEUROD1 are associated with the development of type 2 diabetes mellitus, *Nat Genet* **23**: 323-328.
Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., et al. (2013) RNA-guided human genome engineering via Cas9, *Science* **339**: 823-826.
Manchester, K.L. (1972) Effect of insulin on protein synthesis, *Diabetes* **21**: 447-452.
Marin-Penalver, J.J., Martin-Timon, I., Sevillano-Collantes, C., and Del Canizo-Gomez, F.J. (2016) Update on the treatment of type 2 diabetes mellitus, *World J Diabetes* **7**: 354-395.
Martin, G.R. (1981) Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells, *Proc Natl Acad Sci U S A* **78**: 7634-7638.
Massa, L., Baltrusch, S., Okar, D.A., Lange, A.J., Lenzen, S., and Tiedge, M. (2004) Interaction of 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase (PFK-2/FBPase-2) with glucokinase activates glucose phosphorylation and glucose metabolism in insulin-producing cells, *Diabetes* **53**: 1020-1029.
Mauvais-Jarvis, F., Smith, S.B., Le May, C., Leal, S.M., Gautier, J.F., Molokhia, M., et al. (2004) PAX4 gene variations predispose to ketosis-prone diabetes, *Hum Mol Genet* **13**: 3151-3159.
Maxwell, K.G., Augsornworawat, P., Velazco-Cruz, L., Kim, M.H., Asada, R., Hogrebe, N.J., et al. (2020) Gene-edited human stem cell-derived beta cells from a patient with monogenic diabetes reverse preexisting diabetes in mice, *Sci Transl Med* **12**.
Mayerson, A.B., Hundal, R.S., Dufour, S., Lebon, V., Befroy, D., Cline, G.W., et al. (2002) The effects of rosiglitazone on insulin sensitivity, lipolysis, and hepatic and skeletal muscle triglyceride content in patients with type 2 diabetes, *Diabetes* **51**: 797-802.
McCulloch, L.J., van de Bunt, M., Braun, M., Frayn, K.N., Clark, A., and Gloyn, A.L. (2011) GLUT2 (SLC2A2) is not the principal glucose transporter in human pancreatic beta cells: implications for understanding genetic association signals at this locus, *Mol Genet Metab* **104**: 648-653.
McDonald, T.J., Shields, B.M., Lawry, J., Owen, K.R., Gloyn, A.L., Ellard, S., and Hattersley, A.T. (2011) High-sensitivity CRP discriminates HNF1A-MODY from other subtypes of diabetes, *Diabetes Care* **34**: 1860-1862.
Mendel, D.B., and Crabtree, G.R. (1991) HNF-1, a member of a novel class of dimerizing homeodomain proteins, *J Biol Chem* **266**: 677-680.

Merkle, F.T., Ghosh, S., Kamitaki, N., Mitchell, J., Avior, Y., Mello, C., et al. (2017) Human pluripotent stem cells recurrently acquire and expand dominant negative P53 mutations, *Nature* **545**: 229-233.
Meur, G., Simon, A., Harun, N., Virally, M., Dechaume, A., Bonnefond, A., et al. (2010) Insulin gene mutations resulting in early-onset diabetes: marked differences in clinical presentation, metabolic status, and pathogenic effect through endoplasmic reticulum retention, *Diabetes* **59**: 653-661.
Micallef, S.J., Li, X., Schiesser, J.V., Hirst, C.E., Yu, Q.C., Lim, S.M., et al. (2012) INS(GFP/w) human embryonic stem cells facilitate isolation of in vitro derived insulin-producing cells, *Diabetologia* **55**: 694-706.
Milner, R.D., and Hill, D.J. (1984) Fetal growth control: the role of insulin and related peptides, *Clin Endocrinol (Oxf)* **21**: 415-433.
Misra, S., Hassanali, N., Bennett, A.J., Juszczak, A., Caswell, R., Colclough, K., et al. (2020) Homozygous Hypomorphic HNF1A Alleles Are a Novel Cause of Young-Onset Diabetes and Result in Sulfonylurea-Sensitive Diabetes, *Diabetes Care* **43**: 909-912.
Misra, S., and Owen, K.R. (2018) Genetics of Monogenic Diabetes: Present Clinical Challenges, *Curr Diab Rep* **18**: 141.
Miura, A., Yamagata, K., Kakei, M., Hatakeyama, H., Takahashi, N., Fukui, K., et al. (2006) Hepatocyte nuclear factor-4alpha is essential for glucose-stimulated insulin secretion by pancreatic beta-cells, *J Biol Chem* **281**: 5246-5257.
Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Soria, E. (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements, *J Mol Evol* **60**: 174-182.
Molven, A., Ringdal, M., Nordbo, A.M., Raeder, H., Stoy, J., Lipkind, G.M., et al. (2008) Mutations in the insulin gene can cause MODY and autoantibody-negative type 1 diabetes, *Diabetes* **57**: 1131-1135.
Motzo, C., Contu, D., Cordell, H.J., Lampis, R., Congia, M., Marrosu, M.G., et al. (2004) Heterogeneity in the magnitude of the insulin gene effect on HLA risk in type 1 diabetes, *Diabetes* **53**: 3286-3291.
Munro, S., and Pelham, H.R. (1987) A C-terminal signal prevents secretion of luminal ER proteins, *Cell* **48**: 899-907.
Nair, G.G., Liu, J.S., Russ, H.A., Tran, S., Saxton, M.S., Chen, R., et al. (2019) Recapitulating endocrine cell clustering in culture promotes maturation of human stem-cell-derived beta cells, *Nat Cell Biol* **21**: 263-274.
Najmi, L.A., Aukrust, I., Flannick, J., Molnes, J., Burtt, N., Molven, A., et al. (2017) Functional Investigations of HNF1A Identify Rare Variants as Risk Factors for Type 2 Diabetes in the General Population, *Diabetes* **66**: 335-346.
Nanjo, K., Sanke, T., Miyano, M., Okai, K., Sowa, R., Kondo, M., et al. (1986) Diabetes due to secretion of a structurally abnormal insulin (insulin Wakayama). Clinical and functional characteristics of [LeuA3] insulin, *J Clin Invest* **77**: 514-519.
Naya, F.J., Stellrecht, C.M., and Tsai, M.J. (1995) Tissue-specific regulation of the insulin gene by a novel basic helix-loop-helix transcription factor, *Genes Dev* **9**: 1009-1019.
Naylor, R.N., Greeley, S.A., Bell, G.I., and Philipson, L.H. (2011) Genetics and pathophysiology of neonatal diabetes mellitus, *J Diabetes Investig* **2**: 158-169.
Nerup, J., Andersen, O.O., Bendixen, G., Egeberg, J., and Poulsen, J.E. (1971) Anti-pancreatic cellular hypersensitivity in diabetes mellitus, *Diabetes* **20**: 424-427.
Ng, N.H.J., Jasmen, J.B., Lim, C.S., Lau, H.H., Krishnan, V.G., Kadiwala, J., et al. (2019) HNF4A Haploinsufficiency in MODY1 Abrogates Liver and Pancreas Differentiation from Patient-Derived Induced Pluripotent Stem Cells, *iScience* **16**: 192-205.
Nica, A.C., Ongen, H., Irminger, J.C., Bosco, D., Berney, T., Antonarakis, S.E., et al. (2013) Cell-type, allelic, and genetic signatures in the human pancreatic beta cell transcriptome, *Genome Res* **23**: 1554-1562.
Nishi, M., and Nanjo, K. (2011) Insulin gene mutations and diabetes, *J Diabetes Investig* **2**: 92-100.

Nostro, M.C., Sarangi, F., Ogawa, S., Holtzinger, A., Corneo, B., Li, X., et al. (2011) Stage-specific signaling through TGFbeta family members and WNT regulates patterning and pancreatic specification of human pluripotent stem cells, *Development* **138**: 861-871.
Nostro, M.C., Sarangi, F., Yang, C., Holland, A., Elefanty, A.G., Stanley, E.G., et al. (2015) Efficient generation of NKX6-1+ pancreatic progenitors from multiple human pluripotent stem cell lines, *Stem Cell Reports* **4**: 591-604.
Oakes, N.D., Cooney, G.J., Camilleri, S., Chisholm, D.J., and Kraegen, E.W. (1997) Mechanisms of liver and muscle insulin resistance induced by chronic high-fat feeding, *Diabetes* **46**: 1768-1774.
Odom, D.T., Zizlsperger, N., Gordon, D.B., Bell, G.W., Rinaldi, N.J., Murray, H.L., et al. (2004) Control of pancreas and liver gene expression by HNF transcription factors, *Science* **303**: 1378-1381.
Ohneda, K., Mirmira, R.G., Wang, J., Johnson, J.D., and German, M.S. (2000) The homeodomain of PDX-1 mediates multiple protein-protein interactions in the formation of a transcriptional activation complex on the insulin promoter, *Mol Cell Biol* **20**: 900-911.
Oram, R.A., Patel, K., Hill, A., Shields, B., McDonald, T.J., Jones, A., et al. (2016) A Type 1 Diabetes Genetic Risk Score Can Aid Discrimination Between Type 1 and Type 2 Diabetes in Young Adults, *Diabetes Care* **39**: 337-344.
Orci, L., Baetens, D., Ravazzola, M., Stefan, Y., and Malaisse-Lagae, F. (1976) Pancreatic polypeptide and glucagon : non-random distribution in pancreatic islets, *Life Sci* **19**: 1811-1815.
Orci, L., Ravazzola, M., Storch, M.J., Anderson, R.G., Vassalli, J.D., and Perrelet, A. (1987) Proteolytic maturation of insulin is a post-Golgi event which occurs in acidifying clathrin-coated secretory vesicles, *Cell* **49**: 865-868.
Ostoft, S.H., Bagger, J.I., Hansen, T., Pedersen, O., Holst, J.J., Knop, F.K., and Vilsboll, T. (2014) Incretin effect and glucagon responses to oral and intravenous glucose in patients with maturity-onset diabetes of the young--type 2 and type 3, *Diabetes* **63**: 2838-2844.
Owen, K.R., Thanabalasingham, G., James, T.J., Karpe, F., Farmer, A.J., McCarthy, M.I., and Gloyn, A.L. (2010) Assessment of high-sensitivity C-reactive protein levels as diagnostic discriminator of maturity-onset diabetes of the young due to HNF1A mutations, *Diabetes Care* **33**: 1919-1924.
Oyadomari, S., Koizumi, A., Takeda, K., Gotoh, T., Akira, S., Araki, E., and Mori, M. (2002) Targeted disruption of the Chop gene delays endoplasmic reticulum stress-mediated diabetes, *J Clin Invest* **109**: 525-532.
Pagliuca, F.W., Millman, J.R., Gurtler, M., Segel, M., Van Dervort, A., Ryu, J.H., et al. (2014) Generation of functional human pancreatic beta cells in vitro, *Cell* **159**: 428-439.
Parker, S.C., Stitzel, M.L., Taylor, D.L., Orozco, J.M., Erdos, M.R., Akiyama, J.A., et al. (2013) Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants, *Proc Natl Acad Sci U S A* **110**: 17921-17926.
Pasquali, L., Gaulton, K.J., Rodriguez-Segui, S.A., Mularoni, L., Miguel-Escalada, I., Akerman, I., et al. (2014) Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants, *Nat Genet* **46**: 136-143.
Patel, K.A., Oram, R.A., Flanagan, S.E., De Franco, E., Colclough, K., Shepherd, M., et al. (2016) Type 1 Diabetes Genetic Risk Score: A Novel Tool to Discriminate Monogenic and Type 1 Diabetes, *Diabetes* **65**: 2094-2099.
Pearson, E.R., Flechtner, I., Njolstad, P.R., Malecki, M.T., Flanagan, S.E., Larkin, B., et al. (2006) Switching from insulin to oral sulfonylureas in patients with diabetes due to Kir6.2 mutations, *N Engl J Med* **355**: 467-477.
Pearson, E.R., Starkey, B.J., Powell, R.J., Gribble, F.M., Clark, P.M., and Hattersley, A.T. (2003) Genetic cause of hyperglycaemia and response to treatment in diabetes, *Lancet* **362**: 1275-1281.
Pearson, E.R., Velho, G., Clark, P., Stride, A., Shepherd, M., Frayling, T.M., et al. (2001) beta-cell genes and diabetes: quantitative and qualitative differences in the pathophysiology of hepatic nuclear factor-1alpha and glucokinase mutations, *Diabetes* **50 Suppl 1**: S101-107.
Pedersen, K.B., Chhabra, K.H., Nguyen, V.K., Xia, H., and Lazartigues, E. (2013) The transcription factor HNF1alpha induces expression of angiotensin-converting enzyme 2 (ACE2)

in pancreatic islets from evolutionarily conserved promoter motifs, *Biochim Biophys Acta* **1829**: 1225-1235.
Perez-Alcantara, M., Honore, C., Wesolowska-Andersen, A., Gloyn, A.L., McCarthy, M.I., Hansson, M., et al. (2018) Patterns of differential gene expression in a cellular model of human islet development, and relationship to type 2 diabetes predisposition, *Diabetologia* **61**: 1614-1622.
Permutt, M.A., and Kipnis, D.M. (1972) Insulin biosynthesis. I. On the mechanism of glucose stimulation, *J Biol Chem* **247**: 1194-1199.
Petersen, M.B.K., Azad, A., Ingvorsen, C., Hess, K., Hansson, M., Grapin-Botton, A., and Honore, C. (2017) Single-Cell Gene Expression Analysis of a Human ESC Model of Pancreatic Endocrine Development Reveals Different Paths to beta-Cell Differentiation, *Stem Cell Reports* **9**: 1246-1261.
Pi-Sunyer, X., Astrup, A., Fujioka, K., Greenway, F., Halpern, A., Krempf, M., et al. (2015) A Randomized, Controlled Trial of 3.0 mg of Liraglutide in Weight Management, *N Engl J Med* **373**: 11-22.
Pihoker, C., Gilliam, L.K., Ellard, S., Dabelea, D., Davis, C., Dolan, L.M., et al. (2013) Prevalence, characteristics and clinical diagnosis of maturity onset diabetes of the young due to mutations in HNF1A, HNF4A, and glucokinase: results from the SEARCH for Diabetes in Youth, *J Clin Endocrinol Metab* **98**: 4055-4062.
Piper, K., Brickwood, S., Turnpenny, L.W., Cameron, I.T., Ball, S.G., Wilson, D.I., and Hanley, N.A. (2004) Beta cell differentiation during early human pancreas development, *J Endocrinol* **181**: 11-23.
Plengvidhya, N., Kooptiwut, S., Songtawee, N., Doi, A., Furuta, H., Nishi, M., et al. (2007) PAX4 mutations in Thais with maturity onset diabetes of the young, *J Clin Endocrinol Metab* **92**: 2821-2826.
Pociot, F., Akolkar, B., Concannon, P., Erlich, H.A., Julier, C., Morahan, G., et al. (2010) Genetics of type 1 diabetes: what's next?, *Diabetes* **59**: 1561-1571.
Pontoglio, M., Barra, J., Hadchouel, M., Doyen, A., Kress, C., Bach, J.P., et al. (1996) Hepatocyte nuclear factor 1 inactivation results in hepatic dysfunction, phenylketonuria, and renal Fanconi syndrome, *Cell* **84**: 575-585.
Pontoglio, M., Prie, D., Cheret, C., Doyen, A., Leroy, C., Froguel, P., et al. (2000) HNF1alpha controls renal glucose reabsorption in mouse and man, *EMBO Rep* **1**: 359-365.
Pontoglio, M., Sreenan, S., Roe, M., Pugh, W., Ostrega, D., Doyen, A., et al. (1998) Defective insulin secretion in hepatocyte nuclear factor 1alpha-deficient mice, *J Clin Invest* **101**: 2215-2222.
Porter, A.G., and Janicke, R.U. (1999) Emerging roles of caspase-3 in apoptosis, *Cell Death Differ* **6**: 99-104.
Pouli, A.E., Kennedy, H.J., Schofield, J.G., and Rutter, G.A. (1998) Insulin targeting to the regulated secretory pathway after fusion with green fluorescent protein and firefly luciferase, *Biochem J* **331 (Pt 2)**: 669-675.
Pound, L.D., Sarkar, S.A., Benninger, R.K., Wang, Y., Suwanichkul, A., Shadoan, M.K., et al. (2009) Deletion of the mouse Slc30a8 gene encoding zinc transporter-8 results in impaired insulin secretion, *Biochem J* **421**: 371-376.
Pound, L.D., Sarkar, S.A., Ustione, A., Dadi, P.K., Shadoan, M.K., Lee, C.E., et al. (2012) The physiological effects of deleting the mouse SLC30A8 gene encoding zinc transporter-8 are influenced by gender and genetic background, *PLoS One* **7**: e40972.
Proks, P., Arnold, A.L., Bruining, J., Girard, C., Flanagan, S.E., Larkin, B., et al. (2006) A heterozygous activating mutation in the sulphonylurea receptor SUR1 (ABCC8) causes neonatal diabetes, *Hum Mol Genet* **15**: 1793-1800.
Pugliese, A., Zeller, M., Fernandez, A., Jr., Zalcberg, L.J., Bartlett, R.J., Ricordi, C., et al. (1997) The insulin gene is transcribed in the human thymus and transcription levels correlated with allelic variation at the INS VNTR-IDDM2 susceptibility locus for type 1 diabetes, *Nat Genet* **15**: 293-297.
Qu, X.B., Pan, J., Zhang, C., and Huang, S.Y. (2008) Sox17 facilitates the differentiation of mouse embryonic stem cells into primitive and definitive endoderm in vitro, *Dev Growth Differ* **50**: 585-593.

Raeder, H., Johansson, S., Holm, P.I., Haldorsen, I.S., Mas, E., Sbarra, V., et al. (2006) Mutations in the CEL VNTR cause a syndrome of diabetes and pancreatic exocrine dysfunction, *Nat Genet* **38**: 54-62.
Raeder, H., McAllister, F.E., Tjora, E., Bhatt, S., Haldorsen, I., Hu, J., et al. (2014) Carboxyl-ester lipase maturity-onset diabetes of the young is associated with development of pancreatic cysts and upregulated MAPK signaling in secretin-stimulated duodenal fluid, *Diabetes* **63**: 259-269.
Rafiq, M., Flanagan, S.E., Patch, A.M., Shields, B.M., Ellard, S., Hattersley, A.T., and Neonatal Diabetes International Collaborative, G. (2008) Effective treatment with oral sulfonylureas in patients with diabetes due to sulfonylurea receptor 1 (SUR1) mutations, *Diabetes Care* **31**: 204-209.
Raimondo, A., Chakera, A.J., Thomsen, S.K., Colclough, K., Barrett, A., De Franco, E., et al. (2014) Phenotypic severity of homozygous GCK mutations causing neonatal or childhood-onset diabetes is primarily mediated through effects on protein stability, *Hum Mol Genet* **23**: 6432-6440.
Rajagopal, J., Anderson, W.J., Kume, S., Martinez, O.I., and Melton, D.A. (2003) Insulin staining of ES cell progeny from insulin uptake, *Science* **299**: 363.
Rajan, S., Eames, S.C., Park, S.Y., Labno, C., Bell, G.I., Prince, V.E., and Philipson, L.H. (2010) In vitro processing and secretion of mutant insulin proteins that cause permanent neonatal diabetes, *Am J Physiol Endocrinol Metab* **298**: E403-410.
Ramzy, A., Asadi, A., and Kieffer, T.J. (2020) Revisiting Proinsulin Processing: Evidence That Human beta-Cells Process Proinsulin With Prohormone Convertase (PC) 1/3 But Not PC2, *Diabetes*.
Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., and Vilo, J. (2019) g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update), *Nucleic Acids Res* **47**: W191-W198.
Ravassard, P., Hazhouz, Y., Pechberty, S., Bricout-Neveu, E., Armanet, M., Czernichow, P., and Scharfmann, R. (2011) A genetically engineered human pancreatic beta cell line exhibiting glucose-inducible insulin secretion, *J Clin Invest* **121**: 3589-3597.
Reaven, G.M., Hollenbeck, C., Jeng, C.Y., Wu, M.S., and Chen, Y.D. (1988) Measurement of plasma glucose, free fatty acid, lactate, and insulin for 24 h in patients with NIDDM, *Diabetes* **37**: 1020-1024.
Reber, S., Mechtersheimer, J., Nasif, S., Benitez, J.A., Colombo, M., Domanski, M., et al. (2018) CRISPR-Trap: a clean approach for the generation of gene knockouts and gene replacements in human cells, *Mol Biol Cell* **29**: 75-83.
Rees, S.D., Hydrie, M.Z., O'Hare, J.P., Kumar, S., Shera, A.S., Basit, A., et al. (2011) Effects of 16 genetic variants on fasting glucose and type 2 diabetes in South Asians: ADCY5 and GLIS3 variants may predispose to type 2 diabetes, *PLoS One* **6**: e24710.
Rezania, A., Bruin, J.E., Arora, P., Rubin, A., Batushansky, I., Asadi, A., et al. (2014) Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells, *Nat Biotechnol* **32**: 1121-1133.
Rezania, A., Bruin, J.E., Xu, J., Narayan, K., Fox, J.K., O'Neil, J.J., and Kieffer, T.J. (2013) Enrichment of human embryonic stem cell-derived NKX6.1-expressing pancreatic progenitor cells accelerates the maturation of insulin-secreting cells in vivo, *Stem Cells* **31**: 2432-2442.
Rich, S.S., Akolkar, B., Concannon, P., Erlich, H., Hilner, J.E., Julier, C., et al. (2009) Overview of the Type I Diabetes Genetics Consortium, *Genes Immun* **10 Suppl 1**: S1-4.
Rich, S.S., Concannon, P., Erlich, H., Julier, C., Morahan, G., Nerup, J., et al. (2006) The Type 1 Diabetes Genetics Consortium, *Ann N Y Acad Sci* **1079**: 1-8.
Robertson, C.C., Inshaw, R.J., Onengut-Gumuscu, S., Chen, W.M., Flores Santa Cruz, D., Yang, H., et al. (2020) Fine-mapping, trans-ancestral and genomic analyses identify causal variants, cells, genes and drug targets for type 1 diabetes, *bioRxiv* **2020.06.19.158071**.
Rorsman, P., and Ashcroft, F.M. (2018) Pancreatic beta-Cell Electrical Activity and Insulin Secretion: Of Mice and Men, *Physiol Rev* **98**: 117-214.
Rorsman, P., Braun, M., and Zhang, Q. (2012) Regulation of calcium in pancreatic alpha- and beta-cells in health and disease, *Cell Calcium* **51**: 300-308.

Rossi, A., Kontarakis, Z., Gerri, C., Nolte, H., Holper, S., Kruger, M., and Stainier, D.Y. (2015) Genetic compensation induced by deleterious mutations but not gene knockdowns, *Nature* **524**: 230-233.
Rouhani, F., Kumasaka, N., de Brito, M.C., Bradley, A., Vallier, L., and Gaffney, D. (2014) Genetic background drives transcriptional variation in human induced pluripotent stem cells, *PLoS Genet* **10**: e1004432.
Rubio-Cabezas, O., and Ellard, S. (2013) Diabetes mellitus in neonates and infants: genetic heterogeneity, clinical approach to diagnosis, and therapeutic options, *Horm Res Paediatr* **80**: 137-146.
Rubio-Cabezas, O., Flanagan, S.E., Damhuis, A., Hattersley, A.T., and Ellard, S. (2012) KATP channel mutations in infants with permanent diabetes diagnosed after 6 months of life, *Pediatr Diabetes* **13**: 322-325.
Rubio-Cabezas, O., Minton, J.A., Kantor, I., Williams, D., Ellard, S., and Hattersley, A.T. (2010) Homozygous mutations in NEUROD1 are responsible for a novel syndrome of permanent neonatal diabetes and neurological abnormalities, *Diabetes* **59**: 2326-2331.
Rubio-Cabezas, O., Patch, A.M., Minton, J.A., Flanagan, S.E., Edghill, E.L., Hussain, K., et al. (2009) Wolcott-Rallison syndrome is the most common genetic cause of permanent neonatal diabetes in consanguineous families, *J Clin Endocrinol Metab* **94**: 4162-4170.
Russ, H.A., Parent, A.V., Ringler, J.J., Hennings, T.G., Nair, G.G., Shveygert, M., et al. (2015) Controlled induction of human pancreatic progenitors produces functional beta-like cells in vitro, *EMBO J* **34**: 1759-1772.
Saarimaki-Vire, J., Balboa, D., Russell, M.A., Saarikettu, J., Kinnunen, M., Keskitalo, S., et al. (2017) An Activating STAT3 Mutation Causes Neonatal Diabetes through Premature Induction of Pancreatic Differentiation, *Cell Rep* **19**: 281-294.
Sadur, C.N., and Eckel, R.H. (1982) Insulin stimulation of adipose tissue lipoprotein lipase. Use of the euglycemic clamp technique, *J Clin Invest* **69**: 1119-1125.
Sambandam, N., Abrahani, M.A., St Pierre, E., Al-Atar, O., Cam, M.C., and Rodrigues, B. (1999) Localization of lipoprotein lipase in the diabetic heart: regulation by acute changes in insulin, *Arterioscler Thromb Vasc Biol* **19**: 1526-1534.
Santomauro, A.T., Boden, G., Silva, M.E., Rocha, D.M., Santos, R.F., Ursich, M.J., et al. (1999) Overnight lowering of free fatty acids with Acipimox improves insulin resistance and glucose tolerance in obese diabetic and nondiabetic subjects, *Diabetes* **48**: 1836-1841.
Saponaro, C., Muhlemann, M., Acosta-Montalvo, A., Piron, A., Gmyr, V., Delalleau, N., et al. (2020) Interindividual Heterogeneity of SGLT2 Expression and Function in Human Pancreatic Islets, *Diabetes* **69**: 902-914.
Sarkar, S.A., Kobberup, S., Wong, R., Lopez, A.D., Quayum, N., Still, T., et al. (2008) Global gene expression profiling and histochemical analysis of the developing human fetal pancreas, *Diabetologia* **51**: 285-297.
Schulla, V., Renstrom, E., Feil, R., Feil, S., Franklin, I., Gjinovci, A., et al. (2003) Impaired insulin secretion and glucose tolerance in beta cell-selective Ca(v)1.2 Ca2+ channel null mice, *EMBO J* **22**: 3844-3854.
Schwitzgebel, V.M., Mamin, A., Brun, T., Ritz-Laser, B., Zaiko, M., Maret, A., et al. (2003) Agenesis of human pancreas due to decreased half-life of insulin promoter factor 1, *J Clin Endocrinol Metab* **88**: 4398-4406.
Segerstolpe, A., Palasantza, A., Eliasson, P., Andersson, E.M., Andreasson, A.C., Sun, X., et al. (2016) Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes, *Cell Metab* **24**: 593-607.
Seguin, C.A., Draper, J.S., Nagy, A., and Rossant, J. (2008) Establishment of endoderm progenitors by SOX transcription factor expression in human embryonic stem cells, *Cell Stem Cell* **3**: 182-195.
Sellick, G.S., Barker, K.T., Stolte-Dijkstra, I., Fleischmann, C., Coleman, R.J., Garrett, C., et al. (2004) Mutations in PTF1A cause pancreatic and cerebellar agenesis, *Nat Genet* **36**: 1301-1305.
Sermer, M., Naylor, C.D., Gare, D.J., Kenshole, A.B., Ritchie, J.W., Farine, D., et al. (1995) Impact of increasing carbohydrate intolerance on maternal-fetal outcomes in 3637 women

without gestational diabetes. The Toronto Tri-Hospital Gestational Diabetes Project, *Am J Obstet Gynecol* **173**: 146-156.
Servitja, J.M., Pignatelli, M., Maestro, M.A., Cardalda, C., Boj, S.F., Lozano, J., et al. (2009) Hnf1alpha (MODY3) controls tissue-specific transcriptional programs and exerts opposed effects on cell growth in pancreatic islets and liver, *Mol Cell Biol* **29**: 2945-2959.
Shalem, O., Sanjana, N.E., and Zhang, F. (2015) High-throughput functional genomics using CRISPR-Cas9, *Nat Rev Genet* **16**: 299-311.
Shapiro, E.T., Tillil, H., Rubenstein, A.H., and Polonsky, K.S. (1988) Peripheral insulin parallels changes in insulin secretion more closely than C-peptide after bolus intravenous glucose administration, *J Clin Endocrinol Metab* **67**: 1094-1099.
Shepherd, M., Shields, B., Ellard, S., Rubio-Cabezas, O., and Hattersley, A.T. (2009) A genetic diagnosis of HNF1A diabetes alters treatment and improves glycaemic control in the majority of insulin-treated patients, *Diabet Med* **26**: 437-441.
Shepherd, M., Shields, B., Hammersley, S., Hudson, M., McDonald, T.J., Colclough, K., et al. (2016) Systematic Population Screening, Using Biomarkers and Genetic Testing, Identifies 2.5% of the U.K. Pediatric Diabetes Population With Monogenic Diabetes, *Diabetes Care* **39**: 1879-1888.
Shi, Z.D., Lee, K., Yang, D., Amin, S., Verma, N., Li, Q.V., et al. (2017) Genome Editing in hPSCs Reveals GATA6 Haploinsufficiency and a Genetic Interaction with GATA4 in Human Pancreatic Development, *Cell Stem Cell* **20**: 675-688 e676.
Shibasaki, Y., Kawakami, T., Kanazawa, Y., Akanuma, Y., and Takaku, F. (1985) Posttranslational cleavage of proinsulin is blocked by a point mutation in familial hyperproinsulinemia, *J Clin Invest* **76**: 378-380.
Shield, J.P. (2000) Neonatal diabetes: new insights into aetiology and implications, *Horm Res* **53 Suppl 1**: 7-11.
Shields, B.M., Hicks, S., Shepherd, M.H., Colclough, K., Hattersley, A.T., and Ellard, S. (2010) Maturity-onset diabetes of the young (MODY): how many cases are we missing?, *Diabetologia* **53**: 2504-2508.
Shih, D.Q., Screenan, S., Munoz, K.N., Philipson, L., Pontoglio, M., Yaniv, M., et al. (2001) Loss of HNF-1alpha function in mice leads to abnormal expression of genes involved in pancreatic islet development and metabolism, *Diabetes* **50**: 2472-2480.
Shoelson, S., Fickova, M., Haneda, M., Nahum, A., Musso, G., Kaiser, E.T., et al. (1983) Identification of a mutant human insulin predicted to contain a serine-for-phenylalanine substitution, *Proc Natl Acad Sci U S A* **80**: 7390-7394.
Shoelson, S., Haneda, M., Blix, P., Nanjo, A., Sanke, T., Inouye, K., et al. (1983) Three mutant insulins in man, *Nature* **302**: 540-543.
Simon-Szabo, L., Kokas, M., Mandl, J., Keri, G., and Csala, M. (2014) Metformin attenuates palmitate-induced endoplasmic reticulum stress, serine phosphorylation of IRS-1 and apoptosis in rat insulinoma cells, *PLoS One* **9**: e97868.
Skelin Klemen, M., Dolensek, J., Slak Rupnik, M., and Stozer, A. (2017) The triggering pathway to insulin secretion: Functional similarities and differences between the human and the mouse beta cells and their translational relevance, *Islets* **9**: 109-139.
Smith, S.B., Gasa, R., Watada, H., Wang, J., Griffen, S.C., and German, M.S. (2003) Neurogenin3 and hepatic nuclear factor 1 cooperate in activating pancreatic expression of Pax4, *J Biol Chem* **278**: 38254-38259.
Sone, T., Shin, M., Ouchi, T., Sasanuma, H., Miyamoto, A., Ohte, S., et al. (2019) Dual usage of a stage-specific fluorescent reporter system based on a helper-dependent adenoviral vector to visualize osteogenic differentiation, *Sci Rep* **9**: 9705.
Spracklen, C.N., Horikoshi, M., Kim, Y.J., Lin, K., Bragg, F., Moon, S., et al. (2020) Identification of type 2 diabetes loci in 433,540 East Asian individuals, *Nature* **582**: 240-245.
Steiner, D.F., Cunningham, D., Spigelman, L., and Aten, B. (1967) Insulin biosynthesis: evidence for a precursor, *Science* **157**: 697-700.
Steiner, D.F., and James, D.E. (1992) Cellular and molecular biology of the beta cell, *Diabetologia* **35 Suppl 2**: S41-48.

Steiner, D.J., Kim, A., Miller, K., and Hara, M. (2010) Pancreatic islet plasticity: interspecies comparison of islet architecture and composition, *Islets* **2**: 135-145.
Stepniewski, J., Kachamakova-Trojanowska, N., Ogrocki, D., Szopa, M., Matlok, M., Beilharz, M., et al. (2015) Induced pluripotent stem cells as a model for diabetes investigation, *Sci Rep* **5**: 8597.
Stijnen, P., Ramos-Molina, B., O'Rahilly, S., and Creemers, J.W. (2016) PCSK1 Mutations and Human Endocrinopathies: From Obesity to Gastrointestinal Disorders, *Endocr Rev* **37**: 347-371.
Stoffers, D.A., Ferrer, J., Clarke, W.L., and Habener, J.F. (1997) Early-onset type-II diabetes mellitus (MODY4) linked to IPF1, *Nat Genet* **17**: 138-139.
Stoffers, D.A., Thomas, M.K., and Habener, J.F. (1997) Homeodomain protein IDX-1: a master regulator of pancreas development and insulin gene expression, *Trends Endocrinol Metab* **8**: 145-151.
Stoffers, D.A., Zinkin, N.T., Stanojevic, V., Clarke, W.L., and Habener, J.F. (1997) Pancreatic agenesis attributable to a single nucleotide deletion in the human IPF1 gene coding sequence, *Nat Genet* **15**: 106-110.
Stoy, J., Edghill, E.L., Flanagan, S.E., Ye, H.G., Paz, V.P., Pluzhnikov, A., et al. (2007) Insulin gene mutations as a cause of permanent neonatal diabetes, *Proceedings of the National Academy of Sciences of the United States of America* **104**: 15040-15044.
Stride, A., Ellard, S., Clark, P., Shakespeare, L., Salzmann, M., Shepherd, M., and Hattersley, A.T. (2005) Beta-cell dysfunction, insulin sensitivity, and glycosuria precede diabetes in hepatocyte nuclear factor-1alpha mutation carriers, *Diabetes Care* **28**: 1751-1756.
Sun, J., Xiong, Y., Li, X., Haataja, L., Chen, W., Mir, S.A., et al. (2020) Role of Proinsulin Self-Association in Mutant INS Gene-Induced Diabetes of Youth, *Diabetes* **69**: 954-964.
Sun-Wada, G.H., Toyomura, T., Murata, Y., Yamamoto, A., Futai, M., and Wada, Y. (2006) The a3 isoform of V-ATPase regulates insulin secretion from pancreatic beta-cells, *J Cell Sci* **119**: 4531-4540.
Szopa, M., Ludwig-Galezowska, A.H., Radkowski, P., Skupien, J., Machlowska, J., Klupa, T., et al. (2016) A family with the Arg103Pro mutation in the NEUROD1 gene detected by next-generation sequencing - Clinical characteristics of mutation carriers, *Eur J Med Genet* **59**: 75-79.
Tager, H., Given, B., Baldwin, D., Mako, M., Markese, J., Rubenstein, A., et al. (1979) A structurally abnormal insulin causing human diabetes, *Nature* **281**: 122-125.
Tager, H., Thomas, N., Assoian, R., Rubenstein, A., Saekow, M., Olefsky, J., and Kaiser, E.T. (1980) Semisynthesis and biological activity of porcine [LeuB24]insulin and [LeuB25]insulin, *Proc Natl Acad Sci U S A* **77**: 3181-3185.
Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors, *Cell* **131**: 861-872.
Temple, I.K., Gardner, R.J., Mackay, D.J., Barber, J.C., Robinson, D.O., and Shield, J.P. (2000) Transient neonatal diabetes: widening the understanding of the etiopathogenesis of diabetes, *Diabetes* **49**: 1359-1366.
Teo, A.K., Lau, H.H., Valdez, I.A., Dirice, E., Tjora, E., Raeder, H., and Kulkarni, R.N. (2016) Early Developmental Perturbations in a Human Stem Cell Model of MODY5/HNF1B Pancreatic Hypoplasia, *Stem Cell Reports* **6**: 357-367.
Teo, A.K., Windmueller, R., Johansson, B.B., Dirice, E., Njolstad, P.R., Tjora, E., et al. (2013) Derivation of human induced pluripotent stem cells from patients with maturity onset diabetes of the young, *J Biol Chem* **288**: 5353-5356.
Thanabalasingham, G., Pal, A., Selwood, M.P., Dudley, C., Fisher, K., Bingley, P.J., et al. (2012) Systematic assessment of etiology in adults with a clinical diagnosis of young-onset type 2 diabetes is a successful strategy for identifying maturity-onset diabetes of the young, *Diabetes Care* **35**: 1206-1212.
Thanabalasingham, G., Shah, N., Vaxillaire, M., Hansen, T., Tuomi, T., Gasperikova, D., et al. (2011) A large multi-centre European study validates high-sensitivity C-reactive protein (hsCRP) as a clinical biomarker for the diagnosis of diabetes subtypes, *Diabetologia* **54**: 2801-2810.

Thomas, I.H., Saini, N.K., Adhikari, A., Lee, J.M., Kasa-Vubu, J.Z., Vazquez, D.M., et al. (2009) Neonatal diabetes mellitus with pancreatic agenesis in an infant with homozygous IPF-1 Pro63fsX60 mutation, *Pediatr Diabetes* **10**: 492-496.
Thomsen, S.K., Ceroni, A., van de Bunt, M., Burrows, C., Barrett, A., Scharfmann, R., et al. (2016) Systematic Functional Characterization of Candidate Causal Genes for Type 2 Diabetes Risk Variants, *Diabetes* **65**: 3805-3811.
Thomsen, S.K., Raimondo, A., Hastoy, B., Sengupta, S., Dai, X.Q., Bautista, A., et al. (2018) Type 2 diabetes risk alleles in PAM impact insulin release from human pancreatic beta-cells, *Nat Genet* **50**: 1122-1131.
Thurner, M., van de Bunt, M., Torres, J.M., Mahajan, A., Nylander, V., Bennett, A.J., et al. (2018) Integration of human pancreatic islet genomic data refines regulatory mechanisms at Type 2 Diabetes susceptibility loci, *Elife* **7**.
Tillil, H., and Kobberling, J. (1987) Age-corrected empirical genetic risk estimates for first-degree relatives of IDDM patients, *Diabetes* **36**: 93-99.
Todd, J.A., Walker, N.M., Cooper, J.D., Smyth, D.J., Downes, K., Plagnol, V., et al. (2007) Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes, *Nat Genet* **39**: 857-864.
Tontonoz, P., Hu, E., and Spiegelman, B.M. (1995) Regulation of adipocyte gene expression and differentiation by peroxisome proliferator activated receptor gamma, *Curr Opin Genet Dev* **5**: 571-576.
Toyoda, T., Mae, S., Tanaka, H., Kondo, Y., Funato, M., Hosokawa, Y., et al. (2015) Cell aggregation optimizes the differentiation of human ESCs and iPSCs into pancreatic bud-like progenitor cells, *Stem Cell Res* **14**: 185-197.
Tran, K., Li, Y., Duan, H., Arora, D., Lim, H.Y., and Wang, W. (2014) Identification of small molecules that protect pancreatic beta cells against endoplasmic reticulum stress-induced cell death, *ACS Chem Biol* **9**: 2796-2806.
Trask, B.J. (2002) Human cytogenetics: 46 chromosomes, 46 years and counting, *Nat Rev Genet* **3**: 769-778.
Tsonkova, V.G., Sand, F.W., Wolf, X.A., Grunnet, L.G., Kirstine Ringgaard, A., Ingvorsen, C., et al. (2018) The EndoC-betaH1 cell line is a valid model of human beta cells and applicable for screenings to identify novel drug target candidates, *Mol Metab* **8**: 144-157.
Udler, M.S., Kim, J., von Grotthuss, M., Bonas-Guarch, S., Cole, J.B., Chiou, J., et al. (2018) Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: A soft clustering analysis, *PLoS Med* **15**: e1002654.
Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S., and Gregory, P.D. (2010) Genome editing with engineered zinc finger nucleases, *Nat Rev Genet* **11**: 636-646.
van Burck, L., Blutke, A., Kautz, S., Rathkolb, B., Klaften, M., Wagner, S., et al. (2010) Phenotypic and pathomorphological characteristics of a novel mutant mouse model for maturity-onset diabetes of the young type 2 (MODY 2), *Am J Physiol Endocrinol Metab* **298**: E512-523.
van de Bunt, M., Lako, M., Barrett, A., Gloyn, A.L., Hansson, M., McCarthy, M.I., et al. (2016) Insights into islet development and biology through characterization of a human iPSC-derived endocrine pancreas model, *Islets* **8**: 83-95.
Varshney, A., Scott, L.J., Welch, R.P., Erdos, M.R., Chines, P.S., Narisu, N., et al. (2017) Genetic regulatory signatures underlying islet gene expression and type 2 diabetes, *Proc Natl Acad Sci U S A* **114**: 2301-2306.
Velazco-Cruz, L., Song, J., Maxwell, K.G., Goedegebuure, M.M., Augsornworawat, P., Hogrebe, N.J., and Millman, J.R. (2019) Acquisition of Dynamic Function in Human Stem Cell-Derived beta Cells, *Stem Cell Reports* **12**: 351-365.
Vethe, H., Bjorlykke, Y., Ghila, L.M., Paulo, J.A., Scholz, H., Gygi, S.P., et al. (2017) Probing the missing mature beta-cell proteomic landscape in differentiating patient iPSC-derived cells, *Sci Rep* **7**: 4780.
Vilaro, S., Palacin, M., Pilch, P.F., Testar, X., and Zorzano, A. (1989) Expression of an insulin-regulatable glucose carrier in muscle and fat endothelial cells, *Nature* **342**: 798-800.

Vionnet, N., Stoffel, M., Takeda, J., Yasuda, K., Bell, G.I., Zouali, H., et al. (1992) Nonsense mutation in the glucokinase gene causes early-onset non-insulin-dependent diabetes mellitus, *Nature* **356**: 721-722.
Voight, B.F., Scott, L.J., Steinthorsdottir, V., Morris, A.P., Dina, C., Welch, R.P., et al. (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis, *Nat Genet* **42**: 579-589.
Vujkovic, M., Keaton, J.M., Lynch, J.A., Miller, D.R., Zhou, J., Tcheandjieu, C., et al. (2020) Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis, *Nat Genet* **52**: 680-691.
Wang, H., Maechler, P., Hagenfeldt, K.A., and Wollheim, C.B. (1998) Dominant-negative suppression of HNF-1alpha function results in defective insulin gene transcription and impaired metabolism-secretion coupling in a pancreatic beta-cell line, *EMBO J* **17**: 6701-6713.
Wang, J., Takeuchi, T., Tanaka, S., Kubo, S.K., Kayo, T., Lu, D., et al. (1999) A mutation in the insulin 2 gene induces diabetes with severe pancreatic beta-cell dysfunction in the Mody mouse, *J Clin Invest* **103**: 27-37.
Warren-Perry, M.G., Manley, S.E., Ostrega, D., Polonsky, K., Mussett, S., Brown, P., and Turner, R.C. (1997) A novel point mutation in the insulin gene giving rise to hyperproinsulinemia, *J Clin Endocrinol Metab* **82**: 1629-1631.
Watanabe, N., Hiramatsu, K., Miyamoto, R., Yasuda, K., Suzuki, N., Oshima, N., et al. (2009) A murine model of neonatal diabetes mellitus in Glis3-deficient mice, *FEBS Lett* **583**: 2108-2113.
Wentworth, B.M., Schaefer, I.M., Villa-Komaroff, L., and Chirgwin, J.M. (1986) Characterization of the two nonallelic genes encoding mouse preproinsulin, *J Mol Evol* **23**: 305-312.
Wenzlau, J.M., Juhl, K., Yu, L., Moua, O., Sarkar, S.A., Gottlieb, P., et al. (2007) The cation efflux transporter ZnT8 (Slc30A8) is a major autoantigen in human type 1 diabetes, *Proc Natl Acad Sci U S A* **104**: 17040-17045.
Wesolowska-Andersen, A., Jensen, R.R., Alcantara, M.P., Beer, N.L., Duff, C., Nylander, V., et al. (2020) Analysis of Differentiation Protocols Defines a Common Pancreatic Progenitor Molecular Signature and Guides Refinement of Endocrine Differentiation, *Stem Cell Reports* **14**: 138-153.
Wettergren, A., Schjoldager, B., Mortensen, P.E., Myhre, J., Christiansen, J., and Holst, J.J. (1993) Truncated GLP-1 (proglucagon 78-107-amide) inhibits gastric and pancreatic functions in man, *Dig Dis Sci* **38**: 665-673.
Weyer, C., Bogardus, C., Mott, D.M., and Pratley, R.E. (1999) The natural history of insulin secretory dysfunction and insulin resistance in the pathogenesis of type 2 diabetes mellitus, *J Clin Invest* **104**: 787-794.
Wijesekara, N., Dai, F.F., Hardy, A.B., Giglou, P.R., Bhattacharjee, A., Koshkin, V., et al. (2010) Beta cell-specific Znt8 deletion in mice causes marked defects in insulin processing, crystallisation and secretion, *Diabetologia* **53**: 1656-1668.
Willemsen, G., Ward, K.J., Bell, C.G., Christensen, K., Bowden, J., Dalgard, C., et al. (2015) The Concordance and Heritability of Type 2 Diabetes in 34,166 Twin Pairs From International Twin Registers: The Discordant Twin (DISCOTWIN) Consortium, *Twin Res Hum Genet* **18**: 762-771.
Wilson, C.M., and Cushman, S.W. (1994) Insulin stimulation of glucose transport activity in rat skeletal muscle: increase in cell surface GLUT4 as assessed by photolabelling, *Biochem J* **299 (Pt 3)**: 755-759.
Winter, J., Klappa, P., Freedman, R.B., Lilie, H., and Rudolph, R. (2002) Catalytic activity and chaperone function of human protein-disulfide isomerase are required for the efficient refolding of proinsulin, *J Biol Chem* **277**: 310-317.
Witters, L.A., Watts, T.D., Daniels, D.L., and Evans, J.L. (1988) Insulin stimulates the dephosphorylation and activation of acetyl-CoA carboxylase, *Proc Natl Acad Sci U S A* **85**: 5473-5477.
Wobser, H., Dussmann, H., Kogel, D., Wang, H., Reimertz, C., Wollheim, C.B., et al. (2002) Dominant-negative suppression of HNF-1 alpha results in mitochondrial dysfunction, INS-1 cell apoptosis, and increased sensitivity to ceramide-, but not to high glucose-induced cell death, *J Biol Chem* **277**: 6413-6421.

Wolcott, C.D., and Rallison, M.L. (1972) Infancy-onset diabetes mellitus and multiple epiphyseal dysplasia, *J Pediatr* **80**: 292-297.
Wright, J., Wang, X., Haataja, L., Kellogg, A.P., Lee, J., Liu, M., and Arvan, P. (2013) Dominant protein interactions that influence the pathogenesis of conformational diseases, *J Clin Invest* **123**: 3124-3134.
Wu, H., Esteve, E., Tremaroli, V., Khan, M.T., Caesar, R., Manneras-Holm, L., et al. (2017) Metformin alters the gut microbiome of individuals with treatment-naive type 2 diabetes, contributing to the therapeutic effects of the drug, *Nat Med* **23**: 850-858.
Xie, R., Everett, L.J., Lim, H.W., Patel, N.A., Schug, J., Kroon, E., et al. (2013) Dynamic chromatin remodeling mediated by polycomb proteins orchestrates pancreatic differentiation of human embryonic stem cells, *Cell Stem Cell* **12**: 224-237.
Xu, B., Fan, D., Zhao, Y., Li, J., Wang, Z., Wang, J., et al. (2019) Three-Dimensional Culture Promotes the Differentiation of Human Dental Pulp Mesenchymal Stem Cells Into Insulin-Producing Cells for Improving the Diabetes Therapy, *Front Pharmacol* **10**: 1576.
Yabe, S.G., Fukuda, S., Takeda, F., Nashiro, K., Shimoda, M., and Okochi, H. (2017) Efficient generation of functional pancreatic beta-cells from human induced pluripotent stem cells, *J Diabetes* **9**: 168-179.
Yamada, T., Ishihara, H., Tamura, A., Takahashi, R., Yamaguchi, S., Takei, D., et al. (2006) WFS1-deficiency increases endoplasmic reticulum stress, impairs cell cycle progression and triggers the apoptotic pathway specifically in pancreatic beta-cells, *Hum Mol Genet* **15**: 1600-1609.
Yamagata, K., Furuta, H., Oda, N., Kaisaki, P.J., Menzel, S., Cox, N.J., et al. (1996) Mutations in the hepatocyte nuclear factor-4alpha gene in maturity-onset diabetes of the young (MODY1), *Nature* **384**: 458-460.
Yamagata, K., Nammo, T., Moriwaki, M., Ihara, A., Iizuka, K., Yang, Q., et al. (2002) Overexpression of dominant-negative mutant hepatocyte nuclear fctor-1 alpha in pancreatic beta-cells causes abnormal islet architecture with decreased expression of E-cadherin, reduced beta-cell proliferation, and diabetes, *Diabetes* **51**: 114-123.
Yamagata, K., Oda, N., Kaisaki, P.J., Menzel, S., Furuta, H., Vaxillaire, M., et al. (1996) Mutations in the hepatocyte nuclear factor-1alpha gene in maturity-onset diabetes of the young (MODY3), *Nature* **384**: 455-458.
Yan, W.X., Mirzazadeh, R., Garnerone, S., Scott, D., Schneider, M.W., Kallas, T., et al. (2017) BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks, *Nat Commun* **8**: 15058.
Yang, Q., Yamagata, K., Fukui, K., Cao, Y., Nammo, T., Iwahashi, H., et al. (2002) Hepatocyte nuclear factor-1alpha modulates pancreatic beta-cell growth by regulating the expression of insulin-like growth factor-1 in INS-1 cells, *Diabetes* **51**: 1785-1792.
Yano, H., Kitano, N., Morimoto, M., Polonsky, K.S., Imura, H., and Seino, Y. (1992) A novel point mutation in the human insulin gene giving rise to hyperproinsulinemia (proinsulin Kyoto), *J Clin Invest* **89**: 1902-1907.
Yasunaga, M., Tada, S., Torikai-Nishikawa, S., Nakano, Y., Okada, M., Jakt, L.M., et al. (2005) Induction and monitoring of definitive and visceral endoderm differentiation of mouse ES cells, *Nat Biotechnol* **23**: 1542-1550.
Yore, M.M., Syed, I., Moraes-Vieira, P.M., Zhang, T., Herman, M.A., Homan, E.A., et al. (2014) Discovery of a class of endogenous mammalian lipids with anti-diabetic and anti-inflammatory effects, *Cell* **159**: 318-332.
Yoshihara, E., Wei, Z., Lin, C.S., Fang, S., Ahmadian, M., Kida, Y., et al. (2016) ERRgamma Is Required for the Metabolic Maturation of Therapeutically Functional Glucose-Responsive beta Cells, *Cell Metab* **23**: 622-634.
Yoshinaga, T., Nakatome, K., Nozaki, J., Naitoh, M., Hoseki, J., Kubota, H., et al. (2005) Proinsulin lacking the A7-B7 disulfide bond, Ins2Akita, tends to aggregate due to the exposed hydrophobic surface, *Biol Chem* **386**: 1077-1085.
Yoshioka, M., Kayo, T., Ikeda, T., and Koizumi, A. (1997) A novel locus, Mody4, distal to D7Mit189 on chromosome 7 determines early-onset NIDDM in nonobese C57BL/6 (Akita) mutant mice, *Diabetes* **46**: 887-894.

Yu, S.S., and Kitbachi, A.E. (1973) Biological activity of proinsulin and related polypeptides in the fat tissue, *J Biol Chem* **248**: 3753-3761.
Zhang, D., Jiang, W., Liu, M., Sui, X., Yin, X., Chen, S., et al. (2009) Highly efficient differentiation of human ES cells and iPS cells into mature pancreatic insulin-producing cells, *Cell Res* **19**: 429-438.
Zhang, Q., Chibalina, M.V., Bengtsson, M., Groschner, L.N., Ramracheya, R., Rorsman, N.J., et al. (2014) Na+ current properties in islet alpha- and beta-cells reflect cell-specific Scn3a and Scn9a expression, *J Physiol* **592**: 4677-4696.
Zheng, X., Ren, W., Zhang, S., Liu, J., Li, S., Li, J., et al. (2012) Association of type 2 diabetes susceptibility genes (TCF7L2, SLC30A8, PCSK1 and PCSK2) and proinsulin conversion in a Chinese population, *Mol Biol Rep* **39**: 17-23.
Zhou, G., Myers, R., Li, Y., Chen, Y., Shen, X., Fenyk-Melody, J., et al. (2001) Role of AMP-activated protein kinase in mechanism of metformin action, *J Clin Invest* **108**: 1167-1174.
Zhu, X., Orci, L., Carroll, R., Norrbom, C., Ravazzola, M., and Steiner, D.F. (2002) Severe block in processing of proinsulin to insulin accompanied by elevation of des-64,65 proinsulin intermediates in islets of mice lacking prohormone convertase 1/3, *Proc Natl Acad Sci U S A* **99**: 10299-10304.
Zhu, Z., Li, Q.V., Lee, K., Rosen, B.P., Gonzalez, F., Soh, C.L., and Huangfu, D. (2016) Genome Editing of Lineage Determinants in Human Pluripotent Stem Cells Reveals Mechanisms of Pancreatic Development and Diabetes, *Cell Stem Cell* **18**: 755-768.
Ziegler, A.G., Hummel, M., Schenker, M., and Bonifacio, E. (1999) Autoantibody appearance and risk for development of childhood diabetes in offspring of parents with type 1 diabetes: the 2-year analysis of the German BABYDIAB Study, *Diabetes* **48**: 460-468.
Zinman, B., Lachin, J.M., and Inzucchi, S.E. (2016) Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes, *N Engl J Med* **374**: 1094.
Zuber, C., Fan, J.Y., Guhl, B., and Roth, J. (2004) Misfolded proinsulin accumulates in expanded pre-Golgi intermediates and endoplasmic reticulum subdomains in pancreatic beta cells of Akita mice, *FASEB J* **18**: 917-919.

Additional Bibliography:

American Diabetes Association “Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes—2018” Diabetes Care 2018 Jan; 41(Supplement 1): S13-S2

International Diabetes Federation (2019) IDF Diabetes Atlas 2019. In International Diabetes Federation

WHO (2016). "Diabetes." 2016, from http://www.who.int/mediacentre/factsheets/fs312/en/.

Please cite this article in press as: Wesolowska-Andersen et al., Analysis of Differentiation Protocols Defines a Common Pancreatic Progenitor Molecular Signature and Guides Refinement of..., Stem Cell Reports (2020), https://doi.org/10.1016/j.stemcr.2019.11.010

Stem Cell Reports
Resource
ISSCR
OPEN ACCESS

Analysis of Differentiation Protocols Defines a Common Pancreatic Progenitor Molecular Signature and Guides Refinement of Endocrine Differentiation

Agata Wesolowska-Andersen,[1] Rikke Rejnholdt Jensen,[5] Marta Pérez Alcántara,[1] Nicola L. Beer,[2] Claire Duff,[2] Vibe Nylander,[2] Matthew Gosden,[4] Lorna Witty,[1] Rory Bowden,[1] Mark I. McCarthy,[1,2,3] Mattias Hansson,[5] Anna L. Gloyn,[1,2,3] and Christian Honore[5,*]

[1]Wellcome Centre Human Genetics, University of Oxford, OX3 7BN Oxford, UK
[2]Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, OX3 7LE Oxford, UK
[3]NIHR Oxford Biomedical Research Centre, Churchill Hospital, OX3 7LE Oxford, UK
[4]The MRC Weatherall Institute of Molecular Medicine, University of Oxford, OX3 9DS Oxford, UK
[5]Stem Cell R&D, Novo Nordisk A/S, 2760 Måløv, Denmark
*Correspondence: clfh@novonordisk.com
https://doi.org/10.1016/j.stemcr.2019.11.010

SUMMARY

Several distinct differentiation protocols for deriving pancreatic progenitors (PPs) from human pluripotent stem cells have been described, but it remains to be shown how similar the PPs are across protocols and how well they resemble their *in vivo* counterparts. Here, we evaluated three differentiation protocols, performed RNA and assay for transposase-accessible chromatin using sequencing on isolated PPs derived with these, and compared them with fetal human pancreas populations. This enabled us to define a shared transcriptional and epigenomic signature of the PPs, including several genes not previously implicated in pancreas development. Furthermore, we identified a significant and previously unappreciated cross-protocol variation of the PPs through multi-omics analysis and demonstrate how such information can be applied to refine differentiation protocols for derivation of insulin-producing beta-like cells. Together, our study highlights the importance of a detailed characterization of defined cell populations derived from distinct differentiation protocols and provides a valuable resource for exploring human pancreatic development.

INTRODUCTION

Human pluripotent stem cells (hPSCs) have tremendous potential for modeling human diseases *in vitro* as well as for regenerative medicine in degenerative diseases. However, the realization of both these applications of hPSCs is dependent on the ability to derive the relevant cell lineages from hPSCs by directed differentiation.

In the context of pancreas development, studies in mice have demonstrated that exocrine, ductal, and endocrine lineages all derive from multipotent pancreatic progenitor (PP) cells, defined by co-expression of several transcription factors (TFs), including PDX1, NKX6.1, PTF1a, and SOX9 (Larsen and Grapin-Botton, 2017). Despite noteworthy differences in human pancreas development compared with mouse (Jennings et al., 2015; Nair and Hebrok, 2015), human PPs express a similar core network of TFs, including PDX1 and NKX6.1 (Petersen et al., 2018). When transplanted into immunocompromised mice, the hPSC-derived PPs are able to give rise to all lineages of the pancreas (Kelly et al., 2011; Kroon et al., 2008; Rezania et al., 2012, 2013), supporting their similarity to multipotent PPs observed during development. Knowledge gained from rodent models of pancreas development facilitated many of the advancements in differentiation protocols. For example, retinoic acid and fibroblast growth factor signaling are indispensable for the specification and expansion of PPs during development (Bhushan et al., 2001; Molotkov et al., 2005), and the majority of current differentiation protocols include agonists of these signaling pathways. However, there are also notable differences in protocols reported to differentiate hPSCs to PPs. For example, bone morphogenetic protein (BMP) signaling has been shown to promote a liver fate choice rather than pancreas development (Wandzioch and Zaret, 2009), and thus several protocols include BMP inhibitors during differentiation. However, a recent report argued for the exclusion of BMP inhibitors, since these were shown to promote a premature endocrine differentiation at the expense of PDX1/NKX6.1-positive PPs (Russ et al., 2015). There is also no consensus on inclusion of other pathway modulators, such as epidermal growth factor (EGF) or protein kinase C (PKC) agonists, in the differentiation protocols (Nostro et al., 2015; Rezania et al., 2014; Russ et al., 2015).

As hPSC-derived PPs are often defined by co-expression of a limited set of genes (e.g., *PDX1* and *NXK6.1*), it remains unclear how similar the PP cells derived from various differentiation protocols are and how well they represent embryonic development and subsequent development of more mature cell types of the pancreatic islet.

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Please cite this article in press as: Wesolowska-Andersen et al., Analysis of Differentiation Protocols Defines a Common Pancreatic Progenitor Molecular Signature and Guides Refinement of..., Stem Cell Reports (2020), https://doi.org/10.1016/j.stemcr.2019.11.010

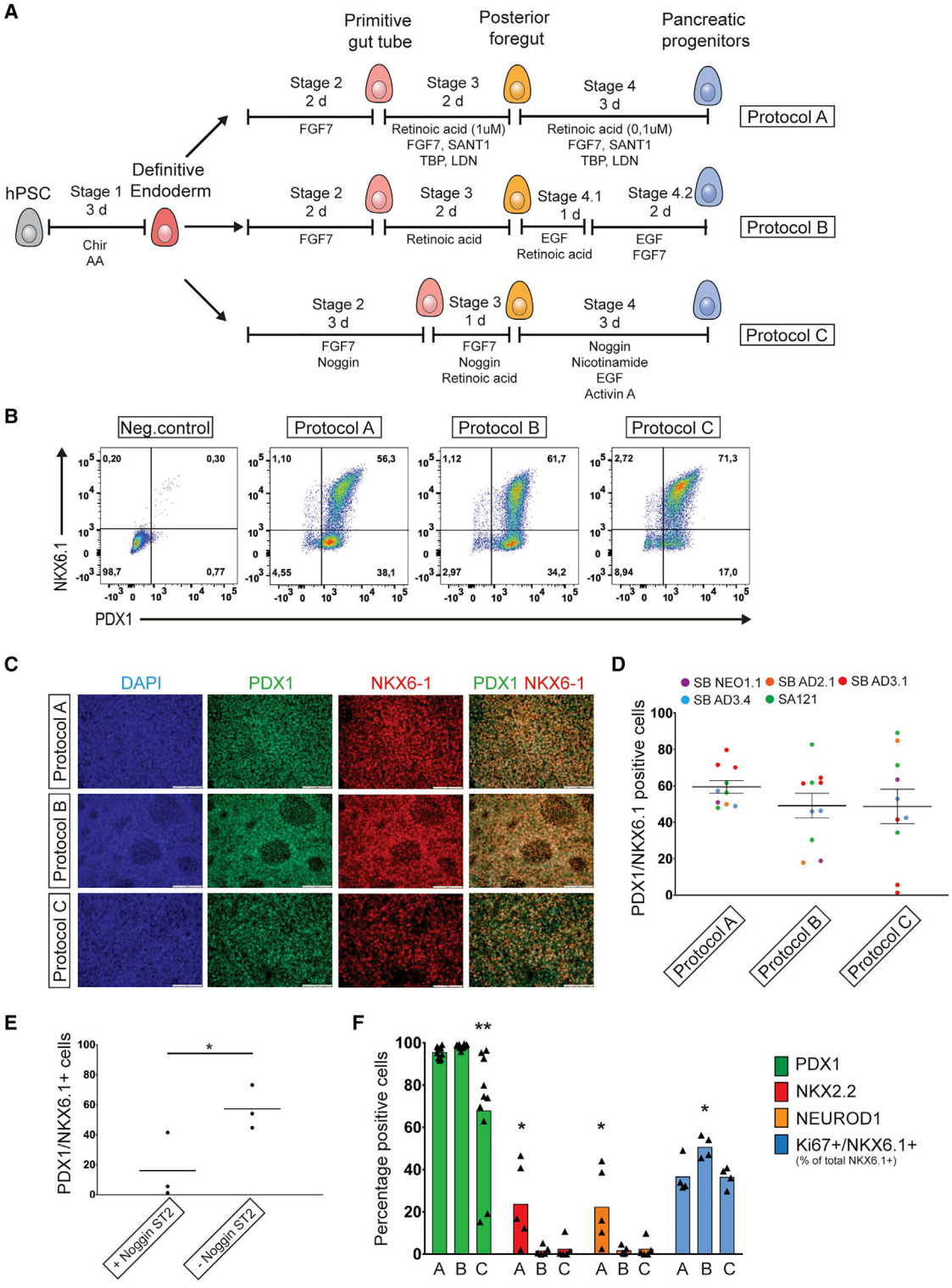


Figure 1. Derivation of PPs from hPSCs Using Multiple Differentiation Protocols

(A) Schematic outline of the three PP protocols applied.

(B) Representative examples of flow cytometry pseudo color dot plots of PPs from the three protocols stained for PDX1 and NKX6.1. Negative control is definitive endoderm cells.

(legend continued on next page)

To address these questions, we performed a detailed characterization of PPs derived using three differentiation protocols adapted from recent publications (Nostro et al., 2015; Rezania et al., 2014; Russ et al., 2015), although with several differences compared with the original publications. We were able to recapitulate several aspects of the reported protocols and achieved efficient differentiation toward PPs across all three protocols. Isolation of PPs allowed us to define their comprehensive gene expression and chromatin accessibility signatures shared across all protocols, which shed light on new endocrine pancreas development biology and will serve as a valuable resource for future studies in disease modeling and cellular replacement therapies. We also highlight several notable differences in the omics profiles of the PPs derived with the different protocols, which translate to differences in their ability to differentiate further toward the endocrine lineage. Finally, we illustrate the utility of these datasets to further optimize various stages of the differentiation protocols to improve the endocrine differentiation and the derivation of beta-like cells from hPSCs.

RESULTS

Efficient Derivation of PPs from Multiple hPSC Lines Using Three Distinct Differentiation Protocols

To compare PP differentiation protocols, we first applied a common protocol for efficient derivation of definitive endoderm (DE) from hPSCs (Rezania et al., 2014; Perez-Alcantara et al., 2018) (Figures S1A and S1B). DE was differentiated toward PPs using protocols outlined in Figure 1A. The three differentiation protocols (A, B, and C) were adapted from previously established protocols (Nostro et al., 2015; Rezania et al., 2014; Russ et al., 2015), but with several differences compared with the original publications (protocol details outlined in Supplemental Experimental Procedures, including key differences from the published protocols). Differentiation efficiency was assessed by flow cytometer and immunofluorescence imaging analysis of expression of the two key PP TFs: PDX1 and NKX6.1 (Figures 1B–1D). Across five hPSC lines we observed an average differentiation efficiency of 59%, 49%, and 48% PDX1/NKX6.1-positive cells derived using protocols A, B, and C, respectively (Figure 1D). A notable variation in the efficiency of PDX1/NKX6.1-positive cells was observed across differentiations (Figure 1D), which may be attributed to the different abilities of individual hPSC lines to differentiate to PPs across these protocols. The hPSC line SB AD3.1 displayed poor differentiation efficiency using protocol C (Figures 1D and S1C), but omission of Noggin during stage 2 of protocol C led to a significant improvement of its differentiation (Figure 1E). Extending this observation, we evaluated the effect of Noggin during stage 2 of all three protocols on the PP differentiation efficiency; however, we found that the effect was not consistent across either cell lines or differentiation protocols (Figure S1C).

A large proportion of the PDX1/NKX6.1-positive cells across the three protocols were proliferating, as demonstrated by the expression of the proliferation marker Ki67 (Figure 1F). Protocols B and C were originally reported to generate a high percentage of PPs, while limiting the commitment to the endocrine lineage (Russ et al., 2015; Nostro et al., 2015). In agreement with this, we observed very few cells expressing the endocrine markers NEUROD1 and NKX2.2 in protocol B and C, whereas protocol A gave rise to a significantly higher number of endocrine cells at this stage of the protocol (Figure 1F). Most of the endocrine progenitors generated were NKX6.1 negative (Figures S1D and S1E), in agreement with previous studies (Nostro et al., 2015; Petersen et al., 2017; Russ et al., 2015). In summary, we are able to recapitulate key aspects of the original reports and we show that all three differentiation protocols efficiently derive PDX1/NKX6.1-positive PPs across multiple hPSC lines.

Global Gene Expression and Chromatin Accessibility Analysis of hPSC-Derived PPs

We next aimed to characterize the PPs derived from the three differentiation protocols in more detail. We differentiated an NKX6.1-GFP reporter human induced PSC (hiPSC) line (Gupta et al., 2018) to PPs using the three

(C) Immunofluorescence images of PPs stained for PDX1 and NKX6.1. Scale bar, 100 μm.

(D) Quantification of PDX1 and NKX6.1 co-expressing cells based on the flow cytometric analysis shown in (B). Graph shows a scatterplot of the mean ± SEM of five individual hPSC lines. Dots are color coded according to individual cell lines (details in Figure S1C). n = 10 independent experiments.

(E) Percentage of PDX1 and NKX6.1 co-expressing cells from the SB AD3.1 hiPSC line differentiated with protocol C with or without 50 ng/mL Noggin included during stage two; n = 3 independent experiments; *p < 0.05, paired t test.

(F) Quantification of PDX1, NKX2.2, NEUROD1, and percentage NKX6.1+ cells co-expressing Ki67. Bars show means and dots represent individual differentiations. One-way ANOVA with the Tukey test for multiple comparisons, *p < 0.05, **p < 0.01, different from the two other groups. PDX1, n = 10 independent experiments, same hPSC lines as in (D). NKX2.2 and NEUROD1, n = 5 independent experiments, one for each of the following hPSC lines: SA121 hESC, SB NEO1.1 hiPSC, SB AD2.1 hiPSC, SB AD3.1 hiPSC, SB AD3.4 hiPSC. Ki67, n = 4 independent experiments, three using SB AD3.1 hiPSC and one using SA121 hESC.

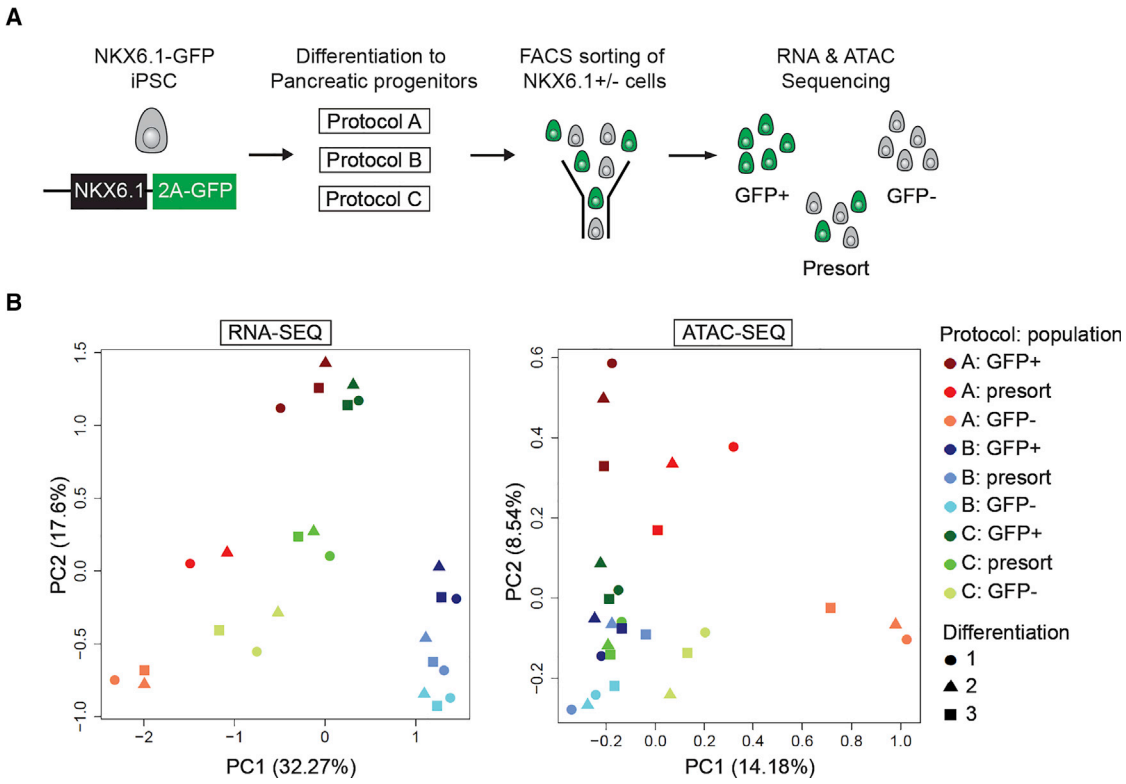


Figure 2. Global Gene Expression and Chromatin Accessibility Analysis of FACS-Isolated PP Populations

(A) Schematic showing the experimental setup. NKX6.1-GFP hiPSCs were differentiated side by side using all three protocols and GFP+ and GFP− cells as well as unsorted cells were collected following FACS for RNA and ATAC sequencing. Cells were collected from three independent differentiations of all three protocols.

(B) Principal component analysis (PCA) of RNA-seq (left) and ATAC-seq data (right). Legend applies to both PCA plots.

differentiation protocols and isolated both GFP+ and GFP− cells using fluorescence-activated cell sorting (FACS) (Figure 2A). Post-sort analysis of live cells demonstrated isolation of highly pure GFP+ and GFP− populations across all three protocols (Figures S2A and S2B) and efficient differentiation of the reporter line to PPs by all protocols was confirmed by flow cytometry analysis of GFP and NKX6.1 expression (Figures S2C and S2D). The isolated populations were processed for genome-wide transcriptome and chromatin accessibility analysis by RNA sequencing (RNA-seq) and assay for transposase-accessible chromatin (ATAC) using sequencing (ATAC-seq), respectively. Global principal component analysis (PCA) of the datasets (Figure 2B) revealed that samples clustered by protocol and by sorted cell population. Samples generated with protocol A were the most distinct and showed the largest separation of the GFP-sorted cell populations in both datasets. For all three protocols, we observed the presort samples clustering in between the sorted populations, as expected. In addition, we confirmed that expression of *NKX6.1* and *PDX1* was enriched in the GFP+ population for all protocols (Figures S2E and S2F).

Cross-Protocol PP Gene Expression and Chromatin Accessibility Signature

We then explored the similarities of the omics profiles of the PPs derived with the different protocols. For that purpose, we compared the samples generated in this study with gene expression and open chromatin profiles of our previously published differentiation model (Perez-Alcantara et al., 2018) across all seven stages of hPSC differentiation toward beta-like cells. The PCA analysis for both the RNA-seq (Figure 3A) and ATAC-seq (Figure 3E) datasets revealed that all the samples characterized in this study clustered, as expected, together with the pancreatic endoderm stage samples from the control dataset. Interestingly, the GFP− cells generated with protocol A clustered out close to the cells from the subsequent differentiation stage, the endocrine progenitors. This was most pronounced in the RNA-seq data, but also apparent in the ATAC-seq data, and is in line with the higher percentage of endocrine cells observed in the NKX6.1− cells with this protocol (Figures 1F, S1D, S1E, S2C, and S2D).

We then sought to derive a PP gene expression signature common to all three protocols by comparing the

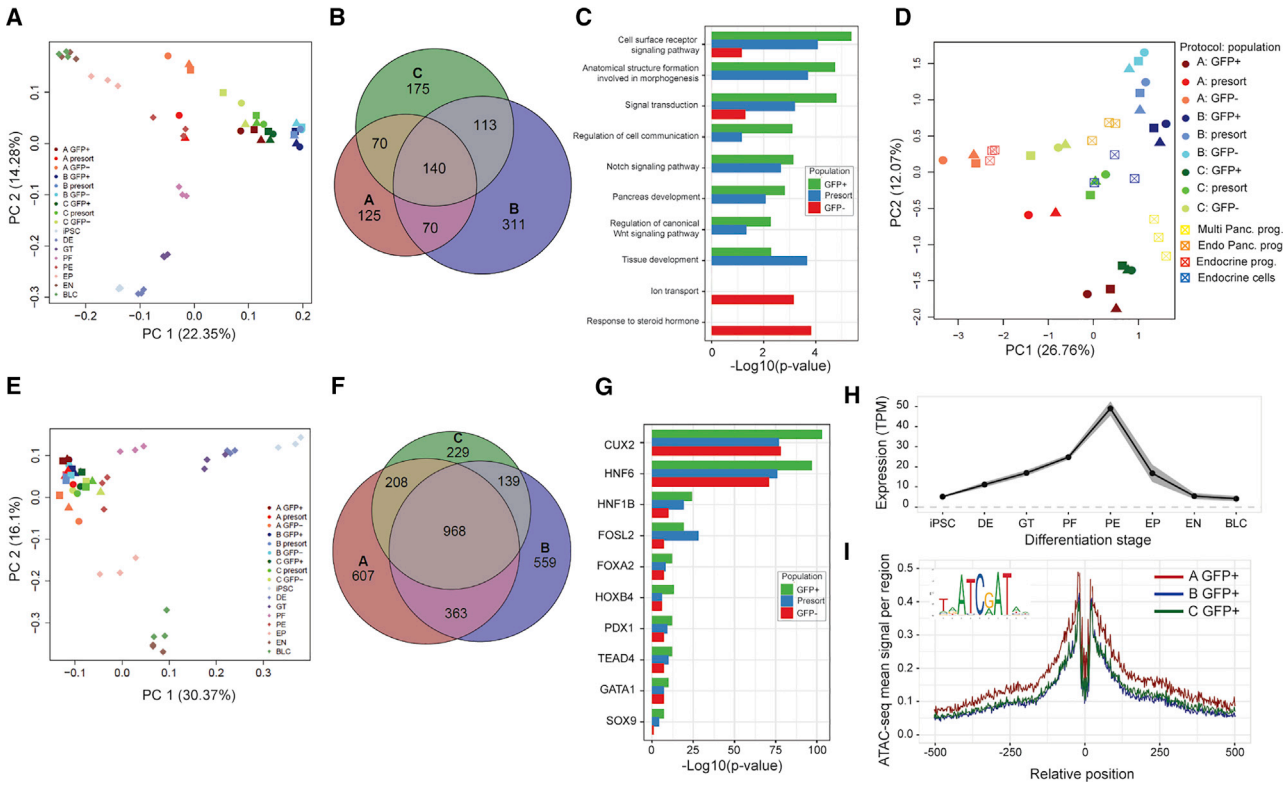


Figure 3. Common Transcriptomic and Epigenomic PP Signatures across Three Differentiation Protocols

(A) PCA of RNA-seq samples together with data collected at all stages of the hPSC differentiation toward beta-like cells generated with protocol A (Perez-Alcantara et al., 2018).

(B) Venn diagram of the RNA-seq GFP+ PP signatures generated separately for each of the protocols by comparison with cells at the other differentiation stages.

(C) Selected gene ontology enrichment of the common PP signature genes across the three protocols; shown separately for signatures derived with presort cells (in blue), GFP+ (in green), and GFP− (in red) cell populations. The length of the bar represents −log10 of the enrichment p value.

(D) PCA of RNA-seq samples from this study, together with transcriptomes of fetal pancreas cell subpopulations (Ramond et al., 2018).

(E) PCA of ATAC-seq samples from this study, together with data collected at all stages of the hPSC differentiation toward beta-like cells, generated with protocol A (Perez-Alcantara et al., 2018).

(F) Venn diagram of the ATAC-seq GFP+ PP signatures generated separately for each of the protocols by comparison with cells at the other differentiation stages.

(G) Enrichment of selected TFs within the common PP signature open chromatin peaks across the three protocols; shown separately for signatures derived with presort cells (in blue), GFP+ (in green), and GFP− (in red) cell populations.

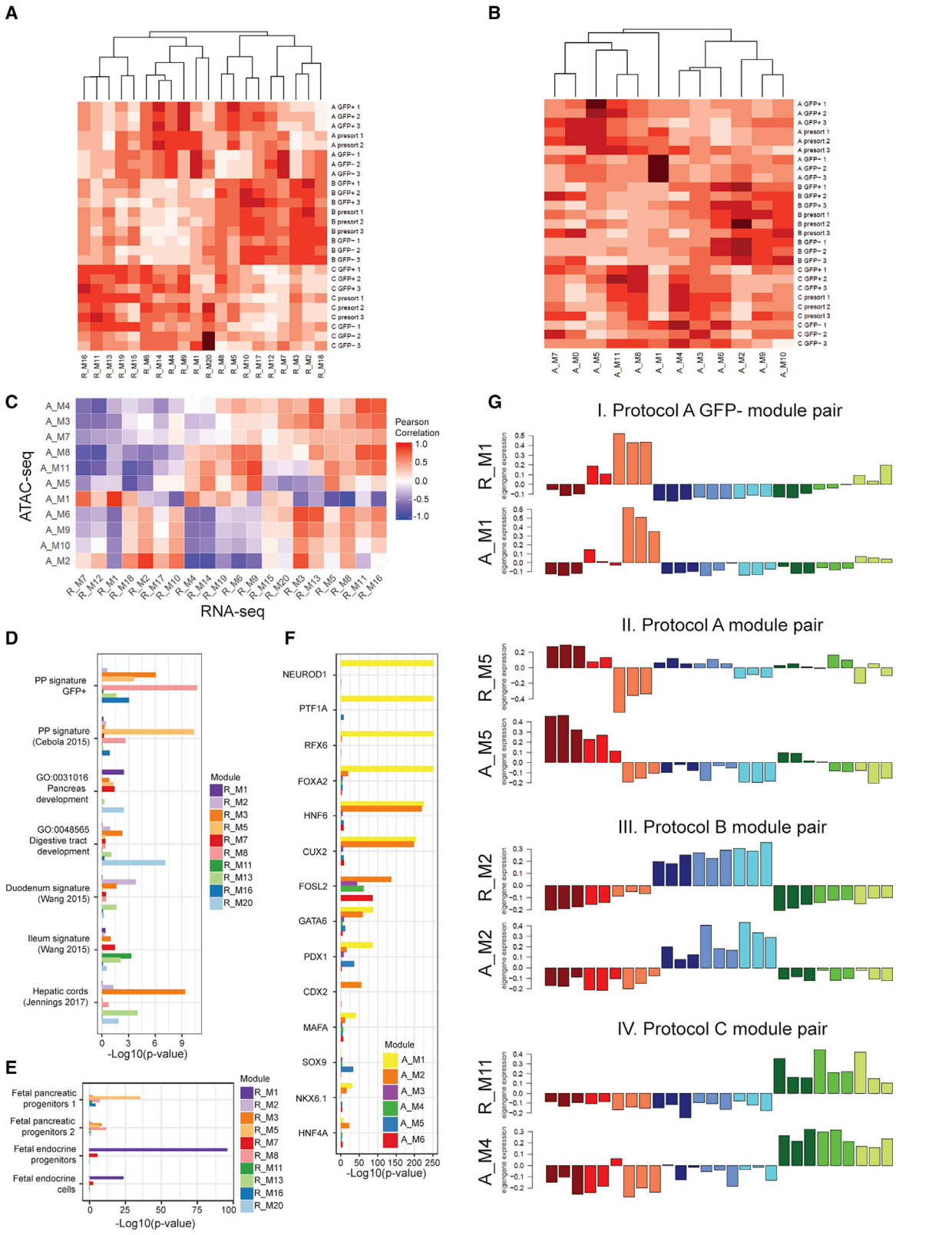
(H) Mean expression of CUX2 gene across all stages of hPSC differentiation toward beta-like cells. Shaded gray area indicates ± SEM (Perez-Alcantara et al., 2018).

(I) Footprinting analysis of CUX2 binding motifs within open chromatin peaks of the GFP+ populations generated with the three differentiation protocols.

BLC, beta-like cells; EN, endocrine cells; ; EP, endocrine progenitors; GT, gut tube; PE, pancreatic endoderm/progenitors; PF, posterior foregut.

transcriptome profiles of the GFP+ cell populations from each of the three protocols separately with the gene expression profiles of the remaining differentiation stages. This resulted in 405, 634, and 498 genes defining the PPs in protocols A, B, and C, respectively, and we next defined the 140-gene common cross-protocol signature by intersecting the three individual protocol-specific signatures (Figure 3B).

In a similar manner, we defined a 112-gene PP signature within the presorted cells, and a 129-gene signature within the GFP− cell populations (Table S1). Functional enrichment analysis revealed that several genes forming the signature were involved in cell surface receptor signaling pathways, including Notch and Wnt pathways, and included several genes previously implicated in pancreas



(legend on next page)

development (Figure 3C, Table S1). In addition, for an unbiased comparison with the transcriptomes of PPs *in vivo*, we compared the expression of the samples generated in this study with recently published transcriptomes of human fetal pancreas subpopulations (Ramond et al., 2018) (Figure 3D). We observed that the GFP− population derived with protocol A resembled closely the fetal endocrine progenitor population, in agreement with the high proportion of endocrine differentiation observed with this protocol. Interestingly, the GFP+ populations of protocols A and C clustered close to the fetal multipotent PPs, suggesting a high level of similarity between these *in vitro*-generated PP and human fetal multipotent PP cells. The other two fetal pancreatic cell subpopulations, comprising endocrine-biased PP and endocrine cells, clustered in between the presorted cell populations generated by protocols B and C.

In a similar manner, we derived the open chromatin PP signature, comprising a total of 968 ATAC-seq peaks within the GFP+ population (Figure 3F, Table S1). These open chromatin regions were significantly enriched in binding sites of CUX2 and HNF6 as well as several other TFs with previously established functions in pancreatic development and links to monogenic diabetes, including HNF1B, FOXA2, and PDX1 (Figure 3G). CUX2 was the most significantly enriched TF; however, it has to our knowledge not been implicated in pancreatic development. *CUX2* was expressed in the PP cell populations derived with all three protocols, and its expression was highest at this stage of differentiation in the control lines (Figure 3H). We have subsequently performed a footprinting analysis of the ATAC-seq peaks containing the CUX2 binding motif (Figure 3I) and observed a clear dip in the sequencing coverage around the investigated CUX2 sites, supporting the hypothesis that this TF is bound within the open chromatin at this stage.

Correlated Modules of Gene Expression and Open Chromatin Highlight Between-Protocol Differences

We next sought to determine the differences in the transcriptional and chromatin accessibility profiles of the PPs derived with the different protocols. We applied weighted gene co-expression network analysis (WGCNA) (Langfelder and Horvath, 2008) to our datasets, and conducted further analyses on the dimensionality reduced sets of modules combining correlated patterns of gene expression and open chromatin. We defined 20 modules of co-expressed genes from the RNA-seq data (Figure 4A, Table S2), as well as 11 modules of co-open chromatin sites from the ATAC-seq data (Figure 4B, Table S2). We then identified highly correlated pairs of RNA-seq and ATAC-seq modules, likely representing the modules of active open chromatin regulating the correlated genes' expression (Figure 4C). For each of the gene expression modules, we performed hypergeometric testing of enrichment for significant overlap with gene signatures of PPs from this and previous studies (Cebola et al., 2015), genes involved in selected developmental processes (Figure 4D, Table S2), as well as gene signatures of fetal pancreas cell subpopulations (Ramond et al., 2018) (Figure 4E, Table S2). For each of the open chromatin modules, we identified TF binding motifs enriched within the module's peaks (Figure 4F). We highlight a few of the most interesting correlated RNA-ATAC gene module pairs in Figure 4G.

The most distinct gene expression and correlated open chromatin pattern observed was for RNA module R_M1 and ATAC module A_M1 (Figure 4GI). Both these modules had the highest eigengene values, corresponding to the first principal component and summarizing the overall pattern of gene expression for all genes in the module, in the GFP− cell population from protocol A. The R_M1 genes were enriched in genes involved in pancreas development,

Figure 4. Protocol-Specific Differences in Transcriptomic and Epigenomic Profiles of PPs Generated with Three Differentiation Protocols

(A) Heatmap of RNA-seq co-expressed gene modules eigengenes across all RNA-seq samples. Higher red color intensities indicate higher eigengene values.

(B) Heatmap of ATAC-seq co-open chromatin modules eigengenes across all ATAC-seq samples. Higher red color intensities indicate higher eigengene values.

(C) Pairwise Pearson correlation heatmap of module eigengene values for RNA-seq and ATAC-seq modules.

(D) Hypergeometric enrichment p values for gene signatures of PP signatures, selected developmental gene ontologies, and gene signatures of selected intestinal and hepatic tissue/cell types. Only modules with enrichment p values <0.01 for any of the selected categories in (D) and (E) are plotted. Full list is available in Table S1.

(E) Hypergeometric enrichment p values for transcriptomic signatures of *in vivo* fetal pancreatic cell subpopulations from Ramond et al. (2018). Only modules with enrichment p values <0.01 for any of the selected categories in (D) and (E) are plotted.

(F) Enrichment p values of selected known TF binding sites within modules of co-open chromatin. Only modules with enrichment p values <1e−20 for any of the selected TFs are plotted.

(G) Selected highly correlated pairs of RNA-seq and ATAC-seq modules eigengenes. The height of the bars represents the eigengene values for each sample in the presented module. Bars are colored by each protocol/cell population, as outlined in figure legends for PCA plots in Figure 2B, and in other subsequent figures in this paper. I, protocol A GFP− module pair; II, protocol A GFP+ module pair; III, protocol B module pair; IV, protocol C module pair.

and very closely resembled the *in vivo* fetal endocrine progenitors, which was in line with the previous PCA analysis (Figure 3D). The A_M1 peaks were significantly enriched in binding sites of several TFs with known roles in pancreatic and endocrine development, including NEUROD1, PTF1A, RFX6, and FOXA2. This evidence corroborates that the GFP− population of protocol A represents a population of early endocrine progenitors, which is absent in the other two protocols.

Genes preferentially expressed in the GFP+ population of protocol A, and correlated open chromatin peaks, were captured in modules R_M5 and A_M5 (Figure 4GII). Importantly, these genes were also expressed in the GFP+ cell populations generated with protocols B and C, albeit at lower depth. The R_M5 genes were enriched in PP signatures derived earlier in this study, as well as in a previous study by Cebola and colleagues (Cebola et al., 2015), and were enriched in the transcriptomic signature of the *in vivo* multipotent PPs (Ramond et al., 2018). Open chromatin in module A_M5 was enriched in binding motifs of PDX1, SOX9, and PTF1a, all known as markers of PPs. This module pair represents the PP signatures preferential to protocol A but also present in the GFP+ populations from the other two protocols.

RNA modules R_M2 and R_M3 were both highly correlated with ATAC module A_M2 (Figure 4GIII) and represented genes preferentially expressed in all cell populations generated with protocol B. We noticed that the corresponding A_M2 open chromatin peaks were enriched in binding sites of HNF6 and CUX2, characteristic of the previously described PP open chromatin signature. However, we also noticed that the A_M2 module was the only module with significant enrichment in CDX2 binding sites. Based on previous reports of CDX2 regulating patterning of the intestinal epithelium (Grainger et al., 2010), we hypothesized that some of the cells differentiated with protocol B might be more likely to assume an intestinal fate. In line with this hypothesis, we found that R_M2 module genes were enriched in the intestinal gene signature of duodenum, and M3 genes in genes involved in digestive tract development and hepatic cords signature.

Finally, RNA gene modules M11, M13, and M16 were preferentially expressed in cells generated with protocol C. All these gene modules were correlated with ATAC module M4, particularly enriched for the FOSL2 binding sites. FOSL2 has not previously been implicated in pancreatic development. M16 module genes were enriched in the transcriptomic signature of the *in vivo* fetal multipotent PPs; however, both modules M11 and M16 showed enrichment in ileum gene signature, and M13 additionally in hepatic cords signature.

Given the magnitude of differences between the PPs derived with the different protocols reported here, we hypothesized that among the differentially expressed genes we might also find genes with well-established roles in pancreatic and endocrine development, and therefore of particular interest to the scientific community. We therefore investigated in more detail the gene expression profiles of genes previously implicated in maturity-onset diabetes of the young and neonatal diabetes (Flanagan et al., 2014; Yang and Chan, 2016) (Figure S3). For several genes we observed significant differences in the magnitude of expression between protocols, but we also found that some genes were only highly expressed in one or two of the protocols (e.g., *NKX2.2* in protocol A, *PPARG* in protocol B, *NEUROG3* in protocols A and C). We also observed some notable differences in the sorted cell population where a gene was preferentially expressed (e.g., the *RFX6* gene was enriched in the GFP− population of protocols A and C, while for protocol B both the GFP+ and GFP− cell population expressed *RFX6*). We anticipate that findings of any hPSC disease modeling studies for the genes highlighted here would be highly dependent on the choice of differentiation protocol.

In summary, analysis of pairs of correlated modules of gene expression and open chromatin provide an objective way to characterize the cell populations common and specific to each protocol. In addition, through gene ontology and transcription motif enrichment analyses, we were able to highlight the distinct developmental programs active in different protocols, as well as the TFs likely to control them.

Variable Efficiency of Differentiation of PPs Toward the Endocrine Lineage

We next sought to evaluate the ability of the PPs derived from the three protocols to differentiate toward the pancreatic endocrine lineage. To this end, we applied stage 5 and stage 6 of the pancreas endocrine differentiation protocol developed by Rezania et al. (2014) (Figure S4A). While this protocol did not give rise to mature endocrine cells in our hands (Figure S4B), we have previously shown it to be useful for studying endocrine development *in vitro* (Petersen et al., 2017; Ramond et al., 2018). The endocrine differentiation was evaluated for expression of the endocrine markers NEUROD1 and NKX2.2 at stage 5 and for expression of C-peptide, a marker of endogenous insulin production, at stage 6. Interestingly, protocol A displayed expression of NEUROD1 and NKX2.2 in a significant percentage of cells after three days of differentiation, whereas the endocrine induction from protocol B and C was significantly lower (Figures S4C and S4D). The poor endocrine induction of protocols B and C was also evident at stage 6 of the differentiation protocol, where only very few C-peptide+ cells were observed from protocols B and C (Figures S4E–S4G). In contrast, protocol A displayed a robust

induction of C-peptide+ cells at stage 6 of the protocol, with several of C-peptide+ cells co-expressing NKX6.1, indicative of differentiation toward beta-like cells (Figures S4E–S4G).

Improved Pancreatic Endocrine Differentiation Following Reduction of the Intestinal Marker CDX2

We next examined whether the RNA-seq and ATAC-seq profiles of the PPs could inform us on how to improve the endocrine differentiation. As described previously, we identified a gene module R_M2 preferentially expressed in protocol B that showed enrichment in an intestinal gene expression signature. We found that the correlated open chromatin module A_M2 was enriched in binding sites of the canonical intestinal TF CDX2. We confirmed that CDX2 was a member of the R_M2 gene module, and its expression was enriched in the NKX6.1+ cells of protocol B compared with protocols A and C (Figures 5A and 5B). We also observed CDX2-positive cells in protocol A, but in contrast to protocol B these were primarily NKX6.1-negative cells (Figures 5A, 5B, and S5). Omission of the BMP antagonist from protocol A led to a substantial increase in the CDX2-positive cells at the expense of NKX6.1-positive cells, suggesting that BMP signaling must be actively inhibited to prevent differentiation toward intestinal lineages. In contrast, addition of EGF during stage 4 of protocol A resulted in a significant reduction in CDX2-positive cells compared with the standard protocol A (Figure S5).

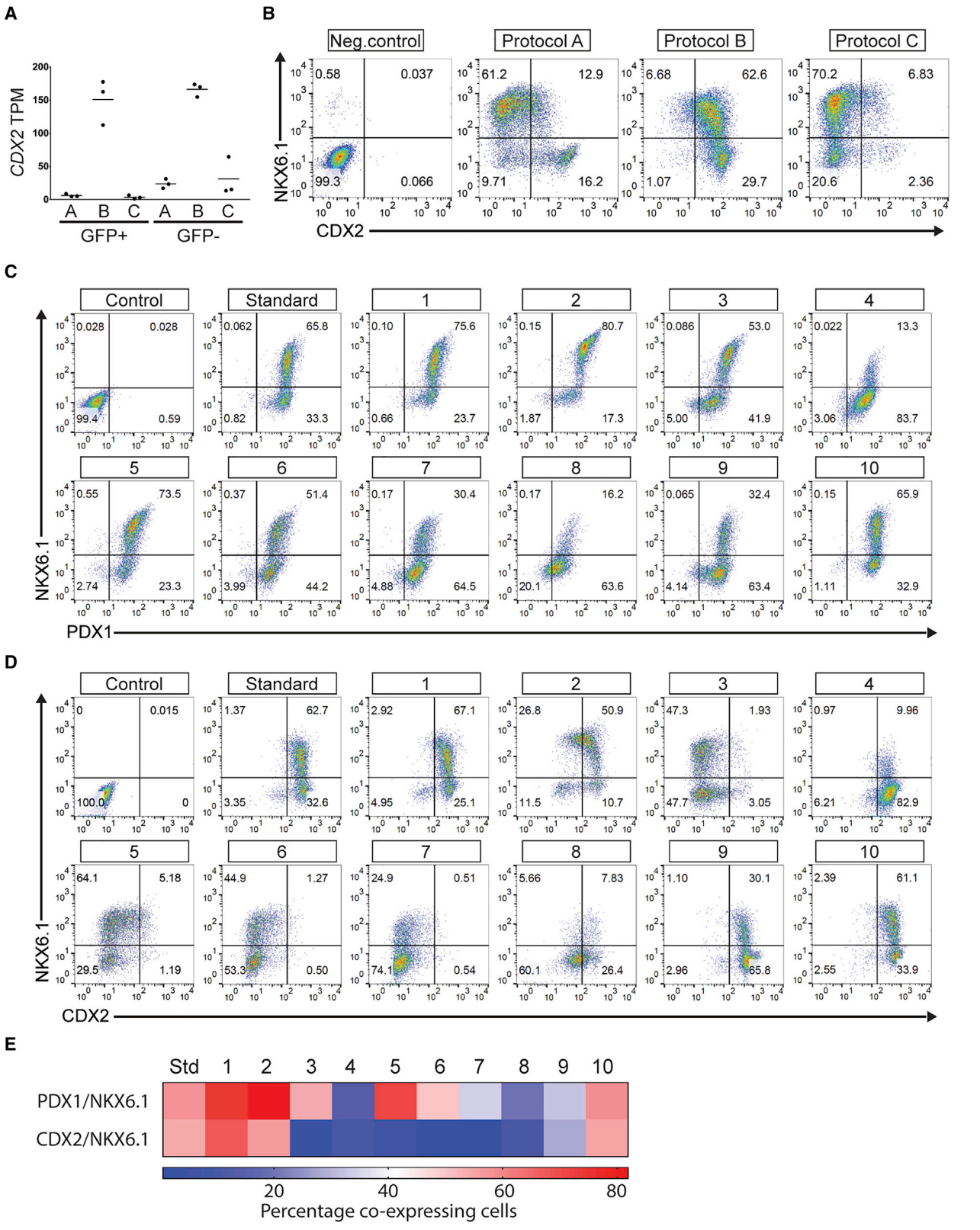
We hypothesized that the intestinal gene expression signature detected in PPs of protocol B could explain the low efficiency of differentiation toward the endocrine lineage (Figures S4C and S4D). To test this hypothesis, we evaluated ten modifications of protocol B, including several conditions with BMP antagonists aiming at maintaining the expression of the PP markers PDX1 and NKX6.1, while reducing the expression of CDX2. Interestingly, several of the modified conditions of protocol B maintained a high proportion of PDX1 and NKX6.1-positive cells (Figures 5C and 5E), but only two of these conditions (condition 3 and 5) also demonstrated a substantial reduction in CDX2 expression (Figures 5D and 5E). We subsequently evaluated the expression of select markers of the pancreatic and intestinal lineage in one of the modifications of protocol B resulting in reduction of CDX2 expression (condition 3). As expected, *CDX2* mRNA expression was downregulated in the modified protocol, as were several other genes from the intestinal gene expression signature, while several genes related to pancreas development were upregulated (Figure S5D).

We then tested whether the reduced CDX2 expression in protocol B resulted in improved ability to differentiate toward the pancreatic endocrine lineage. To this end, we differentiated PPs with protocol B, as well as its two modifications resulting in reduced CDX2 expression toward endocrine progenitors and endocrine cells, as described above (Figure S4A). Interestingly, inclusion of Noggin in stage 2–4 of protocol B (condition 3) resulted in a significant increase in the number of cells expressing the endocrine progenitor markers NEUROD1 and NKX2.2 (Figure 6A), compared with protocol B without the presence of a BMP inhibitor. Despite the reduced CDX2 expression in PPs in condition 5, this did not result in an improved differentiation toward the endocrine lineage (Figure 6A). To test whether the increased endocrine differentiation observed with the inclusion of Noggin in stage 2–4 of protocol B led to an increase in generation of beta-like cells, the cells were differentiated to the end of stage 6 (Rezania et al., 2014) (Figure S4A). We observed a significant increase in cells co-expressing the beta cell markers C-peptide and NKX6.1, when Noggin was included during stage 2–4 of protocol B (condition 3) compared with the standard protocol B (Figures 6B–6D).

DISCUSSION

In the present study, we applied three differentiation protocols for deriving PPs from hPSCs (Nostro et al., 2015; Rezania et al., 2014; Russ et al., 2015). All three protocols allowed for efficient generation of hPSC-derived PPs at levels comparable with the original reports. Interestingly, we noticed that certain hPSC lines displayed varying efficiency depending on the protocol applied. Variation in the ability to differentiate toward various lineages across hPSC lines has previously been reported (Osafune et al., 2008; Rouhani et al., 2014); however, our study highlights that such differences could possibly also be specific to the protocol used for deriving a particular cell lineage. We further demonstrated that modifications can be made to protocols, allowing for increasing the efficiency of otherwise poorly differentiating hPSC lines. Exclusion of Noggin during stage 2 rescued the ability of one hPSC line to generate PPs with protocol C. Previous studies investigated the effect of modulating the BMP signaling pathway during specification to the pancreatic lineage (Nostro et al., 2011; Shahjalal et al., 2014) but our results suggest that BMP signaling may influence the cell lineage choice at even earlier stages of differentiation. The effect of Noggin appeared to be dependent on the differentiation protocols but also of the hPSC lines applied. This suggests that variation in endogenous BMP signaling during the early stages of differentiation can influence the choice of fate of hPSC lines and modulation thereof could be an avenue for further protocol optimization for hPSC lines that prove difficult to differentiate toward the pancreatic lineage.

Please cite this article in press as: Wesolowska-Andersen et al., Analysis of Differentiation Protocols Defines a Common Pancreatic Progenitor Molecular Signature and Guides Refinement of..., Stem Cell Reports (2020), https://doi.org/10.1016/j.stemcr.2019.11.010



(legend on next page)

We performed RNA-seq and ATAC-seq of isolated NKX6.1-positive cells, which allowed us to define a transcriptomic and epigenomic signature for *in vitro*-generated PPs from the three differentiation protocols, through comparison with equivalent data from all the stages of hPSC differentiation toward beta-like cells. Of note, this approach excludes genes expressed at multiple stages of the hPSC differentiation toward beta-like cells and therefore not unique to the PP stage, including *PDX1* and *NKX6.1*, which are continuously expressed from the posterior foregut and pancreatic progenitor stages of the differentiation protocol, respectively. Globally, the gene expression and chromatin accessibility profiles for cells generated with all three protocols clustered together with the pancreatic endoderm stage samples of the control lines. *In vivo* differentiation of hPSC-derived PPs to the specific lineages of the pancreas is considered the most stringent method for evaluating the competence of these cells. Thus, one limitation in the current study is that we did not evaluate the competence of the PPs from the three protocols *in vivo*. Nonetheless, the PPs of protocols A and C showed significant similarities to the *in vivo* PPs derived from human fetal pancreas, suggesting that these protocols recapitulate several aspects of human pancreas development. Among the genes forming the *in vitro* PP signature we found several genes involved in pancreas development, including *ONECUT1* (Jacquemin et al., 2003) and *SOX9* (Seymour et al., 2007), as well as genes involved in Notch and Wnt signaling pathways, with previously well-established functions in pancreatic development (Pan and Wright, 2011). The open chromatin regions forming the epigenomic PP signature were significantly enriched in binding motifs of several known pancreatic TFs, including HNF6, HNF1B, FOXA2, PDX1, and SOX9. We observed that CUX2 motifs were the most significantly enriched in these common open chromatin regions. In line with this, we found that the *CUX2* gene's expression during differentiation peaks at the pancreatic endoderm stage, and we observed evidence for CUX2 being bound within the open chromatin of the samples generated in this study through ATAC-seq footprinting analysis. Interestingly, a previous study demonstrated the expression of *CUX2* in early fetal human dorsal pancreas (Jennings et al., 2017), and more recently the presence of CUX2 motifs in accessible chromatin regions were shown to be differentially enriched in mouse endocrine progenitors across different stages of development (Scavuzzo et al., 2018). Based on this evidence, we suggest CUX2 as a potential novel regulator of pancreatic development.

We then sought to highlight the protocol-specific differences in the omics profiles of the PPs. We focused on modules of co-expressed genes and correlated open chromatin grouping genes preferentially expressed in each of the protocols. We observed a significantly higher endocrine commitment in the NKX6.1-negative compartment of protocol A compared with protocols B and C. This is in agreement with the original reports, where Nostro et al. (2015) and Russ et al. (2015) developed protocols with the specific purpose of limiting the endocrine differentiation in place of PPs. Thus, we were able to recapitulate important features of the three differentiation protocols. Interestingly, the NKX6.1-negative population of protocol A clustered closely together with isolated human fetal endocrine progenitors (Ramond et al., 2018), suggesting that the endocrine differentiation observed *in vitro* shares many similarities with fetal endocrine development. On the other hand, genes with the highest expression in the GFP+ population from protocol A were enriched in gene signatures of the *in vivo* fetal PPs from the same study, as well as in the signatures of *in vitro* PPs from this and a previous study (Cebola et al., 2015).

Further, we noted that some of the genes preferentially expressed in cells generated with protocol B were suggestive of these cells assuming a non-endocrine fate. We noted enrichments in genes forming previously reported intestinal signatures, and we observed that open chromatin regions correlated with expression of these genes were enriched in binding motifs of CDX2, a marker of intestinal development, and *CDX2* gene expression was also most highly expressed in protocol B cells. A recent study

Figure 5. Reduction of Expression of the Intestinal Marker CDX2 in PPs

(A) TPM for *CDX2* in GFP− and GFP+ sorted cell populations from differentiation protocols A, B, and C. Graph shows scatterplot of mean with each dot representing individual differentiations.

(B) Flow cytometry analysis of NKX6.1 and CDX2 of PPs from the three differentiation protocols. DE was used as negative control. Representative pseudo color dot plots of five individual differentiations.

(C and D) Representative pseudo color dot plots of two individual differentiations. Ten modifications of protocol B were assessed for the ability to maintain PDX1 and NKX6.1 expression (C) while simultaneously reducing expression of CDX2 (D). Representative pseudo color dot plots of cells stained for NKX6.1 and PDX1 (C) or NKX6.1 and CDX2 (D). DE cells were used as negative controls.

(E) Heatmap summarizing the percentage of PDX1/NKX6.1 and CDX2/NKX6.1 co-expressing cells. Average percentage of one differentiation each of SB AD3.1 and SB AD3.4 hiPSC lines (n = 2 independent differentiations). Conditions tested were (1) 50 ng/mL Noggin, ST2; (2) 50 ng/mL Noggin, ST2-3; (3) 50 ng/mL Noggin, ST2-4; (4) ST4 only 2 days; (5) 100 nM LDN, ST2; (6) 100 nM LDN, ST2-3; (7) 100 nM LDN, ST2-4; (8) 50 ng/mL Activin A, ST4.2; (9) without KGF, ST4.2; (10) reduced retinoic concentration (0.2 μM) second day of ST3.

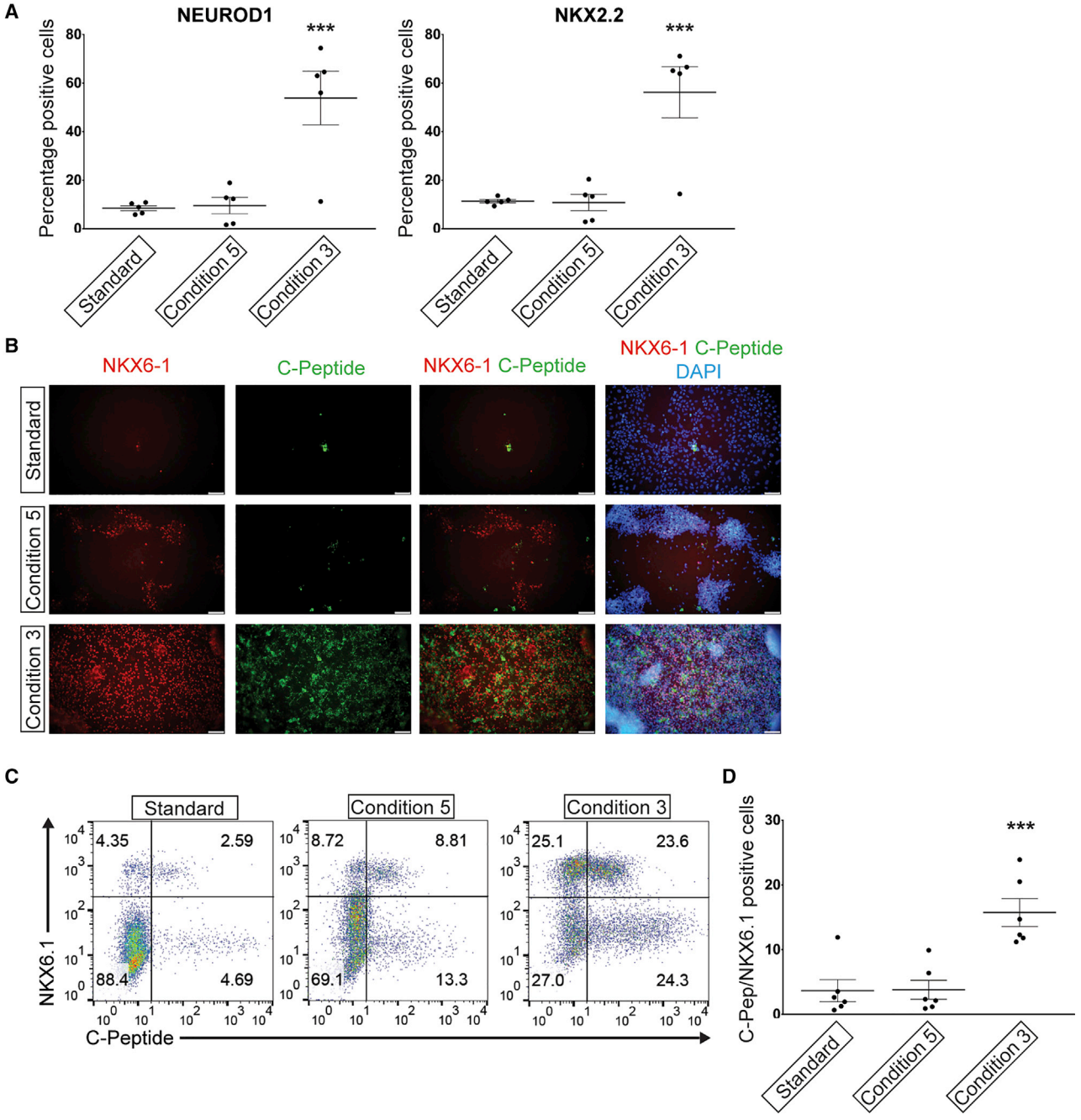


Figure 6. Improved Endocrine Differentiation Following Reduction of CDX2 Expression in PPs

(A) Flow cytometry-based quantifications of stage 5 endocrine progenitor differentiation of protocol B and the two modified conditions (condition 5 and 3). Scatterplots show percentage of NEUROD1 and NKX2.2-positive cells as means ± SEM, n = 5 independent experiments (three with SB AD3.1 hiPSC, one with SB AD3.4 hiPSC, and one with SA121 hESC).

(B) Immunofluorescence microscopy images of stage 6 cells stained with NKX6.1 and C-peptide antibodies. DAPI is used to visualize the nuclei of all cells. Scale bar, 100 μm.

(C) Representative pseudo color dot plots of stage 6 beta-like cells stained for C-peptide and NKX6.1. Numbers mark the percentage of cells in each quadrant.

(legend continued on next page)

reported the expression of *CDX2* in the dorsal pancreas of human fetal tissue (Jennings et al., 2017). However, we have previously performed single-cell gene expression analysis of human fetal pancreata and found no evidence of *CDX2* expression in human PPs (Ramond et al., 2018). This discrepancy could perhaps be explained by the different stages of human pancreas development interrogated in the two studies. Thus, it remains unclear whether PDX1, NKX6.1, and CDX2 co-expressing progenitor cells occur at some point during human pancreas development. CDX2-positive cells were also observed in protocol A but the majority of these were NKX6.1 negative. Excluding the BMP antagonists from protocol A resulted in a significant increase of CDX2-positive cells at the expense of NKX6.1-positive PPs, illustrating that inhibiting BMP signaling is necessary to prevent differentiation toward an intestinal fate. This observation is in agreement with a previous study demonstrating that BMP signaling inhibition during differentiation of hPSCs to PPs resulted in reduced *CDX2* expression (Shahjalal et al., 2014). Interestingly, addition of EGF to the final stage of protocol A resulted in a reduction of the percentage of CDX2-positive cells. These results illustrate how evaluating individual components of a differentiation protocol can elucidate their specific roles in the differentiation and guide improving the protocol efficiency.

We hypothesized that *CDX2* expression and the intestinal gene signature enriched in the PPs from protocol B could suggest a more posterior patterning during the differentiation, which might explain the inability of the PPs from protocol B to differentiate further toward the pancreatic endocrine lineage. In the present study, we found that inclusion of the BMP antagonist Noggin resulted in reduced expression of *CDX2* and other genes specific to the intestinal gene expression module, and subsequently in successful differentiation down the endocrine lineage. Interestingly, another BMP antagonist, LDN193189, also reduced CDX2 expression, but this was not accompanied by an improved differentiation toward the endocrine lineage. The reason for this difference is unclear but may relate to the different mechanisms of the two BMP antagonists. Together, these results illustrate the importance of monitoring CDX2 expression during differentiation toward PPs. Interestingly, the original report describing protocol B argued for the omission of BMP antagonist during pancreatic specification in order for efficient derivation of PDX1/NKX6.1 co-expressing PPs (Russ et al., 2015); however, CDX2 expression was not assessed in this study. It thus remains unclear whether CDX2 expression in the PPs is inherent to this protocol or a consequence of differences in experimental conditions and cell lines applied in this study compared with the original study (Russ et al., 2015). It is also plausible that, since the endocrine differentiation protocol applied in this study was originally developed in combination with protocol A (Rezania et al., 2014), it may not be directly applicable to other protocols. Whether there are differences in the signaling pathways that promote endocrine differentiation from the PPs derived from protocols A, B, and C remains to be shown.

There are many challenges associated with faithfully reproducing differentiation protocols, most recently demonstrated in a study comparing iPSC-derived neurons using a well-defined protocol across five different laboratories (Volpato et al., 2018). In addition, several modifications were made to the protocols applied in the current study compared with the original publications. Thus, it should be emphasized that our study should not be seen as a direct head-to-head comparison of the three protocols. Nonetheless, our work demonstrates the importance of assessing apparently similar cell populations derived using distinct differentiation protocols. This approach allowed us to define a comprehensive transcriptional and epigenomic signature for PPs that will serve as a valuable resource for studying pancreas development. However, our findings also highlighted significant variation in the PPs derived from the three protocols, which should warrant caution for using hPSCs to interrogate roles of specific genes in disease and development. Finally, we demonstrate the utility of benchmarking *in vitro*-derived cell populations to their *in vivo* counterparts, which allows for the identification of markers useful for improving the differentiation of hPSCs toward beta-like cells.

EXPERIMENTAL PROCEDURES

Please refer to Supplemental Experimental Procedures for detailed description of experimental procedures.

Maintenance and Differentiation of hPSC Lines

All hPSCs were maintained on human embryonic stem cell-qualified Matrigel in mTeSR1 medium, except for the NKX6.1-GFP hiPSC line, which was cultured in TeSR1-E8 medium, at 37°C, 5% CO_2, and passaged at 90%–95% confluence using TrypLE select, as previously described (Perez-Alcantara et al., 2018).

(D) Quantification of C-peptide/NKX6.1 double positive by flow cytometry as shown in (C) (conditions 5 and 3). Scatterplot shows percentage of C-peptide/NKX6.1 double-positive cells as means ± SEM, n = 6 independent experiments (four with SB AD3.1 hiPSC, two with SA121 hESC).

(A and D) One-way ANOVA with Tukey test for multiple comparisons, ***p < 0.001, different from the two other groups.

FACS

Briefly, differentiated cells were harvested to a single-cell suspension using TrypLE select, pelleted and resuspended in the stage 4 medium of the respective protocols (without factors) containing 1 μg/mL DAPI solution, and directly proceeded for sorting. Cells were sorted using a BD FACSAria Fusion (BD Biosciences).

RNA-seq

For RNA-seq, cells were pelleted immediately following sorting and medium was removed. Cells were harvested and RNA extracted using TRIzol Reagent (ThermoFisher Scientific, Paisley, UK) as per the manufacturer's guidelines. Smart-Seq2 paired-end RNA-seq libraries were sequenced on Illumina HiSeq4000 to a mean depth of 37.7 (± 1.8) million 75 bp reads pairs per sample.

Transposition Reaction and Purification for ATAC-Seq

Following sorting, cells were pelleted and washed in cold PBS. Cell pellets were gently resuspended in cold lysis buffer (10 mM Tris-HCL, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% IGEPAL CA-630) and immediately pelleted by centrifugation. The supernatant was discarded and the cell pellets were gently resuspended in 50 μL of transposition reaction mix (25 μL 2 of TD buffer, 2.5 μL of Tn5 Transposase [Illumina], 22.5 μL of nuclease-free H_2O) and incubated at 37°C for 30 min. Following the transposition reaction, the transposed DNA was purified using QIAGEN's MinElute Kit according to manufacturer's instructions and eluted in 10 μL of elution buffer (10 mM Tris buffer, pH 8).

Statistics

Statistical analyses were performed using GraphPad Prism (V8.0.2). Means were compared with unpaired Student's t test and one-way ANOVA with Tukey test for multiple comparisons.

ACCESSION NUMBERS

RNA-seq and ATAC-seq raw data have been deposited in the European Genome-phenome Archive, accession number EGAS00001003513. Dataset (RNA-seq): EGAS00001004823. Dataset (ATAC-seq): EGAS00001004824.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.stemcr.2019.11.010.

AUTHOR CONTRIBUTIONS

A.W.A. performed all bioinformatics analysis, interpreted the data, and wrote the manuscript. R.R.J. performed *in vitro* differentiation experiments, data analysis, and interpreted data. M.P.A. provided data for ATAC-seq analysis of control iPSC lines and performed bioinformatics analysis of these. N.L.B., C.D., V.N., M.G., L.W., and R.B. provided experimental and scientific input. M.I.M., M.H. and A.L.G. provided scientific input and edited the manuscript. C.H. performed *in vitro* differentiation experiments, data analysis, interpreted data, and wrote the manuscript. All authors approved the manuscript.

ACKNOWLEDGMENTS

A.L.G. is a Wellcome Trust Senior Fellow in Basic Biomedical Science. M.I.M. is a Wellcome Senior Investigator. This work was funded in Oxford by the Wellcome Trust (095101 [A.L.G.], 200837 [A.L.G.], 098381 [M.I.M.], 212259 [M.I.M.] 106130 [A.L.G., M.I.M.], 203141 (A.L.G., M.I.M.], 203141 [M.I.M.], Medical Research Council (MR/L020149/1) [M.I.M., A.L.G.], European Union Horizon 2020 Programme (T2D Systems) [A.L.G.], and NIH (U01-DK105535; U01-DK085545) [M.I.M., A.L.G.]. The research was funded by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC) [A.L.G., M.I.M.]. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health. Furthermore, the research leading to these results has received support from the Innovative Medicines Initiative (IMI) Joint Undertaking (JU) under grant agreement n° 115439, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and European Federation of Pharmaceutical Industries and Associations (EFPIA) companies' in kind contribution with additional specific support from Eli Lilly. This publication reflects only the authors' views and neither the IMI JU nor EFPIA nor the European Commission are liable for any use that may be made of the information contained therein. In addition, this work was in part supported by the Wellcome Trust via core funding to the Wellcome Centre for Human Genetics (awards 090532/Z/09/Z and 203141/Z/16/Z). M.H., R.R.J., and C.H. are employees of Novo Nordisk A/S and hold shares in Novo Nordisk A/S. N.L.B. is an employee of Novo Nordisk Ltd, but all experimental work was carried out while under the employment of University of Oxford. M.I.M. is a Wellcome Senior Investigator and an NIHR Senior Investigator. He serves on advisory panels for Pfizer, NovoNordisk, Zoe Global; has received honoraria from Merck, Pfizer, NovoNordisk, and Eli Lilly; has stock options in Zoe Global; and has received research funding from Abbvie, Astra Zeneca, Boehringer Ingelheim, Eli Lilly, Janssen, Merck, NovoNordisk, Pfizer, Roche, Sanofi Aventis, Servier, and Takeda. The views expressed in this article are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health.

Received: April 23, 2019
Revised: November 29, 2019
Accepted: November 29, 2019
Published: December 26, 2019

REFERENCES

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

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., et al. (2015). TEAD and YAP regulate the enhancer network of human embryonic pancreatic progenitors. Nat. Cell Biol. *17*, 615–626.

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., et al. (2014). 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*, 146–154.

Grainger, S., Savory, J.G., and Lohnes, D. (2010). Cdx2 regulates patterning of the intestinal epithelium. Dev. Biol. *339*, 155–165.

Gupta, S.K., Wesolowska-Andersen, A., Ringgaard, A.K., Jaiswal, H., Song, L., Hastoy, B., Ingvorsen, C., Taheri-Ghahfarokhi, A., Magnusson, B., Maresca, M., et al. (2018). NKX6.1 induced pluripotent stem cell reporter lines for isolation and analysis of functionally relevant neuronal and pancreas populations. Stem Cell Res. *29*, 220–231.

Jacquemin, P., Lemaigre, F.P., and Rousseau, G.G. (2003). The Onecut transcription factor HNF-6 (OC-1) is required for timely specification of the pancreas and acts upstream of Pdx-1 in the specification cascade. Dev. Biol. *258*, 105–116.

Jennings, R.E., Berry, A.A., Strutt, J.P., Gerrard, D.T., and Hanley, N.A. (2015). Human pancreas development. Development *142*, 3126–3137.

Jennings, R.E., Berry, A.A., Gerrard, D.T., Wearne, S.J., Strutt, J., Withey, S., Chhatriwala, M., Piper Hanley, K., Vallier, L., Bobola, N., et al. (2017). Laser capture and deep sequencing reveals the transcriptomic programmes regulating the onset of pancreas and liver differentiation in human embryos. Stem Cell Reports *9*, 1387–1394.

Kelly, O.G., Chan, M.Y., Martinson, L.A., Kadoya, K., Ostertag, T.M., Ross, K.G., Richardson, M., Carpenter, M.K., D'Amour, K.A., Kroon, E., et al. (2011). Cell-surface markers for the isolation of pancreatic cell types derived from human embryonic stem cells. Nat. Biotechnol. *29*, 750–756.

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

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.

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

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

Nair, G., and Hebrok, M. (2015). Islet formation in mice and men: lessons for the generation of functional insulin-producing beta-cells from human pluripotent stem cells. Curr. Opin. Genet. Dev. *32*, 171–180.

Nostro, M.C., Sarangi, F., Ogawa, S., Holtzinger, A., Corneo, B., Li, X., Micallef, S.J., Park, I.H., Basford, C., Wheeler, M.B., et al. (2011). Stage-specific signaling through TGFbeta family members and WNT regulates patterning and pancreatic specification of human pluripotent stem cells. Development *138*, 861–871.

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

Osafune, K., Caron, L., Borowiak, M., Martinez, R.J., Fitz-Gerald, C.S., Sato, Y., Cowan, C.A., Chien, K.R., and Melton, D.A. (2008). Marked differences in differentiation propensity among human embryonic stem cell lines. Nat. Biotechnol. *26*, 313–315.

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

Perez-Alcantara, M., Honore, C., Wesolowska-Andersen, A., Gloyn, A.L., McCarthy, M.I., Hansson, M., Beer, N.L., and van de Bunt, M. (2018). Patterns of differential gene expression in a cellular model of human islet development, and relationship to type 2 diabetes predisposition. Diabetologia *61*, 1614–1622.

Petersen, M.B.K., Azad, A., Ingvorsen, C., Hess, K., Hansson, M., Grapin-Botton, A., and Honore, C. (2017). Single-cell gene expression analysis of a human ESC model of pancreatic endocrine development reveals different paths to beta-cell differentiation. Stem Cell Reports *9*, 1246–1261.

Petersen, M.B.K., Goncalves, C.A.C., Kim, Y.H., and Grapin-Botton, A. (2018). Recapitulating and deciphering human pancreas development from human pluripotent stem cells in a dish. Curr. Top. Dev. Biol. *129*, 143–190.

Ramond, C., Beydag-Tasoz, B.S., Azad, A., van de Bunt, M., Petersen, M.B.K., Beer, N.L., Glaser, N., Berthault, C., Gloyn, A.L., Hansson, M., et al. (2018). Understanding human fetal pancreas development using subpopulation sorting, RNA sequencing and single-cell profiling. Development *145*. https://doi.org/10.1242/dev.165480.

Rezania, A., Bruin, J.E., Riedel, M.J., Mojibian, M., Asadi, A., Xu, J., Gauvin, R., Narayan, K., Karanu, F., O'Neil, J.J., et al. (2012). Maturation of human embryonic stem cell-derived pancreatic progenitors into functional islets capable of treating pre-existing diabetes in mice. Diabetes *61*, 2016–2029.

Rezania, A., Bruin, J.E., Xu, J., Narayan, K., Fox, J.K., O'Neil, J.J., and Kieffer, T.J. (2013). Enrichment of human embryonic stem cell-derived NKX6.1-expressing pancreatic progenitor cells accelerates the maturation of insulin-secreting cells in vivo. Stem Cells *31*, 2432–2442.

Rezania, A., Bruin, J.E., Arora, P., Rubin, A., Batushansky, I., Asadi, A., O'Dwyer, S., Quiskamp, N., Mojibian, M., Albrecht, T., et al. (2014). Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. Nat. Biotechnol. *32*, 1121–1133.

Rouhani, F., Kumasaka, N., de Brito, M.C., Bradley, A., Vallier, L., and Gaffney, D. (2014). Genetic background drives transcriptional variation in human induced pluripotent stem cells. PLoS. Genet. *10*, e1004432.

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., et al. (2015). Controlled induction of human pancreatic progenitors produces functional beta-like cells in vitro. EMBO J. *34*, 1759–1772.

Scavuzzo, M.A., Hill, M.C., Chmielowiec, J., Yang, D., Teaw, J., Sheng, K., Kong, Y., Bettini, M., Zong, C., Martin, J.F., et al. (2018). Endocrine lineage biases arise in temporally distinct endocrine progenitors during pancreatic morphogenesis. Nat. Commun. *9*, 3356.

Seymour, P.A., Freude, K.K., Tran, M.N., Mayes, E.E., Jensen, J., Kist, R., Scherer, G., and Sander, M. (2007). SOX9 is required for maintenance of the pancreatic progenitor cell pool. Proc. Natl. Acad. Sci. U S A *104*, 1865–1870.

Shahjalal, H.M., Shiraki, N., Sakano, D., Kikawa, K., Ogaki, S., Baba, H., Kume, K., and Kume, S. (2014). Generation of insulin-producing beta-like cells from human iPS cells in a defined and completely xeno-free culture system. J. Mol. Cell Biol. *6*, 394–408.

Volpato, V., Smith, J., Sandor, C., Ried, J.S., Baud, A., Handel, A., Newey, S.E., Wessely, F., Attar, M., Whiteley, E., et al. (2018). Reproducibility of molecular phenotypes after long-term differentiation to human iPSC-derived neurons: a multi-site omics study. Stem Cell Reports *11*, 897–911.

Wandzioch, E., and Zaret, K.S. (2009). Dynamic signaling network for the specification of embryonic pancreas and liver progenitors. Science *324*, 1707–1710.

Yang, Y., and Chan, L. (2016). Monogenic diabetes: what it teaches us on the common forms of type 1 and type 2 diabetes. Endocr. Rev. *37*, 190–222.

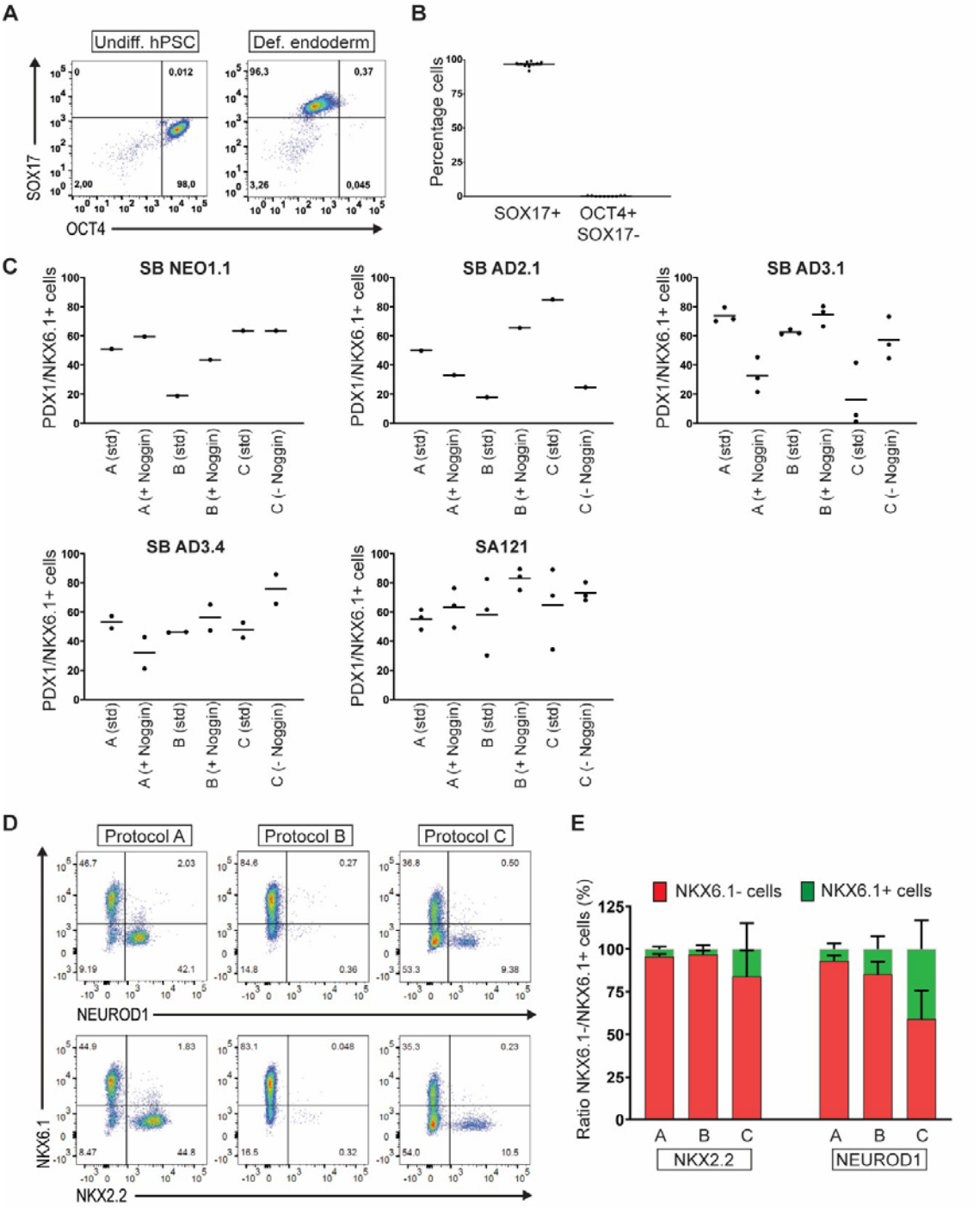
Stem Cell Reports, Volume *14*

Supplemental Information

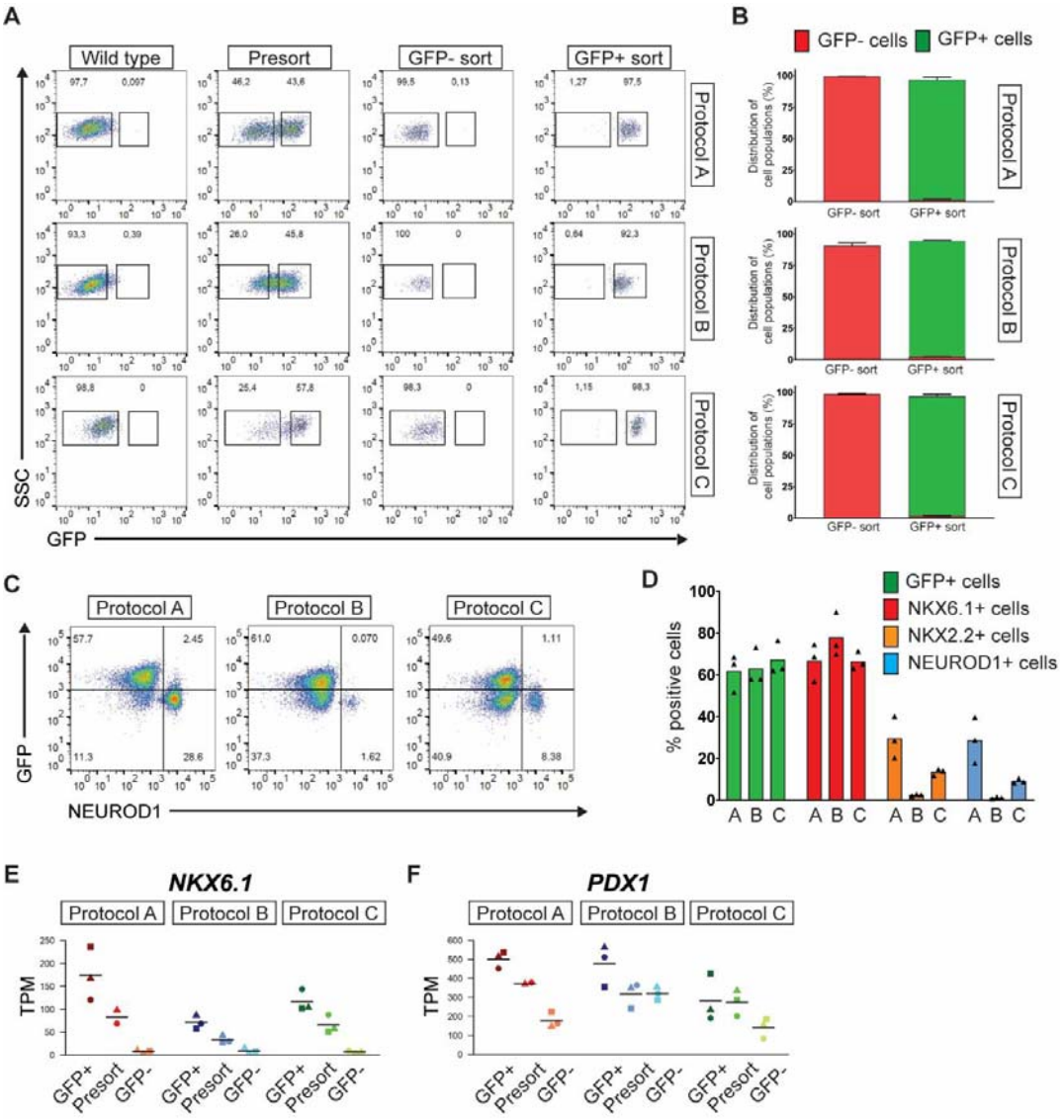
Analysis of Differentiation Protocols Defines a Common Pancreatic Progenitor Molecular Signature and Guides Refinement of Endocrine Differentiation

Agata Wesolowska-Andersen, Rikke Rejnholdt Jensen, Marta Pérez Alcántara, Nicola L. Beer, Claire Duff, Vibe Nylander, Matthew Gosden, Lorna Witty, Rory Bowden, Mark I. McCarthy, Mattias Hansson, Anna L. Gloyn, and Christian Honore

SUPPLEMENTAL FIGURES

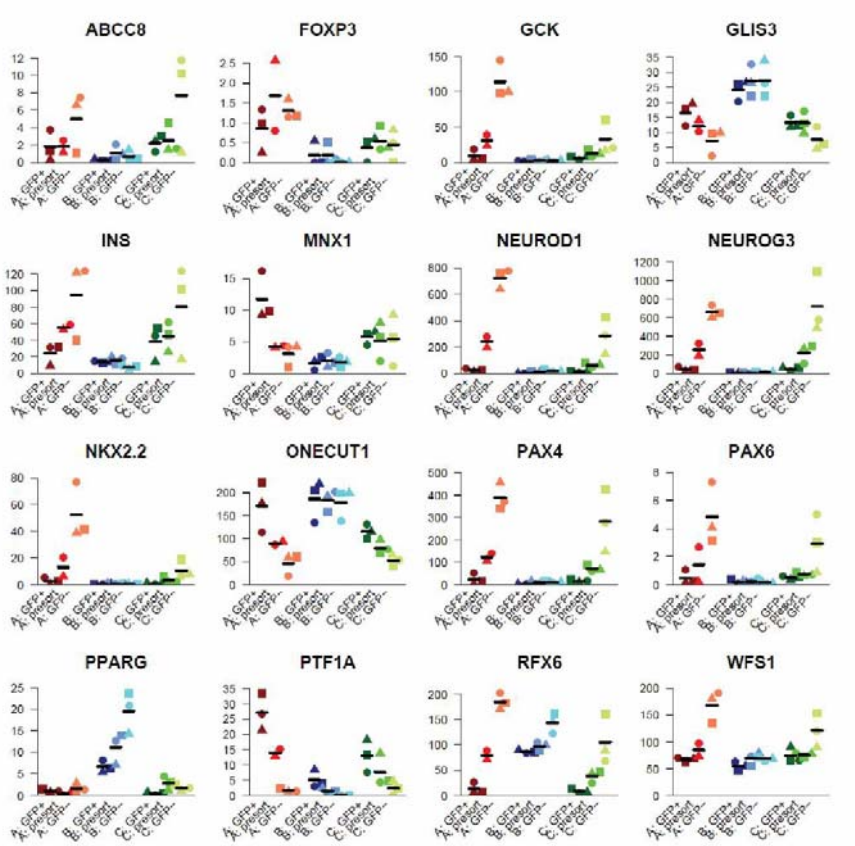


Supplementary figure 1. Definitive endoderm and pancreatic progenitor differentiation efficiency. Related to figure 1.
(A) Representative example of flow cytometry pseudo color dot plots of undifferentiated hPSC and definitive endoderm cells stained for OCT4 and SOX17. (B) Percentage of SOX17+ cells and OCT4+, SOX17- cells at the definitive endoderm stage of the protocol. Data presented as mean ± SEM, n = 10 independent experiments (SA121 hESC n = 3, SB NEO1.1 hiPSC n = 1, SB AD2.1 hiPSC n = 1, SB AD3.1 hiPSC n = 3, SB AD3.4 hiPSC n = 2). (C) Quantification of PDX1 and NKX6.1 co-expressing cells across five individual hPSC lines differentiated with protocol A, B and C. The effect of inclusion/omission of Noggin during stage 2 of each protocol was evaluated side-by-side compared to the standard protocols (outlined in figure 1A). Graphs show scatter plot of mean, n = 1-3 independent experiments, depending on cell line. (D) representative example of flow cytometry pseudo color dot plots of hPSC-derived pancreatic progenitors from protocol A, B and C. Cells are analysed for NKX6.1/NEUROD1 (top) and NKX6,1/NKX2.2 expression (bottom). (E) Ratio of NKX2.2 and NEUROD1 expressing cells negative (red) or positive (green) for NKX6.1 expression. Data presented as mean ± SEM, n = 5 independent experiments.

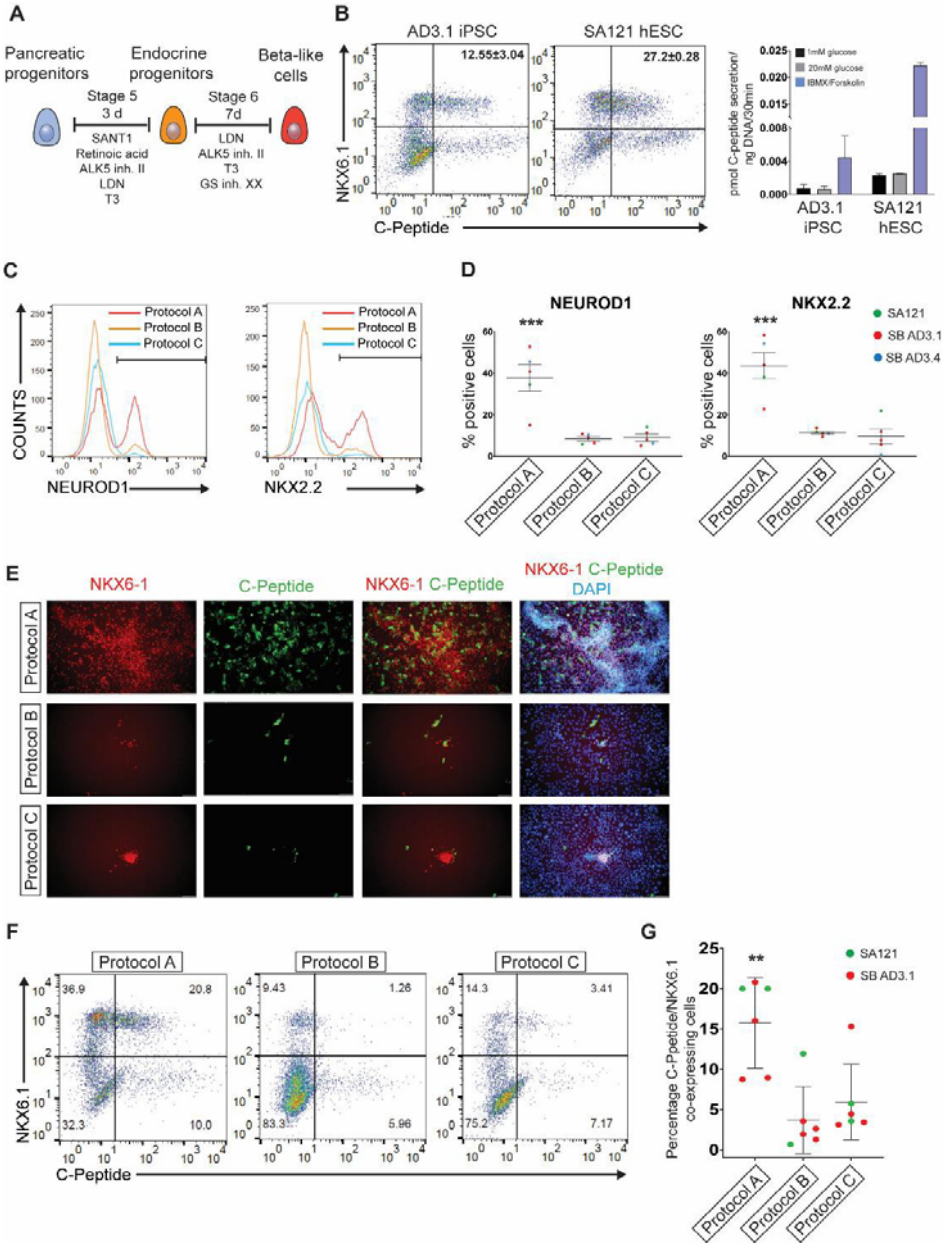


Supplementary figure 2. Confirmation of isolation of NKX6.1+ and NKX6.1- cell populations by FACS. Related to figure 2.

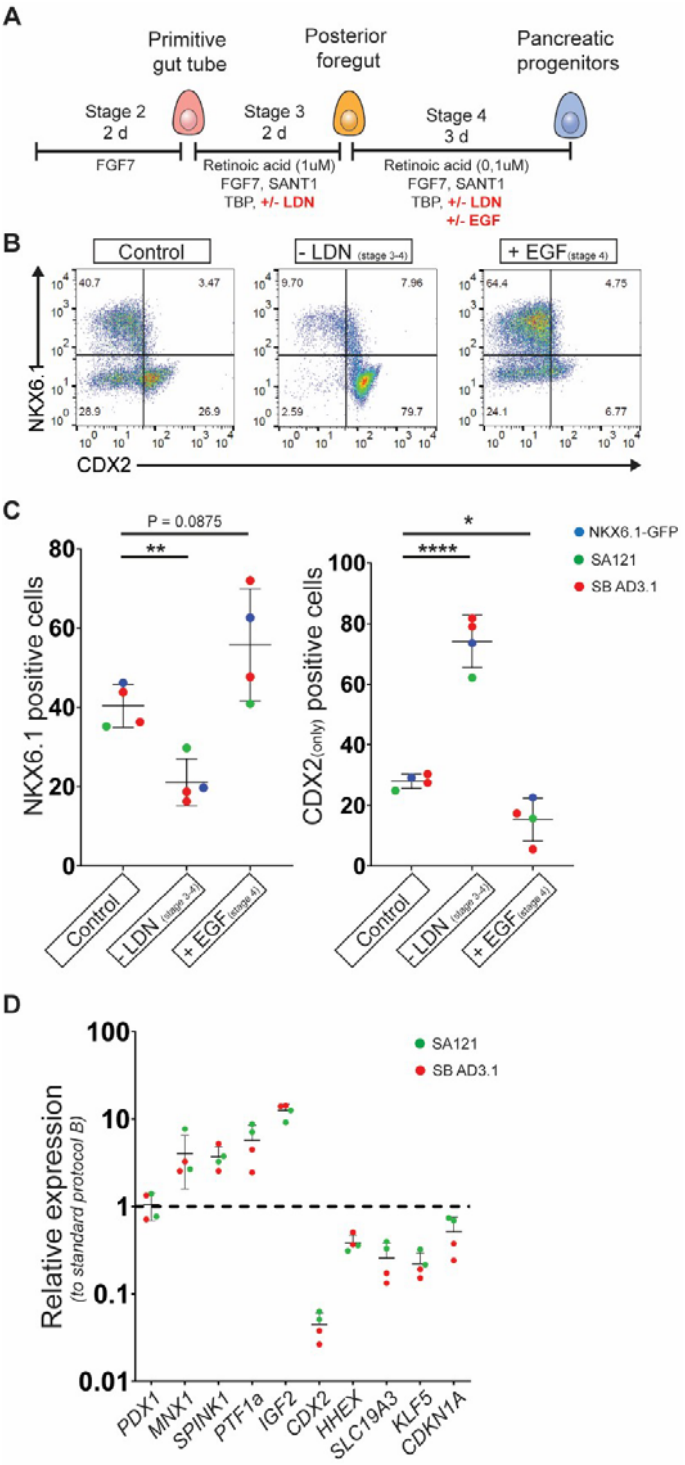
(A) Representative example of flow cytometry pseudo color dot plots of wild type hiPSC line (wild type), presorted NKX6.1-GFP iPSC (presort) and the GFP- and GFP+ sorted fractions from all three protocols analyzed for purity post sorting. Boxes indicates gates labelling GFP+ and GFP- cell fractions (B) Graph showing the percentage of GFP- and GFP+ cells in each of the two sorted fractions across all three differentiation protocols. Graphs shows mean ± SEM of three independent differentiations/sorting experiments (n = 3). (C) Representative examples of pseudo color dot plots of NKX6.1-GFP hiPSC line differentiated to pancreatic progenitors using the three protocols. Cells are analysed for GFP and NEUROD1 expression and illustrates that the endocrine cells (NEUROD1+ cells) are mainly found in the GFP- (NKX6.1-) compartment. (D) Flow cytometry quantification of GFP+, NKX6.1+, NKX2.2+ and NEUROD1+ cells from the three experiments used for the cell sorting experiments. Bars show mean and dots represent individual differentiations (n = 3). GFP+ expression was assessed on live cells, whereas NKX6.1, NKX2.2 and NEUROD1 expression analysis was performed on fixed and permeabilized cells. (E-F) Transcript per kilobase million (TPM) for *NKX6.1* and *PDX1* in presorted, GFP- and GFP+ sorted cell populations across all three differentiation protocols. Graphs shows scatter plot and mean, with each symbol representing an independent differentiation. *NKX6.1*: p_{adjA}=4.05e-101, p_{adjB}=1.60e-27, p_{adjC}=5.97e-64; *PDX1*: p_{adjA} =2.70e-19, p_{adjB} =8.42e-03, p_{adjC} =1.20e-07.



Supplementary figure 3. Expression of selected genes implicated in MODY and neonatal diabetes. Related to figure 4. Expression is shown as transcript per kilobase million (TPM) in pre-sorted, GFP- and GFP+ sorted cell populations across all three differentiation protocols. Graphs shows scatter plot and mean, with each symbol representing an independent differentiation.


A
Pancreatic progenitors
Endocrine progenitors
Beta-like cells
Stage 5
3 d
SANT1
Retinoic acid
ALK5 inh. II
LDN
T3
Stage 6
7d
LDN
ALK5 inh. II
T3
GS inh. XX
B
AD3.1 iPSC
SA121 hESC
12.55±3.04
27.2±0.28
NKX6.1
C-Peptide
1mM glucose
20mM glucose
IBMX/Forskolin
pmol C-peptide secretion/ ng DNA/30min
AD3.1 iPSC
SA121 hESC
C
Protocol A
Protocol B
Protocol C
COUNTS
NEUROD1
NKX2.2
D
NEUROD1
NKX2.2
***
% positive cells
SA121
SB AD3.1
SB AD3.4
Protocol A
Protocol B
Protocol C
E
NKX6-1
C-Peptide
NKX6-1 C-Peptide
NKX6-1 C-Peptide DAPI
Protocol A
Protocol B
Protocol C
F
Protocol A
Protocol B
Protocol C
36.9
20.8
32.3
10.0
9.43
1.26
83.3
5.96
14.3
3.41
75.2
7.17
NKX6.1
C-Peptide
G
**
Percentage C-Ppeticle/NKX6.1 co-expressing cells
SA121
SB AD3.1
Protocol A
Protocol B
Protocol C

Supplementary figure 4. Pancreatic endocrine differentiation of pancreatic progenitors derived from multiple protocols. **Related to figure 5**.
(A) Schematic outline of the protocol for differentiating the pancreatic progenitors from the three protocols towards first endocrine progenitors (stage 5) and subsequently towards beta-like cells (stage 6). (B) NKX6.1/C-Peptide flow cytometry pseudo color dot plots of AD3.1 iPSC and SA121 hESC lines differentiated to stage 6 for subsequent glucose stimulated insulin secretion (GSIS). Numbers in upper right quadrant shows percentage of NKX6.1+, C-Peptide+ cells (mean ± SD, n =2 for each line). Graph shows C-Peptide secretion normalized to DNA content of cells (mean ± SD, n =2 for each line). (C) Representative flow cytometry histograms demonstrating the quantification of NEUROD1 (left) and NKX2.2 (right) positive cells. Gate set based on a negative cell population and unstained cells. Color of the lines mark cells from each of the three protocols. Horizontal bar in both plots shows the gate for positive cells (D) Quantifications of NEUROD1 and NKX2.2 positive cells as shown in (C). Graphs shows percentage of NEUROD1 and NKX2.2 positive cells as mean ± SEM, n = 5 independent experiments (E) Immunofluorescence microscopy images of stage 6 cells derived from pancreatic progenitors from protocol A, B and C. Cells were stained with NKX6.1 and C-Peptide antibodies. DAPI is used to visualize the nuclei of all cells. Scalebar = 100μM. (F) Representative flow cytometry pseudo color dot plots of stage 6 cells derived from each of the three protocols. (G) Quantification of C-Peptide, NKX6.1 co-expressing cells as mean ± SEM, n = 6 independent experiments. For (D) and (G) Dots are color coded according to individual cell lines as indicated in legend of the figure. One-way ANOVA with Tukey test for multiple comparisons, **P < 0.01, ***P < 0.001, different from the two other groups.


A
Primitive gut tube
Posterior foregut
Pancreatic progenitors
Stage 2 2 d
Stage 3 2 d
Stage 4 3 d
FGF7
Retinoic acid (1uM) FGF7, SANT1 TBP, +/- LDN
Retinoic acid (0,1uM) FGF7, SANT1 TBP, +/- LDN +/- EGF
B
Control
- LDN (stage 3-4)
+ EGF (stage 4)
NKX6.1
CDX2
40.7
3.47
28.9
26.9
9.70
7.96
2.59
79.7
64.4
4.75
24.1
6.77
C
P = 0.0875
**
*
****
NKX6.1 positive cells
CDX2(only) positive cells
NKX6.1-GFP
SA121
SB AD3.1
Control
- LDN (stage 3-4)
+ EGF (stage 4)
D
Relative expression
(to standard protocol B)
SA121
SB AD3.1
PDX1
MNX1
SPINK1
PTF1a
IGF2
CDX2
HHEX
SLC19A3
KLF5
CDKN1A

Supplementary figure 5. Evaluation of BMP and EGF signalling in differentiation protocol A and B. Related to figure 5.
(A) Schematic outline of stage 2-4 of protocol A. Red text highlights compounds evaluated. (B) Representative examples of pseudo color dot plots of cells at the end of stage 4 differentiated with protocol A and modifications thereof. (C) Quantification of NKX6.1 positive (left) and $CDX2_{(only)}$ positive cells (right). Graphs shows percentage of positive cells as mean ± SEM, n = 4 independent experiments. Dots are color coded according to individual cell lines as indicated in legend of the figure. Unpaired t test applied to determine differences between individual groups. *P < 0.05, **P < 0.01, ****P < 0.0001. (D) Quantitative gene expression analysis of pancreatic progenitors derived using protocol B or a modified version of protocol B (Condition 3 in figure 5). Expression was normalized to the average expression of two housekeeping genes (*ACTB* and *HPRT1*). Figure shows expression of genes in the modified version of protocol B relative to the standard protocol B (dashed line). Mean ± SEM, n = 4.

RESOURCE TABLE
Reagents, antibodies, assays, cell lines and software applied.

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
NKX6.1 Alexa 647 (FC 1:40)	BD Biosciences	Cat#563338; RRID: N/A
NKX6.1 PE (FC 1:40)	BD Biosciences	Cat#563023; RRID:AB_2716792
PDX1 Alexa 488 (FC 1:40)	BD Biosciences	Cat#562274; RRID:AB_10924596
NeuroD1 PE (FC 1:40)	BD Biosciences	Cat#563001; RRID:AB_2716791
NKX2.2 PE (FC 1:40)	BD Biosciences	Cat#564730; RRID: N/A
Ki-67 Alexa 488 (FC 1:20)	BD Biosciences	Cat#558616; RRID:AB_647087
Sox17 Alexa 488 (FC 1:40)	BD Biosciences	Cat#562205; RRID:AB_10893402
Oct3/4 Alexa 647 (FC 1:10)	BD Biosciences	Cat#560307; RRID:AB_1645319
CDX2 (FC 1:500)	ThermoFisher	Cat#RM-2116; RRID: N/A; clone EPR2764Y
Rabbit IgG (isotype control)	R&D systems	Cat#AB-105-C; RRID:AB_354266
PDX1 (IF 1:8000)	Abcam	Cat#AB47383; RRID:AB_2162359
NKX6.1 (IF 1:200)	DSHB	Cat#F55A12-S; RRID:AB_532379
C-Peptide/Pro-Insulin (IF 1:250)	DSHB	Cat#GN-ID4; RRID:AB_2255626
Donkey Anti-Rabbit Alexa488 (FC 1:500)	ThermoFisher	Cat#A21206; RRID:AB_2535792
Donkey Anti-Goat Alexa488 (IF 1:500)	ThermoFisher	
Donkey Anti-Mouse Alexa594 (IF 1:500)	ThermoFisher	
Donkey Anti-Rat Alexa488 (IF 1:500)	ThermoFisher	
DAPI solution (FC)	BD Biosciences	Cat#564907
LIVE/DEAD™ Fixable Violet Dead Cell Stain Kit	ThermoFisher	Cat#L34955
NucleoSpin RNA/Protein kit	Macherey-Nagel	Cat#740933.50
iScript cDNA Synthesis Kit	Bio-rad	Cat#170-8890
TagMan Gene Expression Master Mix	ThermoFisher Scientific	Cat##PN 4369016
ACTB TaqMan Assay	ThermoFisher Scientific	hs01060665_g1
HPRT1 TaqMan Assay	ThermoFisher Scientific	hs99999909_m1
PDX1 TaqMan Assay	ThermoFisher Scientific	hs00236830_m1
MNX1 TaqMan Assay	ThermoFisher Scientific	hs00907365_m1
PTF1A TaqMan Assay	ThermoFisher Scientific	hs00603586_g1
IGF2 TaqMan Assay	ThermoFisher Scientific	hs04188276_m1
CDX2 TagMan Assay	ThermoFisher Scientific	Hs01078080_m1
HHEX TaqMan Assay	ThermoFisher Scientific	hs00242160_m1
SPINK1 TaqMan Assay	ThermoFisher Scientific	hs00162154_m1
SLC19A3 TaqMan Assay	ThermoFisher Scientific	hs00228858_m1

KLF5 TaqMan Assay	ThermoFisher Scientific	hs00156145_m1
CDKN1A TaqMan Assay	ThermoFisher Scientific	hs00355782_m1
Bacterial and Virus Strains		
Biological Samples		
Chemicals, Peptides, and Recombinant Proteins		
hESC-qualified matrigel	BD biosciences	Cat#354277
Growth factor reduced matrigel	BD biosciences	Cat#356230
mTeSR1 medium	Stemcell technologies	Cat#05850
TeSR1-E8 medium	Stemcell technologies	Cat#05990
MCDB131 medium	ThermoFisher	Cat# 10372019
DMEM/F12 medium	ThermoFisher	Cat# 21331020
DMEM(HG) medium	ThermoFisher	Cat# 31966021
PBS-/- (without calcium and magnesium)	ThermoFisher	Cat#14190094
PBS+/+ (with calcium and magnesium)	ThermoFisher	Cat# 14040133
TrypLE select	ThermoFisher	Cat#12563011
4% Formaldehyde	VWR	Cat#9713.1000; CAS 50-00-0
Triton X-100	Sigma-Aldrich	Cat#T9284; CAS 9002-93-1
Donkey serum	Millipore	Cat#S30-100ML
$NaHCO_3$	ThermoFisher	Cat#25080060
Penicillin-Streptomycin solution	ThermoFisher	Cat#15140122
Glutamax	ThermoFisher	Cat#35050038
Glucose	Fisher scientific	Cat#D16500
BSA, fatty acid free	Proliant	Cat#68700
BSA Fraction V, fatty acid free	Sigma-Aldrich	Cat#10775835001
ITS-X	ThermoFisher	Cat#51500056
Zinc sulphate	Sigma-Aldrich	Cat# Z0251; CAS 7446-20-0
B-27 supplement (50x)	ThermoFisher	Cat#17504044
B-27 supplement (50x), minus vitamin A	ThermoFisher	Cat#12587010
Rock inhibitor	Sigma-Aldrich	Cat#Y0503; CAS 129830-38-2
Activin A	Peprotech	Cat#120-14
CHIR99021	Axon Medchem	Cat#1386; CAS 252917-06-9
L-Ascorbic acid	Sigma-Aldrich	Cat#A4544; CAS 50-81-7
Recombinant human KGF (FGF-7)	Peprotech	Cat#100-19
Retinoic acid	Sigma-Aldrich	Cat#R2625; CAS 302-79-4
SANT-1	Sigma-Aldrich	Cat#S4572; CAS 304909-07-7
LDN-193189	Stemgent	Cat#04-0074; CAS 1062368-24-4
TBP (PKC activator)	Merck Millipore	Cat#565740; CAS 497259-23-1
Heparin	Sigma-Aldrich	Cat#H3149; CAS 9041-08-1
3,3′,5-Triiodo-L-thyronine sodium salt (T3)	Sigma-Aldrich	CAT# T6397; CAS 55-06-1

ALK5 Inhibitor II	Enzo life sciences	Cat# ALX-270-445; CAS 446859-33-2
γ-Secretase Inhibitor XX	Merck Millipore	Cat#565789; CAS 209984-56-5
Recombinant Human EGF	R&D systems	Cat#236-EG
Recombinant Human Noggin	Peprotech	Cat#120-10C
Nicotinamide	Sigma-Aldrich	Cat#N0636; CAS 98-92-0
Critical Commercial Assays		
Deposited Data		
EGA RNA-seq		Study ID: EGAS00001003513 Data set: EGAS00001004823
EGA ATAC-seq		Study ID: EGAS00001003513 Data set: EGAS00001004824
Experimental Models: Cell Lines		
SB NEO1 Cl. 1	https://www.stembancc.org/ https://www.ebisc.org/	DOI: 10.1007/s00125-018-4612-4
SB AD2.1 hiPSC	https://www.stembancc.org/ https://www.ebisc.org/	DOI: 10.1007/s00125-018-4612-4 DOI: 10.1080/19382014.2016.1182276
SB AD3.1 hiPSC	https://www.stembancc.org/ https://www.ebisc.org/	DOI: 10.1007/s00125-018-4612-4
SB AD3.4 hiPSC	https://www.stembancc.org/ https://www.ebisc.org/	DOI: 10.1080/19382014.2016.1182276
SA121 hESC	Takara Bio	Cat# Y00020
NKX6.1-GFP hiPSC (clone 1a-21-3)	https://www.stembancc.org/ https://www.ebisc.org/	DOI: 10.1016/j.scr.2018.04.010
Experimental Models: Organisms/Strains		
Oligonucleotides		
Recombinant DNA		
Software and Algorithms		
FASTQC		https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

STAR (v.2.5.1)	DOI: 10.1093/bioinformatics/bts635	https://github.com/alexdobin/STAR
Picard tools suite (v2.1.1) - MarkDuplicates		https://broadinstitute.github.io/picard/
featureCounts	DOI: 10.1093/bioinformatics/btt656	http://bioinf.wehi.edu.au/featureCounts/
Kundaje lab ATAC-seq pipeline		https://github.com/kundajelab/atac-seq-pipeline
Bowtie2 (v.2.2.6)	DOI: 10.1038/nmeth.1923.	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Bedtools (v.2.26)	DOI:10.1093/bioinformatics/btq033	https://bedtools.readthedocs.io/en/latest/
MACSv2 (2.1.0)	DOI: 10.1186/gb-2008-9-9-r137	https://github.com/taoliu/MACS
DESeq2 (v.1.32.0)	DOI: 10.1186/s13059-014-0550-8	https://www.bioconductor.org/packages/release/bioc/html/DESeq2.html
Sva (v.3.28.0)	DOI: 10.1093/bioinformatics/bts034	http://bioconductor.org/packages/release/bioc/html/sva.html
edgeR (v.3.22.2)	DOI: 10.1093/bioinformatics/btp616	http://bioconductor.org/packages/release/bioc/html/edgeR.html
RUVseq (v.1.14.0)	DOI: 10.1038/nbt.2931	https://bioconductor.org/packages/release/bioc/html/RUVSeq.html
DAVID	DOI: 10.1186/gb-2007-8-9-r183	https://david.ncifcrf.gov/home.jsp
HOMER	DOI: 10.1016/j.molcel.2010.05.004	http://homer.ucsd.edu/homer/
msCentipede	DOI: 10.1371/journal.pone.0138030	https://github.com/rajanil/msCentipede
JASPAR Core	DOI: 10.1093/nar/gkx1126	http://jaspar.genereg.net/collection/core/
FIMO	DOI: 10.1093/bioinformatics/btr064	http://meme-suite.org/tools/fimo
atactk toolkit (v.0.1.6)		https://github.com/ParkerLab/atactk

ggplot2	https://cran.r-project.org/web/packages/ggplot2/citation.html	https://cran.r-project.org/web/packages/ggplot2/index.html
WGCNA (v.1.64.1)	DOI: 10.1186/1471-2105-9-559	https://cran.r-project.org/web/packages/WGCNA/index.html
AmiGO 2	DOI: 10.1093/bioinformatics/btn615	http://amigo.geneontology.org/amigo
Other		

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Maintenance and differentiation of hPSC lines

The following hPSC lines were applied in this study: SA121 hESC, SB NEO1.1 hiPSC, SB AD2.1 hiPSC, SB AD3.1 hiPSC, SB AD3.4 hiPSC and NKX6.1-GFP hiPSC line. Details on source of lines can be found in the resource table in the supplemental information. All hPSC were maintained on hESC-qualified matrigel in mTeSR1 with the exception of the NKX6.1-GFP hiPSC line, which was cultured in TeSR1-E8 medium at 37°C, 5% CO^2 as previously described (Perez-Alcantara et al., 2018). Briefly, cells were passaged every 3-4 days or when reaching 90-95% confluence. Medium was aspirated and the cells were washed once with PBS without calcium and magnesium (PBS-/-) and subsequently incubated with TrypLE select (3-5min, 37°C) to achieve a single cell solution. TrypLe select wash quenched with 3-4 volumes of medium with 5µM Rock inhibitor and cells were pelleted (300xG, 5min). Cell number was determined using a NucleoCounter NC-200 (Chemotec, Denmark) and cells were resuspended to the desired concentration in medium with 5µM Rock inhibitor. Cells were seeded in medium with 5uM Rock inhibitor onto tissue culture flasks (Nunclon delta surfaces, Thermo Scientific, #156367 or #156499) pre-coated with hESC-qualified matrigel for maintenance or onto 12-well tissue culture plates (CellBIND surfaces, Corning #3336) for differentiation. For differentiation, cells were seeded at 300.000-400.000 cells/cm^2 and 24h after seeding the cells, medium was aspirated and cells were washed once with PBS with calcium and magnesium (PBS+/+) before exposing the cells to differentiation medium. Stage 1 (definitive endoderm) was shared across all three protocols. Cells were incubated for 1 day with MCDB131 medium containing 0.5% BSA (w/v), 0.1% (v/v) Penicillin-Streptomycin, 1% (v/v) Glutamax, 10mM glucose, 100ng/ml Activin A and 3uM CHIR99021. For day 2, cells were cultured with MCDB131 medium containing 0.5% BSA (w/v), 0.1% (v/v) Penicillin-Streptomycin, 1% (v/v) Glutamax, 10mM glucose, 100ng/ml Activin A and 0.3uM CHIR99021. For day 3 cells were cultured with MCDB131 medium containing 0.5% BSA (w/v), 1.5 g/l $NaHCO_3$, 0.1% (v/v) Penicillin-Streptomycin (PS), 1% (v/v) Glutamax, 10mM glucose and 100ng/ml Activin A. Following induction of definitive endoderm, cells were differentiated with three pancreatic progenitor protocols outlined below. Importantly, differentiations with the three protocols were always performed side-by-side. Medium was prepared freshly for each stage of the differentiation protocol and cells were replenished with medium daily across all protocols and stages. Small molecules and growth factors were aliquoted to appropriate volumes to reduce the number of freeze-thaw cycles (≤3).

PROTOCOL A adapted from (Rezania et al., 2014)			
Stage	**Medium**	**Compounds**	**Key differences compared to original protocol**
Stage 2 (2 days)	MCDB131, 0.1% PS, 1% glutamax, 1.5g/l $NaHCO_3$ 10mM glucose, 0.5% BSA, 0.25mM ascorbic acid	50ng/ml KGF (FGF7)	• Differences in maintenance of undifferentiated hPSC compared to original study

Stage 3 (2 days)	MCDB131, 0.1% PS, 1% glutamax, 2.5g/l $NaHCO_3$ 10mM glucose, 2% BSA, 0.25mM ascorbic acid , 1:200 (v/v) ITS-X	50ng/ml KGF (FGF7), 1µM Retinoic acid, 0.25µM SANT-1, 100nM LDN-193189, 200nM TBP	• CHIR99021 and Activin A used for DE induction compared to MCX-928 and GDF8 in the original publication. • MCDB131 medium used for stage 3-4 compared to custom made medium (BLAR) reported in the original publication. • Differences in vendors used for various supplements and factors.
Stage 4 (3 days)		2ng/ml KGF (FGF7), 0.1µM Retinoic acid, 0.25µM SANT-1, 200nM LDN-193189, 100nM TBP	

PROTOCOL B adapted from (Russ et al., 2015)			
Stage	**Medium**	**Compounds**	**Key differences compared to original protocol**
Stage 2 (2 days)	MCDB131, 0.1% PS, 1% glutamax, 1.5g/l $NaHCO_3$ 10mM glucose, 0.5% BSA, 0.25mM ascorbic acid	50ng/ml KGF (FGF7)	• Differences in maintenance of undifferentiated hPSC compared to original study • Cells differentiated in 2D format compared to suspension culture in the original publication. • Different protocol for DE induction compared to original publication • Differences in medium and FGF7 concentration in stage 2. • Different retinoic acid receptor agonist applied • Differences in vendors used for various supplements and factors. • Differences in length of stage 4.2 compared to original publication
Stage 3 (2 days)	DMEM(HG) 0.1% PS, 1% B-27 supplement	2µM Retinoic acid	
Stage 4.1 (1 day)		2µM Retinoic acid, 50ng/ml EGF	
Stage 4.2 (2 days)		50ng/ml KGF (FGF7), 50ng/ml EGF	

PROTOCOL C adapted from (Nostro et al., 2015)			
Stage	**Medium**	**Compounds**	**Key differences compared to original protocol**

Stage 2 (3 days)	DMEM/F12, 0.1% PS, 1% Glutamax, 1% N2 supplement, 1% B27 supplement (- Vitamin A), 0.05% BSA, 0.5mM Ascorbic acid	50ng/ml KGF (FGF7), 50ng/ml Noggin* (**cell line dependent*)	• Differences in maintenance of undifferentiated hPSC compared to original study • Different protocol for DE induction compared to original publication • No Wnt3a in stage 2 of protocol • FGF7 used instead of FGF10 throughout • SANT-1 used compared to KAAD-Cyclopamine in original publication • Differences in vendors used for various supplements and factors. • Activin A during stage 4 of protocol.
Stage 3 (1 day)	DMEM(HG) 0.1% PS, 1% B-27 supplement (- Vitamin A), 0.5mM Ascorbic acid	2μM Retinoic acid, 50ng/ml KGF (FGF7), 50ng/ml Noggin	
Stage 4 (3 days)		100ng/ml EGF, 50ng/ml Noggin, 10mM Nicotinamide, 50ng/ml Activin A	

*Omission of Noggin from stage 2 of protocol C resulted in a significantly more efficient differentiation of the SB AD3.1 hiPSC line. The NKX6.1-GFP reporter hiPSC line was derived from the SB AD3.1 hiPSC line and thus, all differentiations of the NKX 6.1-GFP reporter hiPSC line with protocol C was carried out without Noggin during stage 2.

For subsequent endocrine differentiation of the pancreatic progenitors derived from all three protocols, stage 5 and stage 6 of the differentiation protocol outlined below was applied.

ENDOCRINE DIFFERENTIATION adapted from (Rezania et al., 2014)		
Stage	**Medium**	**Compounds**
Stage 5 (3 days)	MCDB131, 0.1% PS, 1% glutamax, 1.5g/l $NaHCO_3$ 20mM glucose, 2% BSA, 0.5% ITS-X, 10uM Zinc sulfate	0.05μM Retinoic acid, 0.25μM SANT-1, 100nM LDN-193189, 10μM ALK5i II, 1μM T3, 10μg/ml heparin
Stage6 (7 days)		100nM LDN-193189, 10μM ALK5i II, 1μM T3, 10μg/ml heparin, 100nM γ-secretase inhibitor XX

Immunofluorescence imaging
Cells were washed once in PBS and fixed directly in the wells of tissue culture plates by incubating with 4% formaldehyde for 20min at room temperature. The fixed cells were subsequently rinsed twice in PBS and then permeabilized for 10min at room temperature in PBS containing 0.5% Triton X-100. Following permeabilization, cells were incubated for 30 min at room temperature with a blocking solution consisting of 0.1 M Tris-HCl pH 7.4, 0.15 M NaCl, and 0.5% Tyramide Signal Amplification (TSA) immunohistochemistry kit blocking reagent. Cells were subsequently incubated with primary antibodies diluted in PBS with 0.1% Triton X-100 overnight at 4°C. Following incubation with primary antibodies, cells were washed three times in PBS and subsequently incubated with secondary antibodies and 4',6-diamidino-2-phenylindole in PBS with 0.1% Triton X-100 for 45min at room temperature. Images were acquired with an inverted fluorescence microscope (Olympus IX-81).

Flow cytometry analysis

Cells were harvested to a single cell suspension using TrypLE select, rinsed once in PBS and stained with Live/Dead cell kit for 10min on ice. Cells were subsequently washed and fixated for 20min at 4°C in 4% formaldehyde. Cells were washed in PBS and incubated for 30min in perm-buffer (PBS +/+ with 0.2% Triton X-100 and 5% Donkey serum) at 4°C. Following permeabilization, cells were pelleted (800 x G, 5min) and resuspended in block-buffer (PBS +/+ with 0.1% Triton X-100 and 5% Donkey serum) containing antibodies. Cells were incubated for 30min at room temperature covered from light and subsequently washed twice in PBS with 1% BSA. Cells were analysed on BD LSR Fortessa cell analyser (BD Biosciences) with 10.000-20.000 events recorded pr. sample.

Fluorescence activated cell sorting
An NKX6.1-T2A-GFP reporter iPSC lines was differentiated side-by-side to the pancreatic progenitor stage using the three differentiation protocols. A wild type iPSC line was also differentiated and used to define the gates for sorting the GFP- and GFP+ fractions. A total of three independent differentiations were performed and from each differentiation both GFP+ and GFP- cells as well as unsorted controls were FACS purified. Briefly, cells were harvested to a single cell suspension using TrypLE select, pelleted and resuspended in the stage 4 medium of the respective protocols (without factors) containing 1µg/ml DAPI solution and directly proceeded for sorting. Cells were sorted using a BD FACSAria Fusion (BD Biosciences) with the 100µM nozzle/20psi configuration. GFP+ and GFP- cells were sorted into Eppendorf tubes containing medium at 4°C. Between 50.000-100.000 cells were sorted for RNA sequencing and 45.000-60.000 cells were sorted for ATAC sequencing.
Real time quantitative PCR (RT-qPCR). RNA was isolated from harvested cells using NucleoSpin RNA/Protein kit. 500ng RNA was converted to cDNA using iScript cDNA Synthesis Kit and adjusted to a final volume of 200ul following reverse transcription. Gene expression was analysed using TagMan Gene Expression Assays (see resource table below). RT-qPCR reactions consisted of 0,5ul Tagman Gene expression assay, 2ul cDNA, 5ul TagMan Gene Expression Master Mix and 2,5ul H_2O. RT-qPCR was performed on a MX3005P qPCR system (Agilent Genomics) using a fast-2-step protocol (first 95°C for 1min, then 40 cycles of 95°C for 10sec, 60°C for 25sec). Expression values were normalized to the average expression of two housekeeping genes (ACTB and HPRT1) and relative expression to unsorted samples was calculated using the ΔΔCt method.

RNA sequencing and processing of sequence data
For RNA sequencing, cells were pelleted immediately following sorting and medium was removed. Cells were harvested and RNA extracted using TRIzol Reagent (ThermoFisher Scientific, Paisley, UK) as per the manufacturer's guidelines. Library preparation and sequencing was performed at the Oxford Genomics Centre (Wellcome Centre for Human Genetics, Oxford, UK) as previously described(van de Bunt et al., 2016). We prepared the Smart-Seq2 paired-end RNA-seq libraries for the 27 samples (3 differentiations x 3 protocols x 3 cell populations: presort, GFP+ and GFP-), and sequenced on Illumina HiSeq4000 to a mean depth of X (+/- X) million 75bp reads pairs per sample. The sequencing quality was assessed with FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and the raw sequencing reads were mapped to the human genome hg19 using STAR version 2.5.1 with default settings. The GENCODE v19 GTF was applied to guide the spliced alignment. Duplicated alignments were marked with the MarkDuplicates script from the Picard tools v2.1.1 suite (http://broadinstitute.github.io/picard). Gene expression was quantified with featureCounts using the GENCODE v19 GTF file. We assessed the library complexities for all the samples by comparing the number of genes detected at >1TPM with the percentage of total reads mapping within the top 100 expressed genes. We discovered that one of the presort samples from protocol A from differentiation 3 had much lower library complexity (X) than the remaining samples (X), therefore we removed it from further analyses.

Transposition reaction and Purification/ATAC-seq
ATAC-seq sample preparation was done as described previously(Buenrostro et al., 2013). Described shortly, following sorting cells were pelleted and washed in cold PBS. Cell pellets were gently resuspended in cold lysis buffer (10mM Tris-HCL, pH 7.4, 10mM NaCl, 3mM MgCl2, 0.1% IGEPAL CA-630) and immediately pelleted by centrifugation. The supernatant was discarded and the cell pellets were gently resuspended in 50µl transposition reaction mix (25µl 2 x TD buffer, 2.5µl Tn5 Transposase (Illumina), 22.5µl Nuclease free H_2O) and incubated at 37°C for 30min. Following the transposition reaction, the transposased DNA was purified using Qiagens MinElute Kit (Qiagen) according to manufacturer's instructions and eluted in 10µl elution buffer (10mM Tris buffer, pH 8). Eluted DNA was PCR amplified using Nextera PCR primers (Illumina) and NEBNext high fidelity PCR master mix (New England Biolabs) with the following cycling 1) 72°C, 5 min, 2) 98°C, 30 sec , 3) 98°C, 10 sec, 4) 63°C, 30 sec, 5) 72°C, 1 min, Repeating steps 3-5, 9-11x. PCR product was purified using Qiagens MinElute Kit (Qiagen) according to manufacturer's instructions and eluted in 20µl elution buffer (10mM Tris buffer, pH 8). Quality of final

library was assessed using Tapestation (agilent). Libraries were sequenced at the High-Throughput Genomics core at the Wellcome Centre for Human Genetics, University of Oxford, UK. Samples were sequenced as multiplexed pools on HiSeq4000 (Illumina) using 75-bp paired-end sequencing.
The ATAC-seq libraries for the 27 samples (3 differentiations x 3 protocols x 3 cell populations: presort, GFP+ and GFP-) were sequenced on Illumina HiSeq4000 to a mean depth of X (+/- X) million 75bp reads pairs per sample. The raw sequencing reads were processed with the Kundaje lab ATAC-seq pipeline (https://github.com/kundajelab/atac-seq-pipeline). The pipeline detects and trims the sequencing adaptors, maps the reads to the human reference genome build hg19 using Bowtie2 version 2.2.6. The duplicated alignments are marked with MarkDuplicates script from the Picard tolos suite, converts the BAM files to tagAlign format using bedtools versión 2.26 and calls the signal peaks using MACSv2 (2.1.0).

Principal component analysis
The global similarities between the samples for both the RNA-seq and ATAC-seq datasets were assessed through principal component analysis (PCA). The gene count tables were transformed with variance stabilizing transformation (vst) in the DESeq2 R package (v.1.32.0)(Love et al., 2014). The batch effects of the two multiplexing pools introduced for sequencing in the ATAC-seq dataset were corrected using the sva R package (v.3.28.0)(Leek et al., 2012). The PCA analysis was performed using the plotMDS function from the edgeR R package (v.3.22.2)(Robinson et al., 2010), using the setting gene.selection="common", and focusing on the top 500 most variable genes. For the PCA analysis of RNA-seq samples from this study together with the RNA-seq dataset from a study of human fetal pancreas populations (Ramond 2018), presented in Fig.3D, we merged the raw gene count tables for the two studies, performed the variance stabilizing transformation in DESeq2 on the merged count table, followed by sva batch correction for the effect of the study of origin.

Defining the common PP signature
To derive common transcriptomic and epigenomic signatures of pancreatic progenitors generated in this study with three differentiation protocols, we compared the data to the equivalent omics profiles of all stages of hPSC differentiation towards beta-line cells (Perez Alcantara 2018). To facilitate the comparison we applied the RUVs normalization implemented in the RUVseq (v.1.14.0) R package(Risso et al., 2014). We first applied the upper-quartile normalization to account for differences in library sizes, followed by the RUVs approach to estimate and remove the factors of unwanted variation. For that purpose we treated the presorted samples generated with protocol A as replicates of the control lines samples from the pancreatic endoderm stage (also generated with protocol A). After the RUVs normalization, we generated the normalized count tables for all the samples and used them to calculate CPM (counts-per-million) expression values using the cpm() function from the edgeR R package. We then calculated mean CPM values for each sample group (GFP cell population from each protocol, and all differentiation stages). We then defined pancreatic progenitor signatures within the GFP+ cell populations for each protocol separately; and then we used the intersection of the three individual protocol-specific signatures to define the common cross-protocol pancreatic progenitor signature. To define the sets of genes forming a pancreatic progenitor signature for each protocol, we required that the genes were expressed at >1 TPM in the GFP+ population from the given protocol, demonstrated stage selectivity (CV>1), as well as specificity to the pancreatic endoderm (Z-score >1) when evaluated together with the remaining stages. We calculated similar PP signatures in the presort, and GFP- populations as well. The PP signatures were derived in the same manner for both the RNA-seq and ATAC-seq datasets. Functional enrichments for the PP signature genes were calculated with DAVID Gene Functional Classification tool(Huang et al., 2007); raw enrichment p-values for selected Gene Ontology categories were then plotted as -log10(p-value) barplots in Fig.3C. Enrichment of known transcription factor binding motifs was calculated with HOMER(Heinz et al., 2010) using the default settings.

Footprinting analysis
msCentipede(Raj et al., 2015) was used to call sites bound by transcription factors. First, transcription factor binding motifs were called genome-wide using position weight matrices from JASPAR Core (non-redundant) 2018 (Khan et al., 2018) using FIMO (version 4.11.2, (Grant et al., 2011)) with default parameters. The model parameters were learned per experiment and GFP status using genome-wide called motifs and replicate bam files. Transcription factor binding motifs were called in each replicate peak file (for high quality peaks with q-value < 10^{-6}; per experiment and GFP status), and subsequently the replicates were merged and used as input for the inference step in msCentipede. Sites with the following scores were considered bound: LogPosOdds > 2, MultLikeRatio>1 and NegBinLikeRatio>1; and the mean counts per site were calculated using make_cut_site function in the atactk toolkit (v. 0.1.6)(https://github.com/ParkerLab/atactk) and plotted using ggplot2 (Wickham, 2016) in R (v. 3.4.3).

DEGs/DOCS

Testing for differential expressed genes (DEGs) and differentially open chromatin sites (DOCS) was performed with DESeq2 R package (v…). We compared data from each differentiation protocol to the data generated with the two other protocols, for each cell population (presort, GFP+ and GFP-) separately. We also compared the data between the GFP+ and GFP- cell populations for each protocol separately. In both RNA-seq and ATAC-seq analyses we adjusted for the effects of the differentiations, and in the ATAC-seq additionally also for the effects of the sequencing multiplex pool. We considered all genes and open chromatic sites with FDR-adjusted p-values<0.05 as
significantly differentially expressed.

WGCNA

We performed weighted gene co-expression analysis (WGCNA)(Langfelder and Horvath, 2008) on both the vst-transformed RNA-seq and ATAC-seq datasets. The RNA-seq analysis was performed on 15,555 protein-coding genes expressed at >=1 TPM in at least 3 samples. The ATAC-seq analysis was limited to 19,739 open chromatin peaks, with >=10 sequencing reads in at least 3 samples, and a coefficient of variation >0.5. Using the criteria of best fit to the scale free topology model we selected power of 6 for soft thresholding of RNA-seq data, and power of 3 for the ATAC-seq dataset, and required a minimum module size of 50 for both analyses. We identified 20 co-expressed gene modules in the RNA-seq dataset, and 11 modules of correlated open chromatin sites. The dimensionality reduction of these large datasets is achieved through assignment of a single eigengene value for each modules in each sample. The eigengene value represents the first principal component value for the given module genes or open chromatin regions. The module eigengene values for the samples with both RNA-seq and ATAC-seq data can then be correlated to derive pairs of highly correlated modules. These are likely to correspond to open chromatin at regulatory elements controlling the gene expression of the correlated genes.

Gene signatures and Gene Ontology functional enrichment testing

For each of the 20 co-expressed gene modules we tested enrichment in selected gene sets using the hypergeometric distribution. We investigated enrichments of the *in vitro* pancreatic progenitor signatures derived in this study, as
described in an earlier section, as well as the 500 gene PP signature derived through comparison of *in vitro* and *in vivo* multipotent pancreatic progenitors from a previous study by Cebola *et al*. 2015. We then tested whether any of the gene modules were enriched in selected developmental Gene Ontology terms: GO:0031016 pancreas development, GO:0001889 liver development , GO:0048565 digestive tract development and GO:0060575 intestinal
epithelial cell differentiation. Genes belonging to these Gene Ontology categories were identified using AmiGO 2 Ontology search function(Carbon et al., 2009), accessed on 10th April 2018. We tested whether any of the gene modules might represent cells with likely intestinal fate through enrichment of previously reported genes differentially and highly expressed in intestinal stem cells derived from duodenum, jejunum and ileum(Wang et al., 2015). We also performed differential gene expression to identify significantly up-regulated biomarkers of dorsal pancreas, hepatobiliary primordium and hepatic cords from laser capture dissection in human embryos(Jennings et
al., 2017). We then tested enrichment for these sets of markers with log2FC>0 and FDR q-value<0.05 in the co-expressed gene modules using the hypergeometric distribution. Enrichment of known transcription factor binding motifs within co-open chromatin modules was calculated with HOMER using the default settings(Heinz et al., 2010).

SUPPLEMENTAL REFERENCES

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.

Carbon, S., Ireland, A., Mungall, C. J., Shu, S., Marshall, B., Lewis, S., Ami, G. O. H. & Web Presence Working, G. (2009). AmiGO: online access to ontology and annotation data. Bioinformatics, 25**,** 288-9.

Grant, C. E., Bailey, T. L. & Noble, W. S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics, 27**,** 1017-8.

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.

Huang, D. W., Sherman, B. T., Tan, Q., Collins, J. R., Alvord, W. G., Roayaei, J., Stephens, R., Baseler, M.

W., Lane, H. C. & Lempicki, R. A. (2007). The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol, 8, R183.
Jennings, R. E., Berry, A. A., Gerrard, D. T., Wearne, S. J., Strutt, J., Withey, S., Chhatriwala, M., Piper Hanley, K., Vallier, L., Bobola, N., et al. (2017). Laser Capture and Deep Sequencing Reveals the Transcriptomic Programmes Regulating the Onset of Pancreas and Liver Differentiation in Human Embryos. Stem Cell Reports, 9, 1387-1394.
Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J. A., van der Lee, R., Bessy, A., Cheneby, J., Kulkarni, S. R., Tan, G., et al. (2018). JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res, 46, D260-D266.
Langfelder, P. & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics, 9, 559.
Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics, 28, 882-3.
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.
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.
Perez-Alcantara, M., Honore, C., Wesolowska-Andersen, A., Gloyn, A. L., McCarthy, M. I., Hansson, M., Beer, N. L. & van de Bunt, M. (2018). Patterns of differential gene expression in a cellular model of human islet development, and relationship to type 2 diabetes predisposition. Diabetologia, 61, 1614-1622.
Raj, A., Shim, H., Gilad, Y., Pritchard, J. K. & Stephens, M. (2015). msCentipede: Modeling Heterogeneity across Genomic Sites and Replicates Improves Accuracy in the Inference of Transcription Factor Binding. PLoS One, 10, e0138030.
Rezania, A., Bruin, J. E., Arora, P., Rubin, A., Batushansky, I., Asadi, A., O'Dwyer, S., Quiskamp, N., Mojibian, M., Albrecht, T., et al. (2014). Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. Nat. Biotechnol, 32, 1121-1133.
Risso, D., Ngai, J., Speed, T. P. & Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. Nat Biotechnol, 32, 896-902.
Robinson, M. D., McCarthy, D. J. & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, 26, 139-40.
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., et al. (2015). Controlled induction of human pancreatic progenitors produces functional beta-like cells in vitro. EMBO J, 34, 1759-72.
van de Bunt, M., Lako, M., Barrett, A., Gloyn, A. L., Hansson, M., McCarthy, M. I., Beer, N. L. & Honore, C. (2016). Insights into islet development and biology through characterization of a human iPSC-derived endocrine pancreas model. Islets, 8, 83-95.
Wang, X., Yamamoto, Y., Wilson, L. H., Zhang, T., Howitt, B. E., Farrow, M. A., Kern, F., Ning, G., Hong, Y., Khor, C. C., et al. (2015). Cloning and variation of ground state intestinal stem cells. Nature, 522, 173-8.
Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.