The role of *RREB1* in human pancreatic beta cell development and function

Katia Kornelia Mattis

Green Templeton College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2019

Memorandum

The data and their interpretation presented in this thesis reflect my own work, unless explicitly stated otherwise.

Experiments were performed at the Oxford Centre for Diabetes, Endocrinology & Metabolism (OCDEM) and the Wellcome Trust Centre for Human Genetics (WCHG) under the supervision of Prof Anna L Gloyn and Dr Ben Davies. Funding of the project was provided by the Wellcome Trust and the Medical Research Council.

Karyotyping of genome-engineered hiPSC lines was done by Dr Daniela Moralli at the Chromosome Dynamics Core (WCHG) (Chapters 3 and 5).

Sequencing data including RNA-Seq and ChIP-Seq libraries (Chapters 3, 4 and 5) were generated by the Oxford Genomics Centre at the WCHG. RNA-Seq data analysis was supported by Marta Pérez Alcántara (WCHG) and Dr Agata Wesolowska-Andersen (WCHG). ChIP-Seq analysis was performed by Dr Toryn Poolman (OCDEM) (Chapter 4). Enrichment analysis of RREB1 binding sites in islet regulatory regions was done by Dr Jason Torres (WCHG) (Chapter 4).

Generation of the *RREB1* KO EndoC-βH1 cell model was performed together with Antje Grotz (OCDEM) and Dr Fernando Abaitua (WCHG) (Chapter 4).

Characterisation of *RREB1* KO EndoC-βH1 cells using electron microscopy (EM) (Chapter 4) was performed by Dr Benoit Hastoy and Dr Anne Clark. Confocal microscopy was done with the assistance of Dr Anne Clark (OCDEM).

Functional islet data for insulin secretion studies in carriers heterozygous for the T2D protective rs9379084 variant (Chapter 5) were generated by the Alberta Diabetes Institute IsletCore, Edmonton and obtained from Prof Patrick MacDonald.

Artwork printed at the beginning of each chapter was created by Felicity Cormack (Oxford Printmakers Co-operative).

This thesis has only been submitted to the University of Oxford for the degree of Doctor of Philosophy in Medical Sciences and not for any other degree at the same or any other university.

This thesis consists of approximately 50,000 words.

Acknowledgements

First, I would like to thank Anna Gloyn and Ben Davies for the opportunity to work on such an exciting project in their labs, their trust in my abilities and ideas, as well as their continuous support, advice, encouragement and guidance. It was a privilege to have them as supervisors.

I would also like to express my gratitude to all members of the Gloyn and Davies groups as well as everybody else in OCDEM and WCHG, who have contributed to a stimulating research environment and provided support in and outside the lab throughout my DPhil.

In particular, I would like to thank Fernando Abaitua for the lively discussions, his patience and open ears in every respect. Thanks to Antje Grotz for sharing the joy of working with EndoC-βH1 cells. I have also valued the discussions with Benoit Hastoy and Anne Clark about beta cell physiology. Thanks to Marta Pérez Alcántara, Agata Wesolowska-Andersen and Jason Torres for helping me with the bioinformatic analyses. I would also like to express my gratitude to Toryn Poolman for sharing his ChIP-Seq expertise and his continuous support with the analysis. Furthermore, I would like to thank my fellow students for sticking with me through all the ups and downs during the course of my DPhil.

I am grateful for having been a student at Green Templeton College, which has created a supportive environment for both my academic and social life.

I would also like to thank Felicity Cormack for the wonderful artwork she has created as part of a collaboration between the Oxford Printmakers Co-operative and the WCHG, representing an artistic interpretation of my research project.

Finally, I would like to express my gratitude to my family and friends, especially to Matthew McGilvray, for their never ending support and believe in me. Thanks for accompanying me during this journey.

Abstract

Human genetics can be used as a tool to gain insights into fundamental aspects of human development, physiology and pathophysiology. Genome-wide association studies have uncovered multiple independent signals which alter risk for type 2 diabetes (T2D) and influence related glycaemic traits at the *RREB1* locus. Fine-mapping revealed that one of these signals is driven by a common coding variant rs9379084 (p.D1171N), highlighting *RREB1* as the causal transcript. Little is known about the function of the zinc finger transcription factor (TF) RREB1 in glucose homeostasis and how changes in its expression and/or activity might alter diabetes risk. The aim of my thesis was to determine the role of RREB1 in human beta cell development and function through transcriptomic and cellular phenotyping of genome-edited human induced pluripotent stem cells (hiPSCs) and EndoC-βH1 beta cells.

To investigate the effect of loss of RREB1 on islet cell development I generated *RREB1* knockout (KO) hiPSC lines using CRISPR/Cas9 and differentiated them along the pancreatic endocrine lineage. Gene expression profiling revealed that loss of RREB1 had a positive impact on the generation of endocrine precursors (EPs). Endocrine progenitor markers (*NEUROG3*, *NKX2.2*, *NEUROD1*) were significantly up-regulated, while pancreas progenitor marker genes (*CPA2*, *NOTCH1*) were markedly down-regulated in hiPSC-derived *RREB1* KO EPs.

RREB1-depleted EndoC-βH1 cells, generated using either RNA interference-mediated knockdown or CRISPR/Cas9-mediated KO of *RREB1*, were characterised by reduced *INS* expression and insulin content, whilst, gene expression profiling pointed to a positive effect of loss of RREB1 on the differentiation state of mature beta cells. Significant up-regulation of genes implicated in beta cell function, connectivity and maturity (*GCK*, *SNAP25*, *GJD2*, *UCN3*) was accompanied by simultaneous down-regulation of beta cell disallowed genes (*PDGFRA*, *IGFBP4*). RREB1 ChIP-Seq analysis in EndoC-βH1 cells identified a subset of these as direct RREB1 target genes and revealed an enrichment of RREB1 binding sites in islet active promoters, highlighting RREB1 as a novel transcriptional activator and repressor in endocrine cells. RREB1 *cis*-regulated genes were enriched for target

genes of the beta cell forbidden TF REST, pointing to a potential cooperative action of these two TFs in endocrine gene regulation.

Transcriptional activities of RFX2 and RFX3 were significantly up-regulated in developing and mature *RREB1* KO beta cells. While *RFX2* was identified as a direct RREB1 target gene showing markedly increased transcript levels in RREB1-deficient beta cells, the mechanism underlying up-regulation of RFX3 motif activity remained unclear.

To explore the impact of the *RREB1* T2D-associated alleles on beta cell development, I generated allele-specific hiPSC lines using CRISPR/Cas9. Characterisation of differentiated hiPSCs homozygous for the *RREB1* T2D-associated coding variant p.N1171 suggested that the T2D protective allele acted as a gain-of-function allele, negatively affecting beta cell differentiation. How this is compatible with a protective effect on T2D risk requires further investigation.

Overall, characterisation of two complimentary *RREB1* KO models showed a novel role for RREB1 in beta cell development and function, contributing to the growing list of studies aiming to facilitate biological insights from T2D-associated genes.

Contents

Appendices

List of Figures

List of Tables

List of Abbreviations

aa . . . . . amino acid

ANOVA . . . . two-way analysis of variance

ATAC . . . . . assay for transposase-accessible chromatin

BLC . . . . . beta-like cells

BMI . . . . . body mass index

BSA . . . . . bovine serum albumin

cAMP . . . . . 3', 5'-cyclic adenosine monophosphate

ChIP . . . . . chromatin immunoprecipitation

CIT . . . . . cold ischaemia time

CNV . . . . . copy number variation

CRISPR . . . clustered regularly interspaced short palindromic repeats

CTCF . . . . . CCCTC-binding factor

DE . . . . . definite endoderm

DEG . . . . . differentially expressed gene

DHS . . . . . DNAse hypersensitive site

DSB . . . . . double-strand break

EM . . . . . electron microscopy

EMSA . . . . . electromobility shift assay

EN . . . . . endocrine-like cells

EP . . . . . endocrine progenitors

ER . . . . . endoplasmic reticulum

FACS . . . . . fluorescence-activated cell sorting

FDR . . . . . false discovery rate

FG . . . . . fasting glucose

GOF . . . . . gain-of-function

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

hPSC . . . . . human pluripotent stem cells

HKG . . . . . housekeeping gene

indel . . . . . insertion/deletion

IP . . . . . immunoprecipitation

$\mathbf{K_{ATP\ channel}}$ . . ATP-sensitive potassium channel

KD . . . . . knockdown

kDA . . . . . kilodalton

KGF . . . . . keratinocyte growth factor

KO . . . . . knockout

LD . . . . . linkage disequilibrium

lincRNA . . . long intergenic non-coding RNA

LOF . . . . . loss-of-function

MAF . . . . . minor allele frequency

ME . . . . . module eigengene

MODY . . . . maturity onset diabetes of the young

NDM . . . . . neonatal diabetes mellitus

NES . . . . . normalised enrichment score

NHEJ . . . . . non-homologous end joining

OR . . . . . odds ratio

PBS(T) . . . . phosphate buffered saline (tween)

PCA . . . . . principal component analysis

PCR . . . . . polymerase chain reaction

PFG . . . . . posterior foregut

PGT . . . . . primitive gut tube

PIC . . . . . proteinase inhibitor cocktail

PP . . . . . pancreatic polypeptide

P/S . . . . . penicillin-streptomyocin

PWM . . . . . position weight matrix

QC . . . . . quality control

RFX . . . . . regulatory factor x

RRE . . . . . Ras-responsive element

RREB1 . . . . Ras-responsive element binding protein 1

RT . . . . . room temperature

RUV . . . . . removal of unwanted variation

SD . . . . . standard deviation

SDM . . . . . site-directed mutagenesis

seq . . . . . sequencing

sgRNA . . . . single guide RNA

siRNA . . . . . small interfering RNA

SNP . . . . . single nucleotide polymorphism

ssODN . . . . single-stranded oligodeoxynucleotides

TF . . . . . transcription factor

TFBS . . . . . transcription factor binding site

TGF-β . . . . . transforming growth factor beta

TIDE . . . . . Tracking of Indels by DEcomposition

TPM . . . . . transcripts per million

T2D . . . . . type 2 diabetes

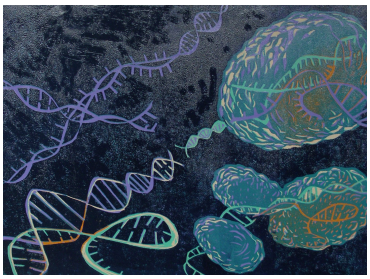
UPR . . . . . unfolded protein response

VST . . . . . variance stabilising transformation

WB . . . . . Western blotting

WGCNA . . . weighted gene co-expression network analysis

WT . . . . . wild type



1 Introduction

1.1 Motivation

Type 2 diabetes (T2D) affects more than 425 million people worldwide (IDF 2017). The pancreas, in particular its endocrine compartment, plays a key role in T2D pathogenesis (DeFronzo et al. 2015). Endocrine cells are arranged in small clusters (islets of Langerhans) scattered throughout the pancreas and secrete different hormones into the bloodstream to regulate glucose homeostasis (Röder et al. 2016) (alpha cells, glucagon (Gromada et al. 2007); beta cells, insulin (Cortizo et al. 1990); delta cells, somatostatin (Weir et al. 1985); PP cells, pancreatic polypeptide (Lonovics et al. 1981; Floyd 1980); epsilon cells, grehlin (Broglio et al. 2006)). Pancreatic beta cells are especially of great importance, as they are the only cell type proficient in lowering blood glucose levels via the release of insulin in response to glucose and other nutrient stimuli (Coore et al. 1964; Crespin et al. 1973; Malaisse et al. 1983; Gerich 1993; JC Henquin et al. 1994; JC Henquin et al. 2006).

T2D is caused by impaired beta cell function and insulin action resulting in hyperglycaemia (Alberti et al. 1998; Gavin et al. 2003). Prolonged elevated blood glucose levels often lead to secondary complications including cardiovascular failure and neuro-, nephro- and retinopathy (IDF 2017). Current available treatment

options are aimed at treating diabetes-related conditions, but unfortunately are not disease modifying. Thus, diabetes imposes a substantial economic burden on our health care system (Herman 2013) and novel targeted therapeutics are in dire need.

T2D is a complex multifactorial polygenic disease. While environmental factors including diet and lifestyle are major contributors to disease risk (Hu 2011), T2D susceptibility is also greatly influenced by genetic factors, as highlighted by studies in twins (Newman et al. 1987; Kaprio et al. 1992; Willemsen et al. 2015) and families (Almgren et al. 2011) as well as the discoveries made by genome-wide association, exome array and sequencing studies over the last 12 years (Fuchsberger et al. 2016; Morris 2018; Mahajan et al. 2018a).

Studying human genetics can help us understand the pathogenic pathways underlying T2D. This in turn will aid the discovery of novel biomarkers to promote early diagnosis, monitor disease progression and drive the development of new drugs to both refine current treatment strategies and uncover new interventions for diabetes prevention.

1.2 Endocrine pancreas development in human and mouse - from common progenitors to mature beta cells

1.2.1 Lessons learnt from rodent beta cell development

Pancreas development has been extensively studied in model organisms and observations have been extrapolated to human organogenesis, due to limited access to human fetal samples (Murtaugh 2007; Jørgensen et al. 2007; FC Pan et al. 2011). In particular, transcription factor (TF)-specific knockout (KO) studies in mice have successfully identified regulatory genes essential for pancreas development and in humans aided the identification of the genetic aetiology underlying monogenic forms of diabetes comprising neonatal diabetes mellitus (NDM, diagnosed within the first six months of life) and maturity onset diabetes of the young (MODY, which

typically manifests before the age of 25 years). For example, inactivating mutations in TF genes important for endocrine cell development including *NEUROD1* (Rubio-Cabezas et al. 2010), *RFX6* (Sansbury et al. 2015), *PDX1* (Stoffers et al. 1997; De Franco et al. 2013), *PAX6* (Solomon et al. 2009), *GLIS3* (Senée et al. 2006), *PTF1A* (Sellick et al. 2004), *NEUROG3* (Rubio-Cabezas et al. 2011), *NKX2.2* (Flanagan et al. 2014), *MNX1* (Flanagan et al. 2014), *HNF4A* (Yamagata et al. 1996b) and *HNF1A* (Yamagata et al. 1996a) have been shown to be responsible for the development of NDM and/or MODY in humans. Patients with homozygous mutations in *NEUROD1*, *PDX1*, *PTF1A*, *RFX6*, *NKX2.2* and *GLIS3* revealed the same phenotype observed in mice with biallelic inactivation mutations in the same genes, highlighting how valuable rodent models have been for our understanding of human beta cell development and the identification of key developmental regulators (Naya et al. 1997; Jonsson et al. 1994; Sellick et al. 2004; Smith et al. 2010; Briscoe et al. 1999; Watanabe et al. 2009).

1.2.2 Humans and mice share key events in pancreas development, but differ in fine details

Insights into human pancreas organogenesis from histological analysis of cadaveric fetal tissue at different developmental time points have revealed some species-specific differences in endocrine pancreas development between human and mouse (Jennings et al. 2013). Inter-species differences are present primarily in the timing and function of islet TFs. For example, compared to mouse, human pancreas and islet formation is characterised by delayed expression of GATA4 and PDX1, prolonged expression of SOX17 and the absence of NKX2.2 before endocrine commitment (Jennings et al. 2013). While in mice a bi-phasic wave of NEUROG3 expression has been observed (Villasenor et al. 2008), there is only a single transient expression phase in humans (Salisbury et al. 2014), which might underlie the different timing of NKX2.2, as NKX2.2 expression has been suggested to depend on NEUROG3 (PS McGrath et al. 2015).

Humans and mice have been shown to differ in the temporal occurrence of endocrine subpopulations (Piper et al. 2004; Johansson et al. 2007; Schwitzgebel et al. 2000). A study using conditional deletion of *Neurod1* in *Nkx2.2* KO mice revealed that repression of *Neurod1* by NKX2.2 in pancreatic progenitors resulted in commitment to the alpha cell lineage (Mastracci et al. 2013), in line with alpha cells having been reported as being the first detectable endocrine cell type in rodents (Johansson et al. 2007; Schwitzgebel et al. 2000). As NKX2.2 is not expressed in human pancreatic progenitors, transcriptional interaction between this TF and NEUROD1 is unlikely and provides a potential explanation why beta cells are the first endocrine cell type arising during human pancreas development (Jeon et al. 2009; Piper et al. 2004).

During endocrine differentiation the presence of polyhormonal cells, expressing insulin and glucagon, have been observed in both human and mouse (Bocian-Sobkowska et al. 1999; Riedel et al. 2012; Herrera 2000). In mice it has been suggested that these polyhormonal cells do not become part of adult islets (Herrera 2000). This is based on the observation that once a cell actives the insulin promoter, it will only develop into a beta cell, but cannot become an alpha cell anymore. Vice versa, once a cell shows glucagon promoter activity it will not become a beta cell anymore (Herrera 2000). In humans, on the other hand, polyhormonal endocrine cells have been reported to express the alpha cell specific TF ARX, but lack beta cell specific TFs, suggesting that polyhormonal endocrine cells represent a transitionary phase of cells likely to adapt alpha cell fate (Riedel et al. 2012).

Finally, adult human beta cells have been shown to maintain MAFB expression in parallel to MAFA (Jeon et al. 2009), while in mice a switch from MAFB to MAFA has been observed perinatally in beta cells (Hang et al. 2011; Hang et al. 2014).

1.2.3 Pancreatic beta cells maintain glucose homeostasis

The central function of mature beta cells is to maintain glucose homeostasis. They synthesise insulin, the only blood glucose-lowering hormone within the human

body, store it in secretory granules and secrete it in response to elevated blood glucose levels (JC Henquin et al. 1994; JC Henquin et al. 2006). Released insulin stimulates glucose uptake in peripheral tissues including muscle, liver and adipose tissue (Dimitriadis et al. 2011). Low glucose levels suppress insulin secretion and trigger glucagon release from alpha cells, which stimulates gluconeogenesis and glycogenolysis in the liver to restore normal blood glucose levels (Gerich 1993; Quesada et al. 2008). Interestingly, while humans possess a single *INS* gene, mice have a two-gene system composed of *Ins1* and *Ins2* (Owerbach et al. 1980; Shiao et al. 2008). In addition, characterisation of insulin promoters in multiple species has revealed profound differences between humans and rodents, in particular in the occurrence of regulatory elements (Hay et al. 2006).

An essential cellular component in insulin synthesis is the endoplasmic reticulum (ER), where proinsulin gets processed to generate the mature form of insulin, which subsequently gets packaged into secretory granules in the Golgi complex (M Liu et al. 2014). The majority of protein flux through the ER can be ascribed to the processing of insulin. Increased glucose levels stimulate insulin as well as overall protein synthesis (Goodge et al. 2000). A quick rise in insulin translation can lead to an accumulation of unfolded proinsulin in the ER causing ER stress, which in turn triggers the unfolded protein response (UPR) pathway (Ellgaard et al. 1999; Kaufman et al. 2010). The UPR regulates various processes with the aim to reduce ER workload, including enhancement of protein folding activity, slowing of translation, stimulation of mRNA degradation and as a last resort triggering of apoptosis (Scheuner et al. 2005; Ron et al. 2007; Oslowski et al. 2010).

Glucose enters pancreatic beta cells via facilitated diffusion through glucose transporters (GLUT) (McCulloch et al. 2011). Once inside the cell, it is phosphorylated by the glucokinase enzyme and metabolised through aerobic glycolysis (De Vos et al. 1995; X Guo et al. 2012). This results in an increased ATP/ADP ratio, which in turn closes K_{ATP} channels (ATP-sensitive potassium channels), causing depolarisation of the beta cell membrane and the opening of voltage-sensitive Ca^{2+} channels (Tarasov et al. 2012; Tarasov et al. 2004; Rutter et al. 2017). Subsequent Ca^{2+}

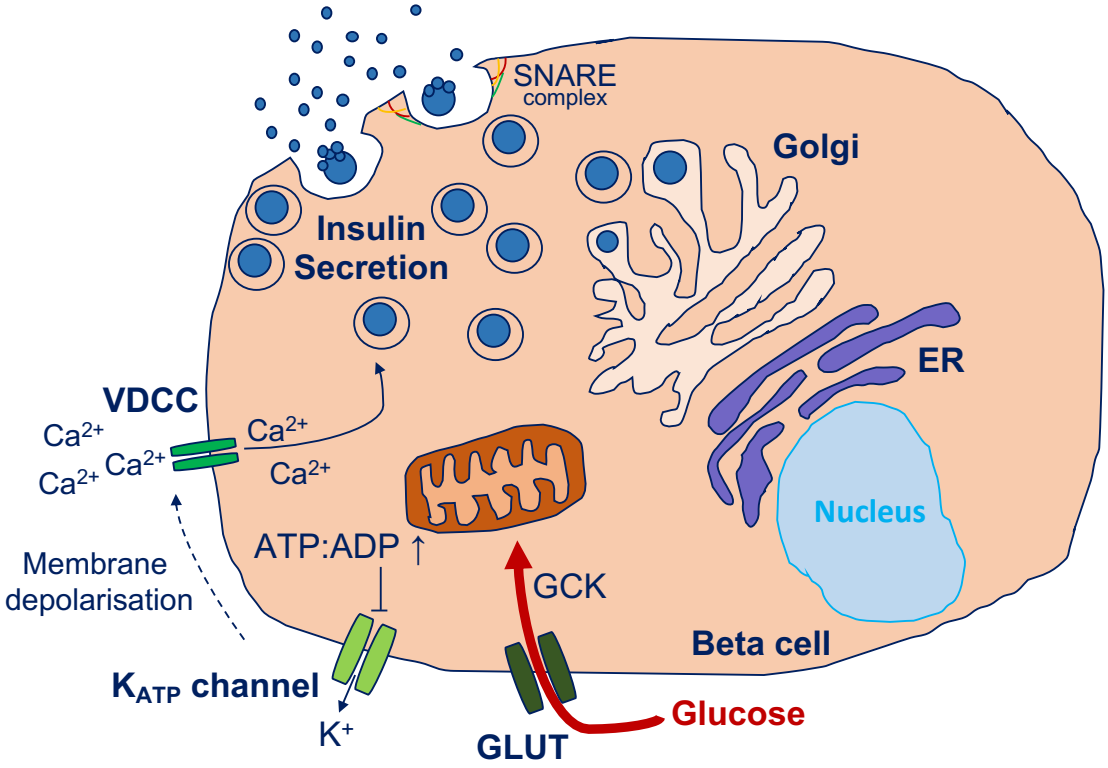


Figure 1.1: Model of glucose-stimulated insulin secretion in the pancreatic beta cell Glucose enters the beta cell through glucose transporters (GLUT). Inside the cell it gets phosphorylated by the glucokinase enzyme (GCK) and metabolised through aerobic glycolysis which leads to the generation of ATP. An increase in the ATP:ADP ratio inhibits K_{ATP} channel activity, resulting in membrane depolarisation and opening of voltage-dependent Ca^{2+} channels (VDCC). Ca^{2+} influx triggers SNARE complex-mediated fusion of docked insulin granules and exocytosis of insulin.

influx stimulates fusion of docked insulin granules mediated via SNARE proteins, leading to exocytosis of insulin (Thurmond 2007; Rutter 2004) (Figure 1.1).

Insulin secretion can be amplified by 3', 5'-cyclic adenosine monophosphate (cAMP), which is generated by an adenylyl cyclase (AC)-catalysed conversion of ATP to cAMP in response to glucose (Delmeire et al. 2003). A rise in cAMP levels activates protein kinase A, augmenting insulin secretion, potentially through direct phosphorylation of proteins that form part of the insulin secretory machinery such as SNAP25 or syntaxin-4 (Ammälä et al. 1993; Insel et al. 2003; Kaihara et al. 2013). The pharmacological agent forskolin can activate ACs by directly interacting with their catalytic subunit, thereby stimulating cAMP generation (Hodson et al. 2014).

1.3 Overview of the current understanding on the genetic basis of T2D

1.3.1 Genome-wide identification of T2D susceptibility loci

Genome-wide association studies (GWAS) interrogating common (minor allele frequency (MAF) >5%) single nucleotide polymorphisms (SNPs) in individuals with and without T2D genome-wide, have identified over 400 signals (mapping to >200 loci) significantly associated with T2D risk (Scott et al. 2017; Mahajan et al. 2014; Fuchsberger et al. 2016; Mahajan et al. 2018a; Spracklen et al. 2019).

The majority of identified genetic variants reside in non-coding regions (introns or intergenic) (Hindorff et al. 2009; Mahajan et al. 2018a). Due to linkage disequilibrium (LD) T2D association signals often span vast genomic regions. Thus, a GWAS lead or index SNP does not necessarily represent the causal variant; it might only tag a LD block (a DNA region transmitted together from one generation to the next) which harbours the causal variant. Pinpointing causal variants and identifying the genes through which variants mediate their effects (effector transcripts), is one of the early challenges in translating T2D genetic discoveries into disease mechanisms.

Genetic fine-mapping can help to localise probable causal variants. It defines a 99% credible set, consisting of the minimum number of SNPs that together account for 99% of the posterior probability of driving the T2D association for each of the identified distinct association signals (Maller et al. 2012). Applying fine-mapping to GWAS loci has shown that independent association signals from multiple causal variants can extend over broad genomic regions. For example, at the *KCNQ1* locus 15 and 16 independent association signals with locus-wide significance have been identified in Europeans and East Asians, respectively (Gaulton et al. 2015; Horikoshi et al. 2016; Scott et al. 2017; Mahajan et al. 2018a; Spracklen et al. 2019).

T2D signals associated with causal coding variants have been valuable in facilitating rapid biological translation, as in most cases, protein-altering variants directly point to the effector transcript (Figure 1.2). Coding variants associated

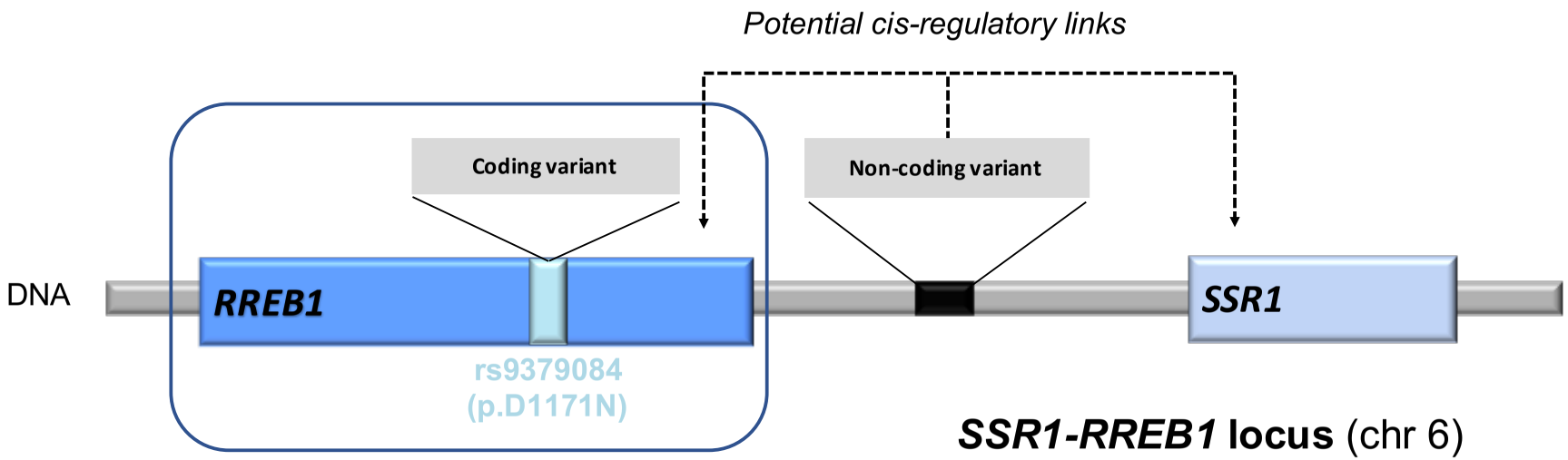


Figure 1.2: Identification of causal variants Non-coding variants located in introns or intergenic regions highly likely exert their effects through gene regulation (regulatory variants). As T2D-associated loci usually contain multiple genes and transcripts, identity of genes, regulatory variants act through, are often unclear. Historically, T2D-associated SNPs have been named after the nearest gene. Identification of coding protein-altering variants can directly point to the effector transcript and facilitate rapid clinical translation. Identification of a coding variant in exon 10 of *RREB1* established this gene as effector transcript at the *SSR1-RREB1* locus on chromosome 6. Image was modified from Thomsen et al. 2014.

with T2D risk have been identified in a small number of genes including *GCKR* (p.P446L) (Orho-Melander et al. 2008), *KCNJ11*/*ABCC8* (*KCNJ11* p.E23K, *ABCC8* p.A1369S) (Gloyn et al. 2003; Florez et al. 2004), *PAM* (p.S539W and p.D563G) (Steinthorsdottir et al. 2014; Fuchsberger et al. 2016; Lek et al. 2016) and *SLC30A8* (p.R325W) (Sladek et al. 2007). Functional studies have established these genes as effector transcripts and highlighted potential T2D disease mechanisms. For example, for *GCKR*, which encodes the glucokinase regulatory protein (GKRP), a coding variant-mediated effect on hepatic glucose metabolism could be shown, i.e. p.P446L increased hepatic glucose production through attenuating GKRP-mediated inhibition of glucokinase in liver, fitting the association with T2D and metabolic traits (Beer et al. 2009). In addition, functional studies in *PAM* demonstrated altered enzyme activity through reduced protein expression for the p.S539W risk allele and decreased catalytic function for the p.D563G missense allele (Thomsen et al. 2018).

Interestingly, coding variant fine-mapping efforts to refine the accuracy of validated target genes showed that a third of interrogated coding variants were highly unlikely to be causal (Mahajan et al. 2018b). Presumably in these cases, non-coding variants residing in one LD block with coding variants are drivers of the GWAS association. Therefore, to avoid incorrect inference of the effector transcript, it is crucial to integrate available GWAS data in fine-mapping efforts at loci containing coding variants.

1.3.2 Epigenomic annotations in human islets aid the identification of causal regulatory variants affecting beta cell function

Genetic regulatory elements comprising promoters, enhancers, insulators, transcribed and repressed regions drive cell type-specific regulation of gene expression, thereby controlling cell-specific processes (ENCODE Project Consortium 2012). Genome-wide chromatin accessibility can be profiled combining next generation sequencing with new or refined biochemical techniques, including ATAC-Seq (assay for transposase-accessible chromatin using sequencing) (Buenrostro et al. 2013),

FAIRE-Seq (formaldehyde-assisted isolation of regulatory elements using sequencing) (Giresi et al. 2007) and DNase-Seq (Thurman et al. 2012). Chromatin state can be further characterised through whole-genome bisulfide sequencing (Busche et al. 2015; Volkov et al. 2017), which defines DNA methylation patterns, chromosome conformation capture sequencing methods (3C, 4C and Hi-C), which can aid the identification of genomic interactions including promoter-enhancer interaction (Javierre et al. 2016; Stevens et al. 2017; Manduchi et al. 2018), and chromatin immunoprecipitation (ChIP)-Seq, which can be used to interrogate histone modifications as well as to identify specific transcription factor binding sites (TFBS) (Visel et al. 2009; Landt et al. 2012). Histone modifications including H3K27ac (histone H3 acetylated at lysine 27), H3K27m3 (histone H3 trimethylated at lysine 27), H3K4me1 (histone H3 monomethylated at lysine 4), H3K4me3 (histone H3 trimethylated at lysine 4) and the CCCTC-binding factor (CTCF) can be used to define subclasses of regulatory elements (Heintzman et al. 2009; Ernst et al. 2010; VW Zhou et al. 2011; ENCODE Project Consortium 2012). For example, active enhancers are marked by H3K4me1 and H3K27ac, while poised enhancers are enriched for H3K4me1 as well as H3K27me3 and inactive enhancers only contain H3K4me1 deposition (Pradeepa 2017).

Integration of GWAS data with annotations of tissue-specific regulatory regions have revealed that complex disease-associated non-coding genetic variants are enriched in enhancer regions of disease relevant cell types (Hindorff et al. 2009; Parker et al. 2013; Trynka et al. 2013; Corradin et al. 2014; Pasquali et al. 2014). As enhancers often contain clusters of binding sites of cell-specific TFs, it has been suggested that genetic variants might act through TF binding motif modulation (Buecker et al. 2012).

Over the last decade multiple research groups have put effort into mapping chromatin states in human islets as well as profiling their transcriptomic landscape (Gaulton et al. 2010; Stitzel et al. 2010; Parker et al. 2013; Pasquali et al. 2014; Thurner et al. 2018). They have identified islet-specific super enhancers (high levels of H3K27ac marks) and stretch enhancers (enhancers $\geq$3kb), clusters of

binding motifs of major human islet TFs (PDX1, NKX6.1, NEUROD1, NKX2.2, MAFB and FOXA2) in enhancer regions and discovered that islet stretch and super enhancers are particularly enriched for T2D- and fasting glucose (FG)-associated GWAS loci, emphasising a primary role for beta cell dysfunction in T2D pathogenesis (Hnisz et al. 2013; Parker et al. 2013; Pasquali et al. 2014; Thurner et al. 2018). For example, integration of 350 SNPs in strong LD with T2D- or FG-associated variants with FAIRE-Seq data, identified 38 SNPs at 10 loci (*TCF7L2*, *CDKAL1*, *CDC123/CAMK1D*, *CDKN2A/CDKN2B*, *FTO*, *G6PC2*, *HNF1B*, *IGF2BP2*, *SLC30A8*, *THADA*) that overlapped open chromatin sites in human islets (Gaulton et al. 2010), suggesting that non-coding variants likely affect gene regulation (regulatory variants) and highlighting the biological importance of characterising non-coding genomic regions in disease-relevant tissues. In addition, Varshney et al. showed an enrichment of T2D-associated SNPs in RFX binding sites mapping to islet accessible chromatin (stretch enhancers) with risk alleles predicted to disrupt the RFX binding motif (Varshney et al. 2017) and a study using DNAse hypersensitive sites (DHSs) sequencing to identify regulatory DNA sequences showed that T2D variants map to binding sites of monogenic diabetes-associated TFs such as HNF4A, HNF1A, HNF1B, NEUROD1, PDX1 and FOXA2 in DHSs (Maurano et al. 2012).

Integration of credible set variants with regulatory annotations in human islets have greatly advanced the identification of causal regulatory variants at T2D GWAS loci and helped prioritise potential effector transcripts for functional follow-up studies (Fuchsberger et al. 2016). At the *ZFAND3* locus fine-mapping identified a variant that displayed allele-specific binding of NEUROD1 and abolished enhancer activity in beta cells (Pasquali et al. 2014). Similarly, fine-mapping of a T2D- and proinsulin-associated genomic locus on chromosome 15 and integration with islet regulatory annotations revealed that rs7163757 overlapped a conserved open chromatin site in human islets (Kycia et al. 2018). The T2D risk allele was associated with increased expression of *C2CD4B*, highlighting this gene as being the likely effector transcript, as well as increased enhancer activity, which was

suggested to be mediated by the TF NFAT, pointing to a potential involvement of the calcineurin-NFAT pathway in T2D susceptibility (Kycia et al. 2018).

Fine-mapping of common variants at the *MTNR1B* locus and integration with islet regulatory annotations revealed that rs10830963 overlapped a FOXA2 binding site in human islets. Functional studies showed that the T2D risk allele increased FOXA2-bound enhancer activity in islet- as well as liver-derived cells and created a new NEUROD1 binding site, which was preferentially bound by the T2D risk allele in beta cells (Gaulton et al. 2015; Tuomi et al. 2016). In addition, using ChIP and electromobility shift assays (EMSAs), Fogarty et al. showed FOXA1 and FOXA2 allele specific-binding of rs11257655 at the *CDC123/CAMK1D* locus (Fogarty et al. 2014). Another study using ATAC-Seq open chromatin data confirmed allelic imbalance for rs11257655 and overlap with a hypomethylated open strong enhancer region (Thurner et al. 2018). While deletion of the enhancer region reduced expression of *CAMK1D* as well as a more distal gene *PPTN*, activation had the opposite effect in beta cells, exemplifying multi-gene effects of T2D-relevant enhancers and supporting *CAMK1D* as effector transcript (Miguel-Escalada et al. 2019).

Epigenomic annotations prioritised rs11708067 as the causal variant at the *ADCY5* locus, since it overlapped an islet enhancer region and co-localised with an islet eQTL (expression quantitative trait locus, T2D risk allele reduced *ADCY5* expression) as well as a methylation QTL (T2D risk allele increased gene methylation) (Thurner et al. 2018). ATAC-Seq and next generation Capture-C in beta cells revealed physical contact of the promoter with a hypomethylated open chromatin enhancer containing rs11708067 (Thurner et al. 2018). These findings were supported by another study observing less acetylated H3K27 ChIP-Seq reads for the rs11708067 T2D risk allele in human islets (Roman et al. 2017). Deletion of the enhancer element in a rodent beta cell line reduced *Adcy5* expression and insulin secretion, further supporting rs11708067 as the likely causal variant (Roman et al. 2017).

Modification of PAX6 binding has been suggested as a potential mechanism underlying T2D-associated variants at the *ARAP1/STARD10* (Kulzer et al. 2014) and *KLHDC5* loci (Kulzer et al. 2014; Thurner et al. 2018). *KLHDC5* rs10842991 was the only variant characterised by allelic imbalance (with the T2D risk allele showing greater chromatin accessibility) and predicted to overlap a PAX6 binding motif (with the T2D risk allele suggested to enhance TF binding) in addition to mapping to an open strong enhancer region in human islets (Thurner et al. 2018). At the *ARAP1/STARD10* locus, allele-specific *ARAP1* transcriptional activity and binding of PAX6 was observed for rs11603334 in rodent beta cells, thus Kulzer et al. favoured *ARAP1* as likely effector transcript at this GWAS locus (Kulzer et al. 2014). Interestingly, dense fine-mapping of variants at the *ARAP1/STARD10* locus, integration with islet chromatin state maps and analysis of promoter-reporter assays identified rs140130268, which is localised in intron two of *STARD10* and overlaps an active islet stretch enhancer, as the only variant causing allelic differences in enhancer activity (Carrat et al. 2017). Subsequent 3C analysis in human beta cells revealed interaction of the enhancer region containing rs140130268 with the promoter region of *STARD10*. As beta cell specific *Stard10*, but not *Arap1* KO mice displayed impaired glucose-stimulated insulin secretion (GSIS) and recapitulated the proinsulin phenotype seen at the human GWAS signal, Carrat et al. strongly supported *STARD10* as effector transcript through which T2D risk variants mediate their effects in beta cells (Carrat et al. 2017).

The first study using single nuclei ATAC-Seq epigenomics in human islets identified almost 250,000 accessible chromatin states in endocrine cells comprising alpha, beta and delta cells, which were further sub-clustered into hormone-high and -low states based on insulin and glucagon promoter accessibility. T2D- and FG-associated variants showed differences in enrichment between hormone-high and -low states of endocrine cells (Chiou et al. 2019). For example, beta cells with increased *INS* promoter accessibility showed stronger enrichment for FG-associated variants than those with lower *INS* promoter accessibility (Chiou et al. 2019). Integration with Hi-C data revealed cell type- and hormone state-specific

co-accessibility of distal chromatin sites with gene promoters (Chiou et al. 2019). For example, the *PDX1* promoter displayed cell type-specific co-accessibility with chromatin sites in beta cells and *INS*-high states showed co-accessible links with active *NEUROD1* or *INS* gene promoters (Chiou et al. 2019). In addition, distal co-accessible chromatin sites were enriched for T2D risk loci. Of 239 fine-mapped T2D-associated signals localised in accessible chromatin in human islets, 97 had co-accessibility with a gene promoter (Chiou et al. 2019). At the *KCNQ1* locus rs231361 mapped to a beta cell specific enhancer with long-range *INS*-high state-specific co-accessibility to the *INS* promoter. In addition, rs231361 was reported to disrupt a *INS*-high state-enriched RFX binding motif in beta cells, highlighting it as a likely causal variant at the *KCNQ1* locus that mediates its effects on T2D risk via beta cell function in a state-dependent way. Deletion of the enhancer region containing rs231361 revealed no effect on beta cell differentiation, i.e. NKX6.1 protein levels were unaffected, but a significant reduction in insulin gene and protein expression, confirming a link between the variant-harbouring enhancer and the *INS* gene promoter (Chiou et al. 2019).

1.3.3 Chromatin states change during endocrine pancreas development - implications for regulatory variants affecting beta cell differentiation

Mapping of regulatory sequences during pancreatic differentiation has revealed that chromatin states are highly dynamic which allows coordinated interaction of lineage-specific TFs and enhancers for precise spatial and temporal control of gene expression during development (Calo et al. 2013). For example, accessible chromatin regions in pancreatic progenitors are enriched for FOXA motifs and FOXA2 has been reported to be crucial for the remodelling of the chromatin architecture as well as the recruitment of GATA6 to enhancers during pancreatic differentiation (K Lee et al. 2019). Furthermore, it was shown that pancreatic enhancers actively acquire a poised state prior to activation (A Wang et al. 2015). For endoderm-derived cell lineages enhancer priming was shown to happen during pancreatic gut tube

(PGT) formation and to be indicative of developmental competence, i.e. PGT cells that acquire poised enhancer states are able to respond to inductive signals during endocrine cell differentiation (A Wang et al. 2015). In addition, binding sites for FOXA, GATA, PDX1, HNF4A, HNF1 and RFX TFs were found to be enriched in pancreatic enhancers (A Wang et al. 2015). TF knockdown (KD) studies have suggested a step-wise enhancer assembly model including acquisition of poised states and subsequent binding of FOXA TFs before induction and recruitment of PDX1 which in turn activates enhancers via histone acetylation (A Wang et al. 2015). Interestingly, similar enhancer priming was observed in pancreatic endoderm (PE) cells. PE poised enhancers were shown to have the competency to activate genes involved in regulating endocrine cell function. For example, six *SLC30A8*-associated enhancers were poised in PE and active in islets (A Wang et al. 2015). In addition, RFX and FOXO1 motifs were enriched at PE-poised enhancers before becoming active in islets, suggesting a role for these TFs in activation of islet enhancers during late endocrine differentiation (A Wang et al. 2015).

Consideration of human genetic disorders caused by defects in pancreas development can also point towards important developmental genomic loci. Integration of whole-genome sequencing data from individuals with isolated pancreatic agenesis with epigenomic annotations in pancreatic progenitors identified novel mutations in a hitherto unknown *PTF1A*-regulating enhancer as disease causative (Weedon et al. 2014). Mutations abolished enhancer activity by disrupting binding of either FOXA2 or PDX1 in pancreatic progenitors (Weedon et al. 2014).

A small subset of T2D-associated GWAS loci are located in or close to genes encoding TFs that govern pancreas development and in which rare mutations are causal for monogenic forms of diabetes including *NEUROG3*, *HNF1A*, *HNF4A*, *FOXA2*, *PAX4* and *NOTCH2* (Vaxillaire et al. 2012; Rubio-Cabezas et al. 2014; Schwitzgebel 2014; Gaulton et al. 2015; Fuchsberger et al. 2016; Bonàs-Guarch et al. 2018), raising the question that T2D-associated variants could exert their effects through modulation of developmental processes. In line with this, T2D risk alleles at the *KCNQ1* locus, which is imprinted during early pancreas development, but

not in adult islets, have been shown to modulate methylation status of regulatory sequences in fetal pancreas, suggesting that T2D risk effects are likely conveyed during beta cell development (Travers et al. 2013).

1.4 Islet cell models and genome editing tools

The majority of GWAS loci have pointed to a central role for pancreatic islets in T2D disease susceptibility, as evidenced by the enrichment of T2D variants in islet *cis*-regulatory elements (Thurner et al. 2018; Fuchsberger et al. 2016; Gaulton et al. 2015; Pasquali et al. 2014; Gaulton et al. 2010) and common association between T2D risk and quantitative islet function measures (Dupuis et al. 2010; Morris et al. 2012; Wood et al. 2017). Especially beta cells are of great importance, as the majority of T2D-associated loci appear to influence insulin secretion (Florez 2008; Ingelsson et al. 2010; Voight et al. 2010; Dupuis et al. 2010; Dimas et al. 2014).

Translation of T2D genetic discovery efforts into potential disease mechanisms has been hampered by the lack of authentic human islet cell systems for study. Table 1.1 gives an overview of the advantages and limitations of the currently available islet and beta cell models.

1.4.1 Rodent cell models

Although animal models and rodent cell lines have been extremely valuable in diabetes disease modelling (Sellick et al. 2004; Watanabe et al. 2009; Flanagan et al. 2014; Carrat et al. 2017), they do not always recapitulate human phenotypes (De Vos et al. 1995; Cabrera et al. 2006; Fiaschi-Taesch et al. 2009; Steiner et al. 2010; Bosco et al. 2010; da Silva Xavier et al. 2013). For example, mice characterised by heterozygous inactivating mutations in *Hnf4a* and *Hnf1a* do not become diabetic, while patients carrying similar mutations in these genes develop MODY (Ryffel 2001). Rare loss-of-function (LOF) variants in *SLC30A8*, which encodes the beta cell zinc transporter ZnT8, have been associated with protection against the development of

Table 1.1: Human pancreatic islet and beta cell models

	Advantages	Limitations
Isolated primary islets[1]	Contain all endocrine cell types	Scarce availability
		Heterogeneous regarding genotype
		Highly variable regarding quality/ viability
		Hard to maintain cell functionality *in vitro*
		Quiescent (genetic manipulation limited to viral expression vectors or shRNA)
Human beta cell lines[2]	Expandable	Immortalised/ proliferative
	Glucose-responsive	Aneuploid
	Replicate functional properties of human primary beta cells	Low clonal efficiency (gene editing at clonal level challenging)
	Transcriptomically similar to human primary beta cells	Genetic manipulation limited (chemical transfer or electroporation with expression vectors or siRNA)
Human pluripotent stem cell-derived beta-like cells[3]	Unlimited expansion capacity	Variable differentiation efficiency between different lines
	Unlimited differentiation capacity	Limited metabolic maturation *in vitro*/ functionally immature
	Can be generated from patient cells	
	Islet cell development studies possible	
	Clonal nature	
	Genome editing possible	
	Can be transplanted into mice to generate 'humanised' mouse models	
	Remarkable capacity to functionally mature *in vivo* after transplantation into rodents	

[1]Lyon et al. 2015; Cooper et al. 2016; Bottino et al. 2004; Ihm et al. 2006; Kulkarni et al. 2014; NJ Hart et al. 2019;
[2]Ravassard et al. 2011; Hastoy et al. 2018; Lawlor et al. 2019; Tsonkova et al. 2017;
[3]Kroon et al. 2008; Bruin et al. 2013; Rezania et al. 2013; Pagliuca et al. 2014; Rezania et al. 2014; Hrvatin et al. 2014; Bruin et al. 2014; Grobarczyk et al. 2015; Kyttälä et al. 2016.

T2D in humans (Flannick et al. 2014; Flannick et al. 2019). Characterisation of multiple *Slc30a8* KO mouse models aiming to investigate how reduced ZnT8 levels confer protection, were inconclusive and did not recapitulate the protective effect associated with *SLC30A8* LOF mutations in humans. *Slc30a8* KO mice either revealed no effect on glucose homeostasis or impaired glucose tolerance (Lemaire et al. 2009; Pound et al. 2009; Pound et al. 2012; Wijesekara et al. 2010), exemplifying the limitations associated with the use of model organisms in human disease modelling.

Observed discrepancies are likely due to species differences in tissue development and organisation as well as whole-body physiology. In particular, human and mice have revealed species discrepancies in endocrine cell development and islet morphogenesis. While fetal human islets resemble the architecture of mouse islets (core-mantle organisation with beta cells forming the core and being surrounded by alpha and delta cells), adult islets consist of intermingled endocrine cells, comprising fewer beta cells and more alpha and delta cells than rodent islets (beta cells: ∼60% vs ∼80%, alpha cells: ∼30% vs ∼20%, delta cells: ∼10% vs <5%) (Cabrera et al. 2006; Steiner et al. 2010). Disease-causing alterations in genes encoding pancreas TFs have shown that some developmental processes are less conserved between species (Gradwohl et al. 2000; German-Diaz et al. 2017; Hancili et al. 2018; Xuan et al. 2012; ZD Shi et al. 2017; N Gao et al. 2008; K Lee et al. 2019). For example, while GATA6 haploinsufficiency is sufficient for pancreas agenesis in humans (Allen et al. 2011), this phenotype is only recapitulated in mice by simultaneous biallelic deletion of *Gata4* and *Gata6* (Xuan et al. 2012; Carrasco et al. 2012), suggesting species-specific roles for GATA6 in pancreas development. Characterisation of patient-specific human pluripotent stem cells (hPSCs) carrying mutations in both alleles of *GATA6* revealed impaired definitive endoderm (DE) formation (Tiyaboonchai et al. 2017). Re-expression of GATA6 or GATA4 rescued DE formation, signifying the TF's essential role in early human beta cell development (Tiyaboonchai et al. 2017). In addition, *GATA6* haploinsufficiency was shown to negatively affect the generation of pancreatic progenitor cells and subsequent formation of mature functional beta cells in humans, highlighting a

potential mechanism leading to pancreas hypoplasia observed in patients carrying heterozygous *GATA6* mutations (ZD Shi et al. 2017; Yorifuji et al. 2012).

A similar observation has been made for *Foxa* factors. Conditional KO of *Foxa2* in either endoderm or pancreatic progenitors was not sufficient to affect pancreas development, only biallelic deletion of *Foxa1* and *Foxa2* led to pancreas hypoplasia in mice (CS Lee et al. 2005; N Gao et al. 2008). Loss of *FOXA2* alone however, was responsible for a reduction in pancreatic progenitor cells in humans, suggesting a species-intrinsic gene dosage requirement for FOXA factors (K Lee et al. 2019).

1.4.2 Human beta cell models

Primary human islets, representing all endocrine cell types, are considered the gold standard for diabetes disease modelling (Kulkarni et al. 2012; Bernal-Mizrachi et al. 2014; NJ Hart et al. 2019). Publicly available gene expression, chromatin state and functional data from large collections of human islets have greatly advanced our understanding of genes with a potential role in human glucose homeostasis (Kaddis et al. 2009; Gaulton et al. 2010; Segerstolpe et al. 2016). Unfortunately, their scarcity, their heterogenity regarding donor genotype and viability/function after isolation and the difficulties associated with their genetic manipulation limits their use (Ihm et al. 2006; Bottino et al. 2004; Kulkarni et al. 2014; NJ Hart et al. 2019). For example, donor age, glycated haemoglobin levels (HbA1c, marker for blood glucose control within an individual over time, diabetes diagnosis criteria >6.5%) and the impact of isolation procedures including cold ischaemia time (CIT, cooling of pancreas under reduced blood supply) have a major influence on islet viability (Lyon et al. 2015; Cooper et al. 2016; Bottino et al. 2004; Hilling et al. 2014).

Immortalised human beta cell lines such as the EndoC-βH family have been widely used in diabetes research (Scharfmann et al. 2019). They represent an authentic model system to study beta cell physiology, as they display many functional properties of mature beta cells, including stimulus-coupled insulin secretion (Ravassard et al. 2011; Andersson et al. 2015; Hastoy et al. 2018; Lawlor

et al. 2019). However, due to their low recombination rates and the inability to expand single clones, genetic manipulation at the clonal level has proven to be challenging. Nonetheless, siRNA-mediated KD, CRISPR/Cas9-mediated KO and gene overexpression have been successfully performed (Thomsen et al. 2016; Grotz et al. 2019; Tsonkova et al. 2017).

1.4.3 Human pluripotent stem cells

A promising recently emerged alternative to existing islet and beta cell models are human pluripotent stem cells (hPSCs), comprising human embryonic stem cells (hESCs) and human induced pluripotent stem cells (hiPSCs). Human iPSCs can be generated by reprogramming any fully differentiated adult somatic cell (Takahashi et al. 2006; Takahashi et al. 2007), thereby circumventing the ethical and legal considerations surrounding the use of embryonic tissue (Hua et al. 2013). The possibility to derive hiPSCs from patient cells also opens new opportunities for diabetes disease modelling, as cell models with the disease-associated genetic background can be established and studied.

1.4.3.1 Evaluation of typical human PSC properties

Human PSCs are characterised by two distinct features, i.e. stemness or self-renewal (potential to replicate indefinitely) and pluripotency (capacity to differentiate into any specialised cell type of the human body) (Evans et al. 1981; GR Martin 1981). In addition, they display a normal human diploid karyotype, which consists of 22 pairs of autosomal and one pair of sex chromosomes, i.e. 46 chromosomes per cell (Evans et al. 1981).

As genome engineering (see Section 1.4.6) and prolonged culture might affect the integrity and identity of hPSCs, edited hPSC lines are routinely evaluated for typical stem cell properties including cellular properties (morphology), pluripotency (stem cell marker expression, teratoma and embryoid body formation) and genomic

integrity (karyotyping, SNP and somatic copy number variation (CNV) analysis). Table 1.2 gives an overview of common techniques used to assess typical hPSC properties.

Table 1.2: Human pluripotent stem cell properties

Category	Property	Examples
Cellular property	Morphology	Colony formation, round shape, large nucleus with pronounced nucleolus, scant cytoplasm
Pluripotency[1]	Stem cell gene expression	*POU5F1*, *SOX2*, *NANOG*
	Stem cell marker expression	SSEA-4, OCT4, GDF3, REX1, TRA-1-60, TRA-1-81
	Teratoma formation	Spontaneous formation of tumours derived from three germ layers (endo-, meso- and ectoderm) when PSCs are injected into immunodeficient mice
	Embryoid bodies	Formation of 3D aggregates containing differentiating hPSCs from all germ layers
Genomic integrity[2]	Karyotyping	Chromosome counting of DAPI-stained or Giemsa-stained (G-banding) metaphases to detect aneuploidy, detection limit: $\geq$5Mb chromosomal rearrangements
	SNP analysis	SNP microarrays (e.g. Illumina HumanCoreExome BeadChip), detection resolution: 100kb
	CNV analysis	High-resolution microarray-based comparative genomic hybridization (aCGH)
	TP53 mutations	SNP arrays

[1]Levine et al. 2006; Son et al. 2013; [2]Spits et al. 2008; Shaffer et al. 2004; Elliott et al. 2010; Maitra et al. 2005 ; Laurent et al. 2011; International Stem Cell Initiative et al. 2011; Wapner et al. 2012; H Wu et al. 2008 ; Martins-Taylor et al. 2011; Merkle et al. 2017; SNP, single nucleotide polymorphism; CNV, copy number variation.

A study screening protein-coding genes of >200 independent hPSC lines revealed that hPSC lines can acquire and expand mutations in the tumour suppressor gene *TP53*, which are commonly found in human cancers (Merkle et al. 2017). In addition, during genetic manipulation, endonuclease-induced DNA double-strand breaks (DSBs) have been shown to cause a p53-dependent toxic response in hPSCs (Ihry et al. 2018), suggesting that genome editing could result in a selection against cells with a functional p53 pathway, as mutated hPSCs might have a growth advantage in culture. As non-functional p53 has been shown to be involved in tumorigenesis, genome-edited hPSCs should therefore be carefully monitored for *TP53* mutations, especially, if cells might be utilised in the clinic (Rivlin et al. 2011).

1.4.4 Modelling beta cell development using human PSCs

1.4.4.1 *In vitro* differentiation of human PSCs along the endocrine lineage into beta-like cells

The last 10 years have seen unprecedented progress in *in vitro* differentiation protocols making it possible to generate beta-like cells from hPSCs, mimicking human beta cell development in a culture dish. In 2006, the first protocol was established by D'Amour et al. 2006 and has been refined since by multiple research groups (Kroon et al. 2008; Rezania et al. 2013; Rezania et al. 2014; Pagliuca et al. 2014; Aguayo-Mazzucato et al. 2015; Nostro et al. 2015; Veres et al. 2019). To date, *in vitro* beta cell differentiation represents seven key developmental stages of the endocrine pancreas comprising definitive endoderm (DE), primitive gut tube (PGT), posterior foregut (PFG), pancreatic endoderm (PE), endocrine precursors (EP), endocrine-like (EN) and beta-like cells (BLCs).

Figure 1.3 provides an overview of the step-wise *in vitro* beta cell differentiation protocol including culture conditions and duration of individual stages, used in this thesis and described in Perez-Alcantara et al. 2018.

Formation of DE is achieved by addition of Activin A, a member of the transforming growth factor beta (TGF-β) family, and CHIR99021, a highly selective

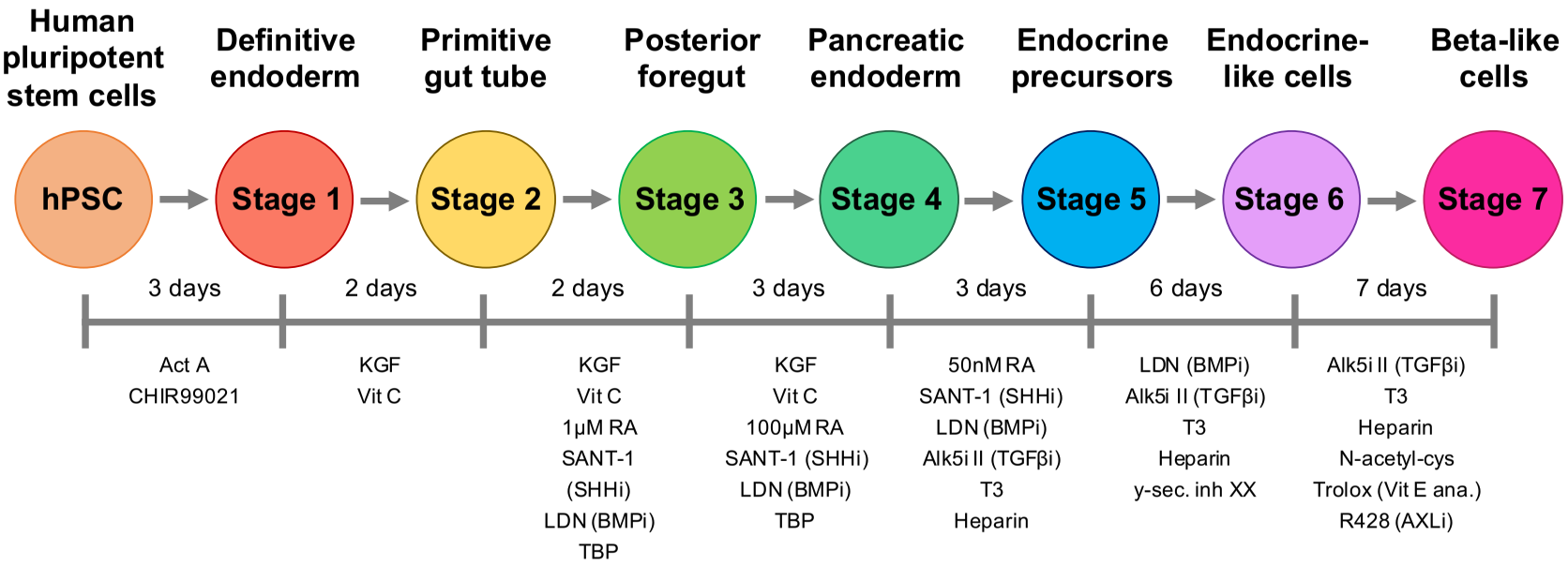


Figure 1.3: Overview of the *in vitro* beta cell differentiation protocol Schematic summarising the step-wise *in vitro* beta cell differentiation protocol from hPSCs to BLCs. Distinct stages are represented by individual circles. Timely addition of specific growth factors and small molecules (listed underneath) enable directed differentiation. Duration of incubations are indicated between stages. Act A, Activin A; CHIR99021, GSK-3 inhibitor; KGF, keratinocyte growth factor; Vit C, vitamin C; RA, retinoic acid; i, inhibitor; T3, thyroid hormone; R428, receptor tyrosine kinase AXL inhibitor (Rezania et al. 2014; Perez-Alcantara et al. 2018).

GSK-3 inhibitor, which modulates the Wnt signalling pathway that is involved in the maintenance of stem cell fate (McLean et al. 2007; X Xu et al. 2011). Induction of PGT cells is stimulated by keratinocyte growth factor (KGF) administration. Pancreatic progenitors (PFG and PE) are obtained by addition of KGF and retinoic acid (RA) as well as inhibition of various signalling pathways. For example, SANT-1 acts as a sonic hedgehog (SHH) signalling antagonist and the small molecule LDN is a potent inhibitor of the bone morphogenetic (BMP) pathway. Addition of TBP, a protein kinase C activator, negatively affects TGF-β signalling. Combined addition of these factors as well as vitamin C, which stimulates extracellular matrix secretion and cell proliferation (Choi et al. 2008) and negatively affects gene expression of the endocrine master regulator NEUROG3, has been shown to enhance the yield of PE cells (Rezania et al. 2014). Premature induction of *Neurog3* in mouse pancreatic progenitors has been reported to result in the generation of preferentially glucagon-expressing or polyhormonal endocrine cells (Johansson et al. 2007). Therefore, correctly timed expression of *NEUROG3* is important for endocrine lineage commitment and can be induced by administration of a TGF-β receptor inhibitor (e.g. ALK5 inhibitor II) to EP cells during stage 5 (Nostro et al. 2011; Rezania et al. 2012; Rezania et al. 2013; Rezania et al. 2014; Pagliuca et al. 2014). As heparin has been shown to improve viability of differentiated cells, it is often added to differentiating cells during stages 5, 6 and 7 (Rezania et al. 2014). Further differentiation into hormone-expressing EN cells has been reported to be achieved by inhibition of the Notch pathway through addition of γ-secretase inhibitor XX (Rezania et al. 2014; Pagliuca et al. 2014). Notch signalling plays a role in pancreatic endocrine/exocrine fate choice (Apelqvist et al. 1999) and ectopic activation of Notch signalling in pancreatic progenitors has been shown to block pancreas differentiation (Murtaugh et al. 2003; Zhu et al. 2016). Combined with the addition of thyroid hormone (T3) it was suggested to drive the formation of BLCs co-expressing NKX6.1 and insulin (Rezania et al. 2014). During the final stage of *in vitro* beta cell development, addition of the antioxidants N-acetyl cysteine (Harmon et al. 2005; Harmon et al. 2009) and the vitamin E analogue trolox, as

well as the small molecule R428, a receptor tyrosine kinase AXL inhibitor (Haase et al. 2013), ALK5 inhibitor II and T3 induces the expression of the key beta cell gene *MAFA* in BLCs (Rezania et al. 2014).

Mature beta cells are characterised by the expression of genes implicated in insulin secretion such as *ABCC8* and *KCNJ11*, encoding the subunits SUR1 (sulphonylurea receptor 1) and KIR6.2 (inwardly rectifying potassium channel) of the beta cell K_{ATP} channel, which is a key component in coupling metabolic signals to electrical activity leading to insulin secretion (Ashcroft 1988). Another maturity marker gene is *GJD2*, which encodes the gap junction protein delta 2 (or connexin 36) and has been implicated in beta cell electrical coupling playing an important role in physiological insulin release (Le Gurun et al. 2003; Serre-Beinier et al. 2009; Head et al. 2012; Farnsworth et al. 2014). Expression of *ABCC8* and *KCNJ11* have been observed in hPSC-derived BLCs, albeit at levels which are considerably lower compared to those in primary human islets (Rezania et al. 2014; Bruin et al. 2014).

In line with this, gene expression profiling of hPSC-derived BLCs, human fetal pancreata at gestational weeks 15-16 and adult human islets showed that the transcriptome of BLCs was more similar to fetal rather than adult beta cells (Hrvatin et al. 2014). For example, while genes important for beta cell fate (*PDX1*, *NKX6.1*, *MAFA*) and functional maturity (*PCSK1*, *SLC30A8*, *CHGB*) were significantly lower, genes implicated in alpha cell fate (*ARX*, *FOXA1*, *IRX2*) were markedly higher expressed in BLCs compared to adult beta cells (Hrvatin et al. 2014). Interestingly, functionally immature BLCs have been reported to have the capacity to further mature *in vivo*, when transplanted under the kidney capsule, and successfully correct diabetes in immunocompromised mice (Kroon et al. 2008; Bruin et al. 2013; Rezania et al. 2013; Pagliuca et al. 2014; Rezania et al. 2014). This suggests that yet unknown factors are present in the *in vivo* environment that are critical for functional maturation of BLCs. Identification of these factors will be important to generate authentic high quality functional beta cells from hPSCs in the future.

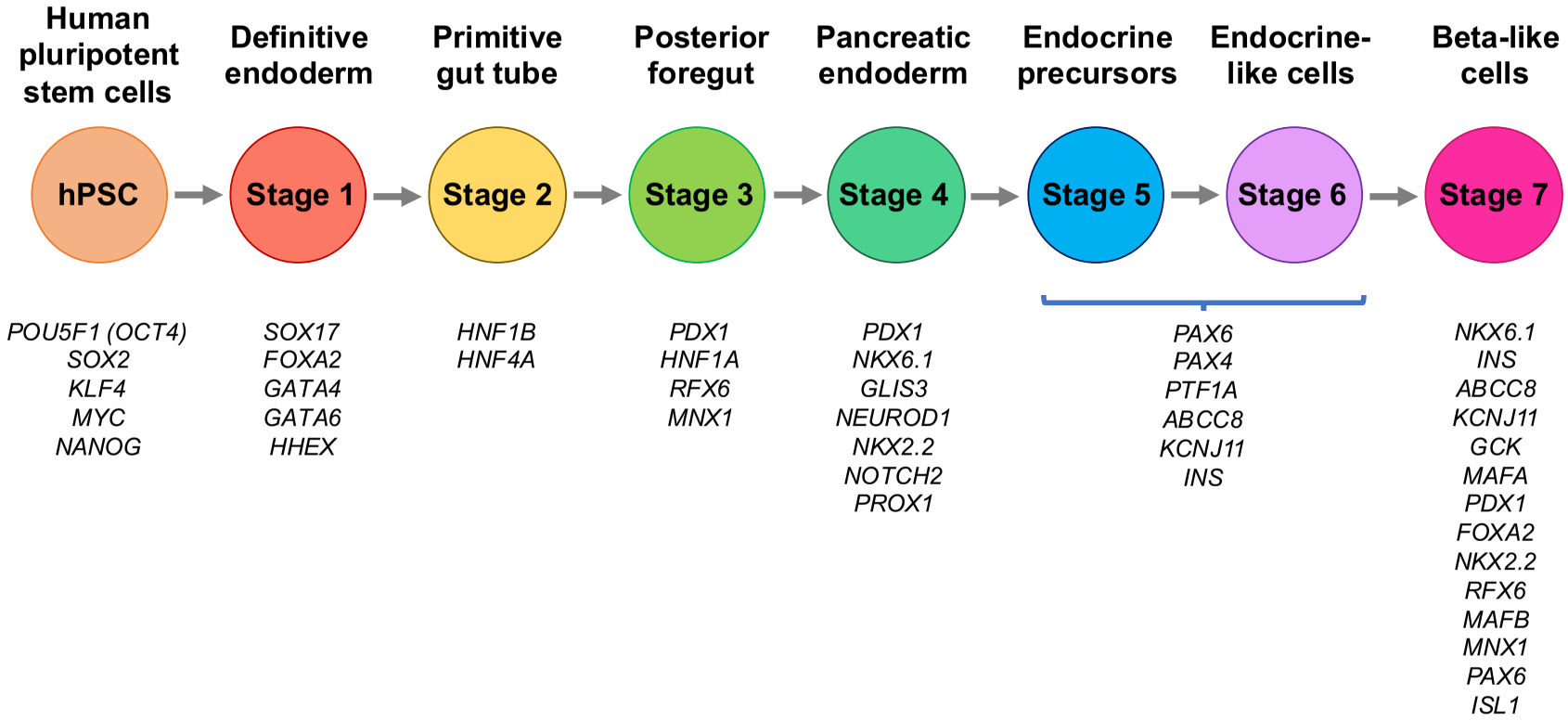


Figure 1.4: Discrete developmental stages during endocrine pancreas development Step-wise differentiation of human pluripotent stem cells (hPSCs) through distinct stages (represented by individual circles) of endocrine pancreas and beta cell development. Definitive endoderm and pancreatic endocrine precursors resemble early fetal pancreas tissue. They can differentiate into all pancreatic endocrine cell types as well as ductal cells. Stage 7 cells consist mainly of insulin-positive beta-like cells, but can also contain a small fraction of glucagon-positive cells. Genes encoding transcription factors and proteins which perform crucial biological functions in beta cell development and maturation or are important for beta cell identity are listed according to the stage(s) they are expressed in (Bruin et al. 2014; Rezania et al. 2014; Perez-Alcantara et al. 2018).

1.4.4.2 Complex TF networks govern pancreatic endocrine cell differentiation and are important for the maintenance of beta cell identity and functional maturity

Human pancreatic endocrine cells derive from common progenitors (Jennings et al. 2015). Complex TF networks initiate and maintain distinct gene expression programs to establish the five different endocrine cell subtypes, alpha, beta, delta, epsilon and PP cells. Extensive evaluation of pancreas development in mice revealed that endocrine differentiation does not follow a linear cascade of gene activation and repression, as the majority of TFs are expressed at more than one time and therefore often exert more than one role during development (Wilson et al. 2003).

Table 1.3 lists a subset of endocrine pancreas TFs including their temporal expression patterns during human pancreas development (Jennings et al. 2015) as well as observed expression profiles during hPSC-based *in vitro* beta cell differentiation (Bruin et al. 2014; Rezania et al. 2014; Perez-Alcantara et al. 2018) (Figure 1.4).

Early endoderm/Stage 1 - definitive endoderm

During human pancreas development early endoderm is marked by FOXA2 and SOX17 expression (Dufort et al. 1998). In line with this, formation of DE in mouse and human PSCs is initiated by the TFs SOX17 (Yasunaga et al. 2005; Qu et al. 2008; Séguin et al. 2008; D'Amour et al. 2005), HHEX (Rankin et al. 2011; Fisher et al. 2017) as well as GATA6 (Fisher et al. 2017; Tiyaboonchai et al. 2017; Chia et al. 2019) and marked by the expression of the chemokine receptor CXCR4 (KE McGrath et al. 1999; D'Amour et al. 2005).

Distal foregut endoderm/Stage 2 - primitive gut tube

Members of the hepatocyte nuclear factor TF family have been implicated in early pancreas development and are associated with MODY (Nishigori et al. 1998; Lindner et al. 1999; Yamagata et al. 1996b; Ng et al. 2019). Mice deficient of *Hnf1b* die before gastrulation due to impaired formation of endoderm (Barbacci et al. 1999;

Haumaitre et al. 2005). In addition, lack of HNF1B during development has been reported to result in a reduced number of pancreatic progenitors and defective generation of NEUROG3-expressing endocrine precursors in mice, suggesting a crucial role for HNF1B in early pancreas development (De Vas et al. 2015). The hepatocyte nuclear factor 4 (HNF4A) has a well established role in liver development and regulation of liver-specific gene expression (Lau et al. 2018). LOF mutations in *HNF4A* cause MODY, which is characterised by defective GSIS (Yamagata et al. 1996b; Byrne et al. 1995). Using *HNF4A*-MODY patient-derived iPSCs in a model of liver and pancreas development, it was shown that *HNF4A* haploinsufficiency affected foregut development and subsequent gene expression profiles in hepatocyte-like and beta-like cells (Ng et al. 2019).

Pancreatic-specified endoderm/Stage 3 - posterior foregut

FOXA2 has recently been identified as a pioneer factor in human pancreas specification (K Lee et al. 2019). Using an *in vitro* directed endocrine differentiation model in combination with ATAC-Seq open chromatin profiling, it was shown that FOXA2 expression was crucial for the generation of pancreatic progenitors from hPSCs and required for chromatin architecture remodelling necessary for the recruitment of pancreas-specific TFs during pancreatic differentiation (K Lee et al. 2019). Loss of FOXA2 in hPSCs led to a reduced number of PDX1-positive pancreatic progenitor cells (without affecting DE and PGT cell numbers) and impaired the recruitment of the TF GATA6 to pancreatic enhancers (K Lee et al. 2019).

In addition, FOXA2 has been reported to activate *PDX1*, which encodes the pancreatic and duodenal homeobox 1 protein, a master regulator of pancreas development and crucial in beta cell functional maturation in humans and mice (Gerrish et al. 2000; Stoffers et al. 1997; Jonsson et al. 1994; Ahlgren et al. 1998; Yang et al. 2011). In human pancreas development PDX1 expression is first observed around gestational week 4 and remains stably expressed in mature beta cells (Piper et al. 2004; Jennings et al. 2013). KO of *Pdx1* in beta cells leads to beta cell loss and

diabetes (Gannon et al. 2008). In addition, it has been shown that *Pdx1*-depleted beta cells can adopt alpha cell fate expressing *Mafb* and glucagon (T Gao et al. 2014). Thus, PDX1 seems to act as a master regulator of beta cell lineage specification through initiating expression of genes specific for beta cell identity while repressing genes important for alpha cell identity. Human PSC-based *in vitro* beta cell differentiation models have revealed a similar temporal expression pattern of PDX1 (Bruin et al. 2014). Expression is detected throughout the end of stage 3 (PFG) to stage 7 (BLC) and mutations in the transactivation domain in *PDX1* were shown to impair human endocrine progenitor and beta cell development (X Wang et al. 2019).

Endocrine differentiation - pancreatic bud (multipotent) and trunk (bipotent) progenitors/Stages 4, 5 and 6 - pancreatic endoderm, endocrine precursors and endocrine-like cells

The basic helix-loop-helix TF NEUROG3 is essential for endocrine lineage specification, i.e. its transient expression in endocrine precursor cells is crucial for the commitment of multipotent pancreatic progenitors into an endocrine cell fate, comprising all five endocrine subtypes (Rukstalis et al. 2009; Apelqvist et al. 1999; G Gu et al. 2002; Salisbury et al. 2014). In human pancreas development NEUROG3 expression is initiated in gestational week 8 and peaks at around week 11 (G Gu et al. 2002; Jennings et al. 2013; Salisbury et al. 2014). As NEUROG3 itself does not co-express with insulin, glucagon or somatostatin (Gradwohl et al. 2000), it can be used as an endocrine progenitor marker in hPSC-based beta cell differentiation models. *NEUROG3* expression is positively regulated by HNF1, FOXA and HNF6 (ONECUT1) and negatively regulated by HES-1, a downstream factor of Notch signalling, in humans and mice (Jacquemin et al. 2000; Jensen et al. 2000; JC Lee et al. 2001)

While mice deficient for *Neurog3* do neither form endocrine precursors nor any of the endocrine cell subtypes and die postnatally from diabetes (Gradwohl et al. 2000), humans carrying compound heterozygous mutations in *NEUROG3* have

Table 1.3: Expression patterns of transcription factors during endocrine pancreas development

TF	Full name	TF family	Expression during human pancreas development[1]	Expression during *in vitro* beta cell differentiation[2]
SOX17	SRY-box 17	HMG-box	distal foregut endoderm, pancreatic-specified endoderm	DE, PGT
HHEX	hematopoietically-expressed homeobox protein	homeobox		DE
HNF1B	hepatocyte nuclear factor 1 homeobox B	hepatocyte nuclear factor		PGT
HNF4A	hepatocyte nuclear factor 4 alpha	hepatocyte nuclear factor		
FOXA2 (HNF3β)	forkhead box A2	hepatocyte nuclear factor	distal foregut endoderm, pancreatic-specified endoderm, pancreatic bud (multipotent progenitors), trunk (bipotent) progenitor, beta cell	DE-BLC
GATA6	Gata-binding protein 6	zinc finger	pancreatic bud (multipotent progenitors)	DE
GATA4	Gata-binding protein 4	zinc finger	pancreatic-specified endoderm, pancreatic bud (multipotent progenitors)	DE-PE
PDX1	pancreatic and duodenal homoebox 1	parahox home-odomain	pancreatic-specified endoderm, pancreatic bud (multipotent progenitors), trunk (bipotent) progenitor, beta cell	PFG-BLC
SOX9	SYR-box 9	HMG-box	pancreatic-specified endoderm, pancreatic bud (multipotent progenitors), trunk (bipotent) progenitor	PFG, PE
NKX6.1	NK-homeobox	homoebox	pancreatic bud (multipotent progenitors), trunk (bipotent) progenitor, beta cell	PE-BLC
HNF6 (ONECUT1)	hepatocyte nuclear factor 6	hepatocyte nuclear factor		PFG
NEUROG3	neurogenin 3	bHLH	transiently expressed in endocrine progenitors, correlates with loss of SOX9	PE-EN, peaks at EP

Continued on next page

TF	Full name	TF family	Expression during human pancreas development[1]	Expression during *in vitro* beta cell differentiation[2]
NKX2.2	NK-homeobox		downstream of NEUROG3, beta cell	PE-BLC
PAX4	paired box gene 4	paired homeodomain	downstream of NEUROG3	EP
NEUROD1	neurogenic differentiation factor 1	bHLH	downstream of NEUROG3	PE-BLC, peaks at EN
PAX6	paired box 6	paired homeodomain	beta cell	EN, BLC
ISL1	insulin gene enhancer protein	LIM-homoedomain	downstream of NEUROG3, beta cell	EN, BLC
RFX6	regulatory factor X 6		beta cell	
RFX3	regulatory factor X 3		early endocrine progenitors, beta cell	
MAFB	V-maf musculoaponeurotic fibrosarcoma oncogene homolog B	bZIP	alpha and beta cell	EP-BLC
MAFA	V-maf musculoaponeurotic fibrosarcoma oncogene homolog A	bZIP	beta cell	BLC
HES-1	hairy and enhancer of split 1	bHLH	anti-endocrine, pro-exocrine (activated by Notch signalling, inhibits *Neurog3*)	
REST	RE-1 silencing transcription factor	Krueppel-like zinc finger		PGT-PE

[1]Human pancreas development and TF network reviewed in Jennings et al. 2015; [2]TF expression patterns during *in vitro* beta cell differentiation derived from Bruin et al. 2014; Rezania et al. 2014 and Perez-Alcantara et al. 2018; bHLH, basic helix-loop-helix; HMG, high mobility group; bZIP, basic leucine zipper; RE-1, repressor element 1; SRY, sex determining region Y; NK, Nierenberg and Kim; DE, definitive endoderm; PGT, primitive gut tube; PFG, posterior foregut; PE, pancreatic endoderm; EP, endocrine precursors; EN, endocrine-like cells; BLC, beta-like cells.

functional insulin-secreting endocrine cells despite lacking intestinal enteroendocrine cells (German-Diaz et al. 2017; Hancili et al. 2018). KO of *NEUROG3* in hESC lines and subsequent differentiation along the endocrine lineage *in vitro* revealed normal formation of pancreatic progenitors, but absence of endocrine cells, even after engraftment of pancreatic progenitor cells into NOD mice (PS McGrath et al. 2015). Interestingly, a reduction of *NEUROG3* expression of up to 90% was not sufficient to replicate the complete loss of *NEUROG3* phenotype, thus a

small amount of NEUROG3 seems to be adequate for the formation of pancreatic endocrine cells in humans (PS McGrath et al. 2015). Human *NEUROG3* mutations are likely hypomorphic and residual activity sufficient for normal endocrine cell function (PS McGrath et al. 2015).

NEUROG3 activates the TF NEUROD1 which plays an important role in endocrine lineage specification and differentiation (Naya et al. 1997; Huang et al. 2000; Gasa et al. 2008). Expression of NEUROD1 persists in mature beta cells, where it is involved in the regulation of the *INS* gene promoter in cooperation with PDX1 and ISL1 (Glick et al. 2000; H Zhang et al. 2009; C Gu et al. 2010; W Wang et al. 2016). Mice lacking *Neurod1* specifically in beta cells, revealed the importance of this TF in the maintenance of functional maturity and the differentiated state of beta cells (C Gu et al. 2010). Islets of *Neurod1* KO mice showed increased basal insulin secretion, oxygen consumption and expression of glycolytic genes, resembling the metabolic profile of immature beta cells (C Gu et al. 2010).

The paired homeodomain TF PAX4 plays a crucial role in endocrine pancreas differentiation. It is required for the development of both beta and delta cells by acting as a transcriptional repressor of alpha cell specific gene expression, thereby opposing ARX function (Sosa-Pineda et al. 1997; Collombat et al. 2003). PAX4 is only transiently expressed in the developing pancreas (Smith et al. 1999). Its expression is activated by pancreas-specific TFs, including NEUROG3 and HNF1A (Smith et al. 2003) and turned off, once sufficient protein levels are reached. In mice PAX4 has been shown to exerts its effects in parallel with NKX2.2 leading to increased levels of PDX1 and insulin (Junfeng Wang et al. 2004).

Similar to NEUROD1, NKX2.2 is initiated downstream of NEUROG3 and promotes the commitment of cells to alpha, beta and PP, but not delta cell lineages (Sussel et al. 1998; Lyttle et al. 2008; Jennings et al. 2013; Churchill et al. 2017). While in mice *Nkx2.2* is already expressed upstream of *Neurog3*, in human pancreas development it emerges after NEUROG3 and in hPSC-based differentiation models it has first been detected in PE cells with expression persisting in mature beta cells

(Jennings et al. 2013; Jennings et al. 2015; Rezania et al. 2014; Perez-Alcantara et al. 2018). Both NKX2.2 and PDX1 have been suggested to directly regulate *Nkx6.*1 expression in mice (Watada et al. 2000).

The TF NKX6.1 is expressed successive to PDX1 in cells committed to the beta cell lineage (Sander et al. 2000). Co-expression of both TFs gets later restricted to beta cells. NKX6.1 expression has been shown to be crucial for beta cell fate (Taylor et al. 2013). While ectopic expression of *Nkx6.1* drives endocrine precursors towards the beta cell lineage, they get directed to alternative endocrine lineages as a consequence of loss of *Nkx6.1* in mice (Schaffer et al. 2013).

ISL1, PAX6 and members of the RFX TF family represent further TFs required for the progressive differentiation into mature endocrine cells as well as their maintenance and survival. Expression of the LIM homeodomain protein ISL1 is initiated after NEUROG3 emergence and has been reported to be required for the generation and survival of endocrine cells as well as pancreatic islet cell proliferation (Ahlgren et al. 1997; T Guo et al. 2011). The paired homeodomain factor PAX6 is expressed in all islet cells and implicated in insulin, glucagon and somatostatin gene expression (Sander et al. 1997). *Pax6* KO mice die at birth and have been shown to contain reduced numbers of all endocrine cell subtypes (Sander et al. 1997). Studies of *Pax6* deficiency in mice and rodent beta cells suggest that PAX6 is important for the maintenance of beta cell function through direct transcriptional regulation of genes involved in insulin synthesis and secretion including *Ins1*, *Ins2* and *Pcsk1* (Ashery-Padan et al. 2004; Gosmain et al. 2012).

The RFX TF family comprises eight members characterised by a highly conserved DNA binding domain containing a winged helix motif (X-box motif) (Aftab et al. 2008; Sugiaman-Trapman et al. 2018). So far, only RFX3 and RFX6 have been implicated in islet cell differentiation, regulation of insulin secretion and maintenance of beta cell functional identity in humans and mice (Ait-Lounis et al. 2007; Ait-Lounis et al. 2010; Soyer et al. 2010; Smith et al. 2010; Taleb et al. 2011; Chandra et al. 2014; Piccand et al. 2014).

RFX6 has been shown to be essential for islet cell development in zebrafish (Soyer et al. 2010), xenopus (Pearl et al. 2011), mice (Smith et al. 2010; Soyer et al. 2010; Piccand et al. 2014) and humans (Smith et al. 2010; Pearl et al. 2011; Spiegel et al. 2011; Concepcion et al. 2014; Zhu et al. 2016). Mutations in *RFX6* are associated with the Mitchell-Riley syndrome, which is characterised by NDM, pancreatic hypoplasia and intestinal atresia among other abnormalities (Mitchell et al. 2004; Kambal et al. 2019). Loss of RFX6 in human or mouse beta cells was shown to negatively affect insulin secretion through down-regulation of *cis*-regulated genes including the glucose-sensor *GCK*, *ABCC8* and genes encoding voltage-dependent Ca^{2+} channel subunits (Chandra et al. 2014; Piccand et al. 2014). The beta cell maturation marker *Ucn3*, which represents an indicator of functional glucose sensing as well as metabolism and was reported to have a positive effect on GSIS (C Li et al. 2007; Blum et al. 2012), was down-regulated in islets of *Rfx6* KO mice (Piccand et al. 2014). While RFX6-depleted human beta cells were additionally characterised by reduced *INS* transcript levels, lower cellular insulin content, decreased insulin promoter activity and RFX6 was indicated as transcriptional regulator of human *INS* transcription (Chandra et al. 2014), in mice, loss of *Rfx6* only reduced expression of one of the two insulin transcripts (*Ins1*) and did not affect cellular insulin content (Piccand et al. 2014), suggesting that RFX6 might not be essential for insulin production in adult rodent beta cells.

The TF RFX3 has been identified as an important regulator of endocrine differentiation in mice (Ait-Lounis et al. 2007; Ait-Lounis et al. 2010). It has been shown that as a consequence of loss of RFX3, islets of *Rfx3* KO mice contained significantly less alpha and beta cells, while the number of PP cells was markedly increased and delta cells remained unchanged (Ait-Lounis et al. 2007). This was caused by a block in functional maturation of beta cells, thus *Rfx3*-depleted endocrine cells remained in an incomplete differentiated state. While displaying normal expression of beta cell TFs NKX6.1 or PDX1, expression of proteins important for beta cell function (GLUT2, GCK and insulin) were markedly reduced in *Rfx3* KO embryos (Ait-Lounis et al. 2010).

Adult beta cells/Stage 7 - beta-like cells

Adult beta cells are characterised by the expression of *MAFA*, which is absent in developing beta as well as other endocrine cells (Jennings et al. 2015). *MAFA* encodes a basic leucine zipper TF and has been shown to be critical in establishing beta cell function and maturity in islets. *Mafa* KO mice are characterised by compromised GSIS and islet architecture (Artner et al. 2010; Hang et al. 2014). In addition, glucose-responsiveness in hPSC-derived BLCs has been associated with the appearance of *MAFA*, thus current optimisation efforts focus on identifying compounds that induce expression of MAFA (Kroon et al. 2008; Rezania et al. 2014).

The unique beta cell phenotype has been suggested to be defined through expression of genes necessary for beta cell function (De Vos et al. 1995; Chimienti et al. 2006; T Gao et al. 2014; Taylor et al. 2013; Swisa et al. 2017) as well as a subset of genes specifically repressed in beta cells (Quintens et al. 2008; Pullen et al. 2013). To date more than 60 so called 'disallowed' or 'forbidden' genes have been identified in rodent islet cells by comparison of gene expression in islets relative to other tissues (Pullen et al. 2010; Thorrez et al. 2011). The core of these is formed by seven genes, identified in two independent studies (Pullen et al. 2010; Thorrez et al. 2011), including *Slc161a* (Pullen et al. 2012), *Ldha* (Sekine et al. 1994) and *Oat*, which have been implicated in metabolic-secretion coupling and insulin exocytosis as well as pro-proliferative genes *Igfbp4* (Durai et al. 2006), *Pdgfra* (H Chen et al. 2011) and *Cxcl12* (Singh et al. 2011) and *Cd302*, encoding a cell surface receptor that might be involved in actin cytoskeletal dynamics (Kato et al. 2007). Suppression of disallowed genes has been proposed to be achieved by either epigenetic silencing, i.e. through changes in chromatin structure and modifications of histones during cellular differentiation (E Li 2002), by beta cell specific transcriptional repression or through the effect of microRNAs (short non-coding RNAs, which influence mRNA stability and translation) (Selbach et al. 2008). In mice for example, RFX6 has been shown to repress beta cell disallowed genes including *Ldha*, *Slc16a1*, *Igfbp4* and *Pdgfra* (Piccand et al. 2014).

One of the more extensively studied beta cell forbidden genes is *REST*, which encodes the transcriptional repressor RE-1 (repressor element 1) silencing TF, a Krueppel-like zinc finger protein which is ubiquitously expressed except in mature neurons as well as pancreatic alpha and beta cells (Chong et al. 1995; Schoenherr et al. 1995; Atouf et al. 1997; Hohl et al. 2005; D Martin et al. 2008). In line with this, REST has recently been shown to be a key repressor of beta cell differentiation, likely functioning upstream of or in parallel to NEUROG3, blocking endocrine lineage initiation (D Martin et al. 2015). *REST* is expressed in pancreatic progenitors, but expression levels drop as cells differentiate into endocrine progenitor and endocrine cells and eventually becomes a disallowed gene in mature beta cells (D Martin et al. 2015; D Martin et al. 2017). Suppression of *REST* is likely mediated through selective Polycomb (PcG)-mediated H3K27me3, which *Rest* has been shown to acquire after the pancreatic precursor stage during beta cell development in mice and which coincides with the initiation of derepression of PcG targets allowing beta cells to activate a neural gene activity program (van Arensbergen et al. 2010). In addition, three promoters, six enhancers and two repressor regions have been identified to be involved in cell type-specific transcription of *REST*, however, exact transcriptional modulators are yet to be determined (Koenigsberger et al. 2000).

REST itself contains a cluster of eight zinc fingers that have been shown to bind to the 21bp RE-1 element (Schoenherr et al. 1995). To inhibit gene expression REST recruits co-repressor complexes that modify the chromatin structure. While the N-terminus of REST interacts with the mSin3 complex, which contains histone deacetylases (HDACs) and retinoblastoma-associated proteins (Grimes et al. 2000; Naruse et al. 1999; Roopra et al. 2000; Fleischer et al. 2003), the C-terminus recruits the CoREST complex, comprising HDACs, the chromatin-modelling enzyme BRG1 which stabilises and increases REST occupancy at RE-1 binding sites, the lysine-specific histone demethylase 1A (LSD1) and the H3K9 methylase G9a (Andrés et al. 1999; You et al. 2001; Battaglioli et al. 2002; Y Shi et al. 2003; Garriga-Canut et al. 2006; Y Shi et al. 2004; Avtar Roopra et al. 2004). Chromatin-modifying enzymes remove histone marks associated with transcriptional activation and replace them

with repression marks, thereby inhibiting or silencing gene transcription (Andrés et al. 1999; Grimes et al. 2000; Naruse et al. 1999; Roopra et al. 2000; Avtar Roopra et al. 2004; You et al. 2001). REST-mediated transcriptional repression was also suggested of being achieved through direct modulation of the transcriptional machinery. For example, interaction with TBP prevents the formation of the preinitiation complex and binding to small C-terminal domain phosphatases inhibits the activity of RNA polymerase II (Murai et al. 2004; Yeo et al. 2005).

REST target genes are essential to the differentiation and function of islet cells. They include genes encoding TFs and proteins playing a role in endocrine differentiation (*NEUROG3*, *NEUROD1*, *RFX3*, *MYT1*, *MFNG* (Svensson et al. 2009; Y Xu et al. 2006), *DNER* (Hald et al. 2012) and *CELSR3* (Cortijo et al. 2012)), genes involved in exocytosis and synaptic signalling (*SNAP25*, *SYT4*, *CHGB*, *SCG2* (Hohl et al. 2005), *NRXN1* and *NRXN2* (Mosedale et al. 2012) as well as *GRIA1*, *GRIA2* and *GRIK2* (ZY Wu et al. 2012)), genes with pro-survival beta cell activity (*MAPK8IP1* (Bonny et al. 2000; Haefliger et al. 2003), *PTPRN* (Mziaut et al. 2008) and *CDK5R2* (Tang et al. 1995; D Martin et al. 2012)) and genes crucial for cell-to-cell communication (*GJD2* (Klee et al. 2011; Hohl et al. 2005; D Martin et al. 2003)). Thus, REST has to be actively excluded from mature beta cells to ensure normal cell function.

1.4.5 Identification and characterisation of subpopulations during human pancreas differentiation

1.4.5.1 Pancreatic progenitor signatures

Characterisation of human embryonic pancreatic bud tissue and hESC-derived pancreatic progenitors has shed light on the regulatory network governing the specification of pancreatic progenitors from gut endoderm. Assessment of gene expression profiles in both progenitor models and comparison to transcriptome profiling data of 22 non-pancreatic tissues identified a set of 500 genes differentially enriched in pancreatic progenitor cells (Cebola et al. 2015). A subset of those contained genes

encoding TFs with a well-established role in pancreas development including *PDX1*, *NKX6.1*, *HNF1A*, *HNF1B*, *HNF4A*, *FOXA2*, *MNX1* and *ONECUT1*. In addition, there was an enrichment of genes involved in Wnt signalling such as *CELSR2*, *FZD2*, *WNT8B* and *VANGL2* (Cebola et al. 2015).

Human pancreatic bud tissue and hESC-derived pancreatic progenitors shared common FOXA2 binding sites and H3K4me1 enhancer marks near the identified 500 pancreas progenitor-enriched genes (Cebola et al. 2015). In addition, almost half of the identified active transcriptional enhancers in pancreatic progenitors did not overlap with those detected in adult human islets, highlighting the importance of choosing the appropriate developmental stage when investigating T2D-associated variants. Pancreatic progenitor-specific enhancers showed a substantial co-occupancy of the TFs PDX1, FOXA2, GATA6, HNF1B, ONECUT1 and TEAD1 (Cebola et al. 2015).

TEAD TFs are part of the Hippo pathway, a signalling cascade involved in pancreas development, in particular in the regulation of pancreatic progenitor cell proliferation as well as beta cell growth and survival (Ardestani et al. 2018). TEADs are recruited by the transcriptional co-activator YAP (Yes-associated protein) which is negatively regulated by the Hippo pathway, i.e. Hippo signalling-mediated phosphorylation of YAP leads to cytoplasmic sequestration and inhibition of transcriptional activity (Zhao et al. 2008; Zhao et al. 2011). While YAP has been shown to be highly expressed in pancreatic progenitors, expression was undetectable in NEUROG3-expressing endocrine progenitor cells and *YAP* was identified as disallowed gene in beta cells (George et al. 2015; Cebola et al. 2015; Pullen et al. 2010; Pullen et al. 2017). Disruption of the YAP/TEAD complex in pancreatic progenitor led to a reduction in expression of genes associated with enhancers bound by TEAD1 including *SOX9*, a gene encoding a TF important for progenitor cell growth regulation (Lynn et al. 2007; Seymour et al. 2007). Thus, it was suggested that the YAP/TEAD complex is key in activating a stage-specific gene regulatory programme in pancreatic progenitors and that down-regulation and inactivation of YAP in endocrine progenitors leads to inhibition of pancreatic progenitor-specific enhancers during further differentiation (Cebola et al. 2015).

1.4.5.2 Pancreatic progenitors, endocrine progenitors and endocrine cells are characterised by distinct gene expression profiles in humans

To improve our understanding of human endocrine pancreas development Ramond et al. used cell-specific surface markers in combination with fluorescence-activated cell sorting (FACS) to distinguish and enrich specific cell populations along the endocrine differentiation path (Ramond et al. 2017; Ramond et al. 2018).

Four distinct populations were identified in human fetal pancreata at gestational week 9. Population A was representative of multipotent pancreatic progenitors, expressing *SOX9* and *ONECUT1*, but also acinar cell markers including *CPA1* and *CPA2* (Q Zhou et al. 2007). Population B contained cells that initiated endocrine differentiation, but retained expression of pancreas progenitor markers. Population C was enriched in expression of endocrine progenitor markers including *NEUROG3* and its downstream targets such as *NEUROD1*, *NKX2.2*, *PAX6*, *RFX6* and *CELSR3*. Simultaneously, pancreas progenitor markers were down-regulated and gene expression of the hormones insulin (*INS*), glucagon (*GCG*) and somatostatin (*SST*) was initiated at low levels. Increased hormone transcript levels, as well as beta cell specific gene expression (*MAFA*, *G6PC2* and *SLC30A8*) marked population D as early endocrine cells.

Comparison of gene expression profiles between endogenous cell types and *in vitro* differentiated hPSCs revealed that early PE cells (stage 4) clustered closest to population B while end-of stage 4 (PE) and stage 5 (EP) cells clustered around population C (endocrine progenitors) (Petersen et al. 2017). Thus, subpopulation-specific gene expression patterns provide a valuable reference to assess the quality of *in vitro* hPSC-derived pancreatic cells.

1.4.6 CRISPR/Cas9 genome engineering in human PSCs

Human PSCs are highly proliferative. Their clonal nature and high genomic recombination rate allows stable genomic manipulation via site-specific nucleases including zinc finger nucleases (Bobis-Wozowicz et al. 2011), TALENs (transcription activator-like effector nucleases) (Hockemeyer et al. 2011) and CRISPR/Cas9 (Grobarczyk et al. 2015).

The genome editing CRISPR/Cas9 system capitalises on a bacterial adaptive immune system response to invading pathogens (Cong et al. 2013). CRISPR are clustered regularly interspaced short palindromic repeats, which consist of repeated DNA stretches separated by short spacer DNA sequences. These spacers are acquired from past invading bacteriophages and are characterised by a common protospacer adjacent motif (PAM) at one end (Mojica et al. 2005; Bolotin et al. 2005; Barrangou et al. 2007). Spacer sequences are transcribed into sequence-specific guide RNAs (gRNAs), which in turn direct the endonuclease Cas9 to invading phage and plasmid DNA (Brouns et al. 2008). Cleavage by Cas9 from *Streptococcus pyogenes* introduces DNA DSBs three nucleotides upstream of the PAM, which consists of the three nucleotide sequence 5'-NGG-3' (with N = any nucleotide, followed by two guanines) and is required for target site recognition (Garneau et al. 2010; Gasiunas et al. 2012; C Anders et al. 2014).

Heterologous expression of synthetic single gRNAs (sgRNAs), containing 20nt custom target sequences, and Cas9 have been successfully used for stable genomic manipulation in various cell lines including hPSCs (Grobarczyk et al. 2015; Jinek et al. 2012; Cong et al. 2013; Zhu et al. 2016; Saarimäki-Vire et al. 2017). Cas9-mediated DNA DSBs can be re-ligated through the cell's non-homologous end joining (NHEJ) repair mechanism, which can lead to the introduction of insertion/deletion (indel) mutations, and thus can be utilised to generate gene KO models. The presence of a repair template activates the homology directed repair (HDR) pathway allowing precise editing such as the exchange of a single nucleotide (Mali et al. 2013; Ran et al. 2013), opening new opportunities in disease modelling. For example, patient-specific

mutations can be introduced in healthy control hPSC lines or candidate variants can be corrected in patient-derived hiPSCs. Resulting isogenic cell line pairs share the same genetic background except for the mutation of interest, circumventing any confounding effects caused by genetic heterogeneity. Differentiation efficiencies can vary between different hPSC lines caused by heterogeneity in their genetic background (Kyttälä et al. 2016; Rouhani et al. 2014). Concerns have been expressed that this could obscure the investigation of variants which manifest their effects as changes in differentiation efficiency. The use of isogenic cell line pairs bypasses these concerns.

1.4.7 Diabetes disease modelling using human PSCs

Human PSC-derived beta cell models have not only promoted the identification of TFs crucial for human beta cell development, but have also been successfully applied in monogenic diabetes disease modelling and shed light on potential molecular mechanisms underlying pathogenesis (Hua et al. 2013; Teo et al. 2013; Stepniewski et al. 2015).

In a remarkable effort, Zhu et al. systematically analysed the role of eight TFs in human pancreas development (Zhu et al. 2016). Using an inducible Cas9 hESC line and CRISPR/Cas9 they generated TF-specific KO lines and showed a novel role for RFX6 in the formation of pancreatic progenitors likely via regulating expression of PDX1 (Zhu et al. 2016). In addition, characterisation of heterozygous PDX1 hESC-derived BLCs revealed that loss of one functional *PDX1* allele (haploinsufficiency) reduces PDX1 protein expression, compromising endocrine differentiation and raising the question whether T2D-associated mutations in or near *PDX1* might compromise beta cell development rather than affecting beta cell function (Hani et al. 1999; Macfarlane et al. 1999). Furthermore, Zhu et al. confirmed that activation of Notch signalling in pancreatic progenitors prevents endocrine differentiation and highlighted a potential divergent role for NEUROG3 in humans and mice (Zhu et al. 2016). While *Neurog3* KO mice were characterised by complete absence of

insulin-positive cells, differentiated *NEUROG3* null allele hESC lines and those carrying a patient-specific mutation, contained very few insulin-positive BLCs, which might explain why patients carrying *NEUROG3* mutations are often born without diabetic symptoms (Jiafang Wang et al. 2006).

To investigate potential disease mechanisms underlying *HNF1A*-MODY, another study used CRISPR/Cas9 to delete either one or two alleles of *HNF1A* in different hESC lines (Cardenas-Diaz et al. 2019). Differentiation into BLCs revealed increased glucagon and grehlin expression, while insulin transcript levels were down-regulated in heterozygous and homozygous *HNF1A* KO cells. This was accompanied by an increase in alpha cell-specific genes including *ARX* and a decrease in *PAX4* expression, implicating a role for HNF1A as suppressor of alpha cell fate (Cardenas-Diaz et al. 2019). In addition, GSIS was impaired in *HNF1A*-deficient BLCs, replicating defective insulin secretion observed in islets derived from a patient with *HNF1A*-MODY (Haliyur et al. 2019). Interestingly, only heterozygous *HNF1A* KO BLCs were characterised by defective mitochondrial respiration, a phenotype that was also seen in hESC-derived BLCs depleted of the human-specific lncRNA *LINC01139*, which represents a HNF1A target gene and was suggested to regulate genes important for mitochondrial function (Cardenas-Diaz et al. 2019). As a *LINC01139* homolog was not identified in the mouse genome, the authors suggested that it might contribute to species-specific phenotypes associated with loss of HNF1A in humans and mice, highlighting the value of hPSC-based disease models (Pontoglio et al. 1998; Garcia-Gonzalez et al. 2016).

Three elegant proof-of-concept studies have recently been undertaken in NDM and MODY disease modelling involving patient-derived hiPSCs. Human iPSCs were generated from a patient carrying an activating mutation in *STAT3* and step-wise directed differentiation along the endocrine lineage revealed that mutant cells up-regulated *NEUROG3* causing premature endocrine differentiation. CRISPR/Cas9-mediated correction of the *STAT3* mutation rescued the NDM disease phenotype (Saarimäki-Vire et al. 2017). A similar study derived hiPSCs from individuals carrying NDM-associated mutations in the *INS* gene. Differentiation into endocrine

cells and characterisation of BLCs revealed increased ER stress and reduced cell proliferation. Transplantation of mutant *INS* BLCs into immunodeficient mice showed accumulation of proinsulin, reduced expression of PDX1 and smaller beta cells, implicating insufficient development of a functional beta cells mass (Balboa et al. 2018). Using *HNF4A*-MODY patient-derived iPSCs in a model of liver and pancreas development, another study showed that *HNF4A* haploinsufficiency affected foregut development and subsequent gene expression profiles in hepatocyte-like and beta-like cells (Ng et al. 2019).

While these studies exemplify how stem cell-based diabetes models have been successful in identifying genes affecting beta cell differentiation, identification of T2D variants acting via regulation of insulin secretion remain challenging. As *in vitro* derived BLCs are still functionally immature (Bruin et al. 2014), phenotypes associated with impairments of metabolic regulation of insulin secretion might be obscured and not detectable. Therefore, human beta cell lines such as EndoC-βH1 remain a good complementary model when studying mature beta cell function.

1.5 GWAS identify T2D risk variants at the *SSR1/RREB1* locus

1.5.1 Discovery of *RREB1* as potential effector transcript

Multiple independently conducted GWAS have identified several independent SNPs at the *SSR1/RREB1* locus associated with either FG levels or T2D risk (Scott et al. 2012; Mahajan et al. 2014). Since those signals were non-coding, i.e. variants were located in either introns (rs17762454, FG SNP (Scott et al. 2012)) or in intergenic regions (rs9505118 and rs9502570, T2D SNPs (Mahajan et al. 2014)), there was uncertainty over the effector transcript. Exome array-based genotyping, conducted by the T2D-GENES (T2D Genetic Exploration by Next-Generation Sequencing) and GoT2D (Genetics of Type 2 Diabetes) consortia, identified a common coding variant p.S1499Y (rs35742417, MAF=21.1%, OR=1.04), associated with FG in

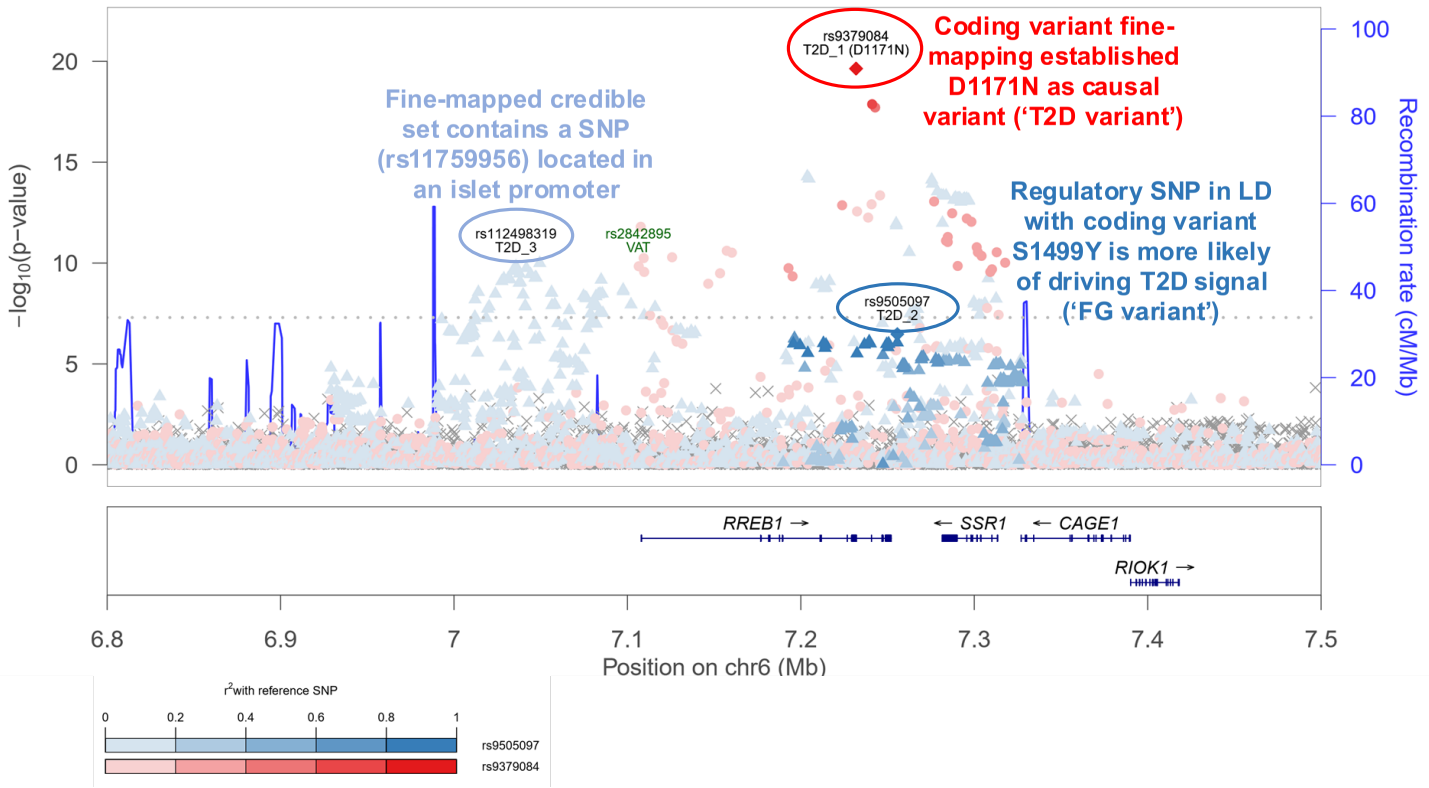


Figure 1.5: *RREB1* locus zoom plot The *RREB1* locus contains three independent signals for T2D. While coding variant fine-mapping identified rs9379084 (p.D1171N) as causal variant, establishing *RREB1* as effector transcript, rs35742417 (p.S1499Y) (represented by index SNP rs9505097) could not be confirmed as disease-causative. The FG-association signal is more likely attributable to one of the 20 fine-mapped non-coding variants in LD with p.S1499Y. The third signal represented by rs112498319 comprises a credible set of 200 variants. Epigenomic annotation revealed that rs11759956 is located in an islet promoter suggesting it as most likely causal regulatory variant. In addition, an ectopic fat distribution association was identified at the *RREB1* locus, which is represented by rs2842895 (VAT, visceral adipose tissue). Locus zoom plot was generated by Dr Anubha Mahajan and modified to highlight *RREB1* T2D-associated signals.

RREB1, highlighting this gene as a likely effector transcript at the *SSR1/RREB1* locus (Mahajan et al. 2015). Additionally, Fuchsberger et al. identified a novel signal for T2D in the form of a common coding SNP p.D1171N (rs9379084, MAF=11%, OR=1.09) in *RREB1*, encoding a non-synonymous variant in which asparagine replaces aspartic acid at position 1171 (Fuchsberger et al. 2016) (see Figure 1.2).

To date, three independent signals for T2D risk have been identified at the *SSR1/RREB1* locus, represented by rs9379084 (p.D1171N), rs9505097 (in LD with rs35742417 (p.S1499Y)) and rs112498319 (Fuchsberger et al. 2016; Mahajan et al. 2018b; Mahajan et al. 2014; Mahajan et al. 2015 and Anubha Mahajan, personal communication) (Figure 1.5). Coding variant fine-mapping revealed that p.D1171N accounted for 92% of the posterior probability of driving the T2D association at the *SSR1/RREB1* locus, establishing *RREB1* as the effector transcript (Mahajan et al. 2018b). The minor allele p.N1171 (T2D protective allele) is associated with reduced T2D risk and reduced FG levels, suggesting it likely exerts its effects through islet-cell function. Fine-mapping could not establish causality for p.S1499Y, thus it is likely that the FG association is attributable to a non-coding regulatory variant in LD with p.S1499Y (Mahajan et al. 2018b). As the credible set at this signal includes 20 variants, of which none has revealed high posterior probability of driving the T2D association, there is still uncertainty over the causal variant (Anubha Mahajan, personal communication). Fine-mapping efforts at the third independent T2D signal (rs112498319) identified a credible set of 200 variants. Integration with epigenomic annotations revealed that rs11759956 is located in an islet promoter, and therefore represents the regulatory variant with the highest probability (∼50%) of driving the T2D association at the third independent signal (Anubha Mahajan, personal communication).

1.5.2 *RREB1* is widely expressed

The *RREB1* gene is located on chromosome 6p25 (Thiagalingam et al. 1997). It encodes the zinc finger TF Ras-responsive element binding protein 1, which has been

shown to bind to Ras-responsive elements (RRE) of gene promoters to regulate gene expression (Thiagalingam et al. 1996; L Zhang et al. 1999). For example, RREB1 was first isolated as a protein bound to a RRE in the promoter region of the calcitonin gene (*CALCA*) in human medullary thyroid cancer cells (Thiagalingam et al. 1996).

Figure 1.6 gives an overview of the *RREB1* expression profile in different tissues, cell types and during distinct stages of *in vitro* beta cell development (GTEx Consortium 2015; Nica et al. 2013; van de Bunt et al. 2015; Perez-Alcantara et al. 2018). *RREB1* is expressed in diabetes-relevant tissues including adipose tissue, liver, muscle cells and most critically in human pancreatic islets (GTEx Consortium 2015; Nica et al. 2013; van de Bunt et al. 2015). Pooled human islet expression data obtained from two different centres (Oxford and Geneva) show similar median expression levels of *RREB1* (Nica et al. 2013; van de Bunt et al. 2015). While fluorescence-activated cell sorted beta cells express *RREB1* to a similar extent as pooled human islets (Nica et al. 2013), high expression levels have also been detected in other endocrine cell types (i.e. alpha and delta cells). Interestingly, brain represents the only tissue, where *RREB1* expression has not been detected. Examination of *RREB1* expression at different stages of *in vitro* differentiated hiPSCs revealed increased abundance during stages 1 (DE) to 5 (EP) compared to hiPSCs, with *RREB1* expression peaking at stage 4 (PE) (Perez-Alcantara et al. 2018) (Figure 1.6 B).

Alternative splicing of the *RREB1* gene yields three RREB1 protein isoforms, whose specific functions have not been characterised (Figure 1.7). Protein isoform 1 is encoded by a 8733bp transcript (NM_001003699.3), encoding 13 exons. It is the largest isoform with 1742aa (amino acids). Isoform 2 is encoded by two transcript variants (NM_001168344.1 and NM_001003698.3) which differ in the 5'UTR (5' untranslated region, exon 1). Compared to isoform 1, isoform 2 misses exon 11. The third protein isoform is the smallest with 1476aa due to exclusion of exon 12 (NM_001003700.1).

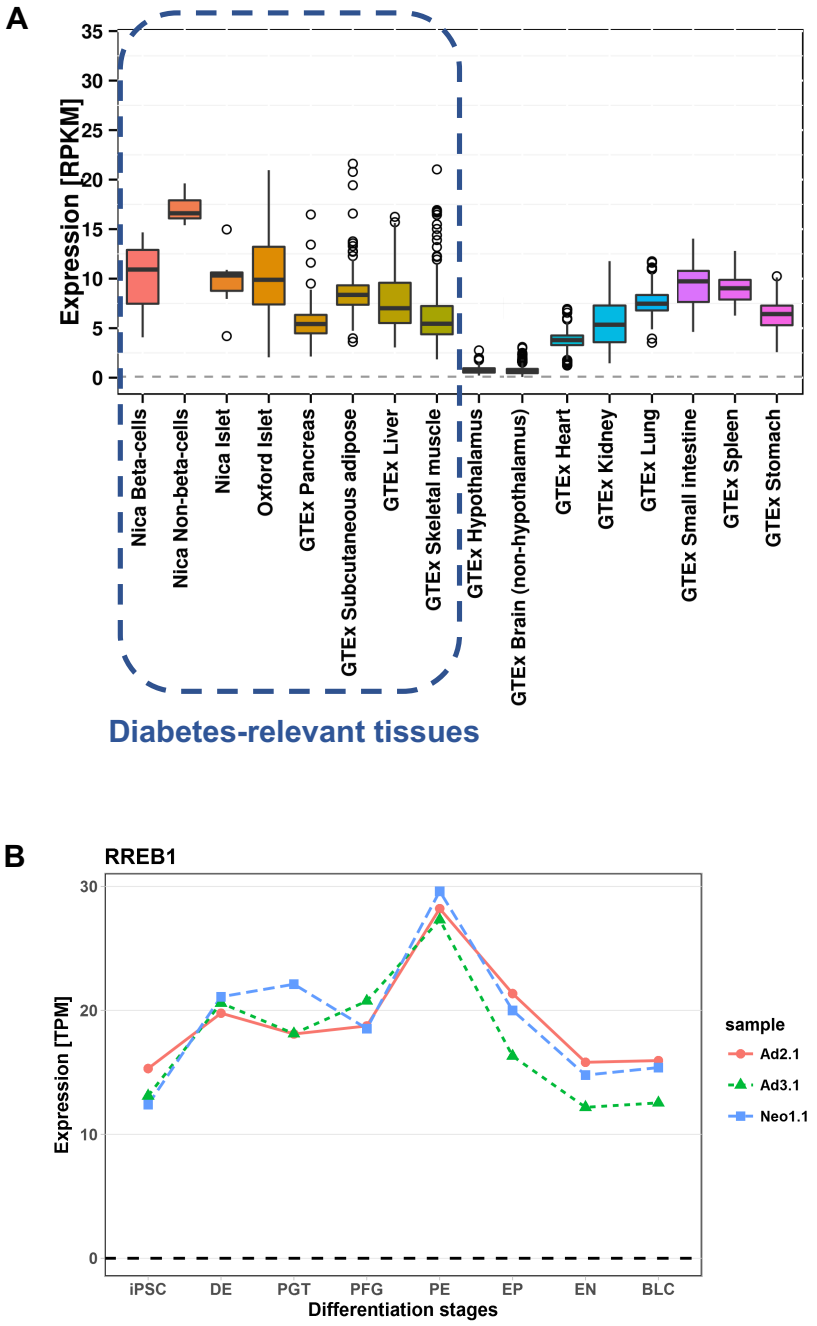


Figure 1.6: Expression profile of *RREB1* in different tissues, cell types and distinct stages of *in vitro* beta cell development (A) Expression data is shown in RPKM (Reads Per Kilobase of transcript per Million mapped reads). Box plots display median as well as 25th and 75th percentiles. Points were defined as outliers, if they were above or below 1.5 times of the interquartile range. Data was obtained from GTEx Consortium 2015; Nica et al. 2013; van de Bunt et al. 2015. (B) *RREB1* expression (in Transcripts Per Million mapped reads, TPM) at different beta cell developmental stages from three independent SB (StemBancc) hiPSC control lines (blue, green and red). Data acquired from Perez-Alcantara et al. 2018. iPSC, induced pluripotent stem cells; DE, definitive endoderm; PGT, primitive gut tube; PFG, posterior foregut; PE, pancreatic endoderm; EP, endocrine precursors; EN, endocrine-like cells; BLC, beta-like cells.

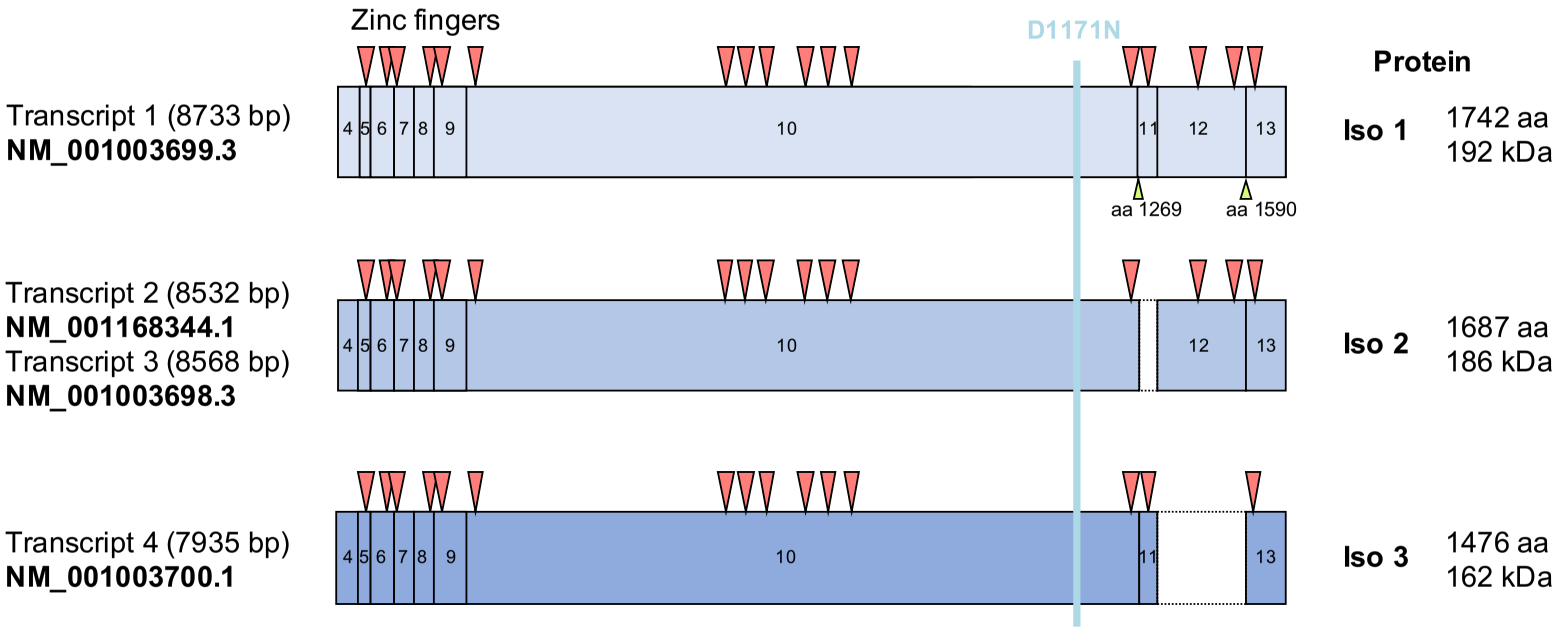


Figure 1.7: Schematic of the different *RREB1* splice variants Three RREB1 protein isoforms are encoded by four transcript variants. Exon 11 is excluded from isoform 2, exon 12 is missing in isoform 3. The *RREB1* T2D-associated variant p.D1171N is common to all three isoforms. Boxes indicate exons, triangles represent zinc fingers. bp, base pair; aa, amino acid; kDa, kilodalton.

1.5.3 The two facets of RREB1

RREB1 has been shown to both activate and repress gene expression. For example, it directly binds to and increases gene expression of *CALCA* (Thiagalingam et al. 1996), *TP53* which encodes the tumour suppressor p53 (H Liu et al. 2009), and the cholecystokinin gene (Yamane et al. 2013). Additionally, RREB1 has been shown to associate with NEUROD1 to potentiate *Sct* gene expression in hamster insulin tumour cells (Ray et al. 2003; Ray et al. 2014). *SCT* encodes the peptide hormone secretin produced by S-cells in the small intestine. It was suggested to trigger insulin release in humans (Kraegen et al. 1970; Glaser et al. 1988) and promote growth of the pancreas in rodents (Dembinski et al. 1980). Genes reported to be transcriptionally repressed by RREB1 include *CDKN2A*, which encodes the tumour suppressor p16 and has been identified as a potential T2D-associated gene (S Zhang et al. 2003; Zeggini et al. 2007; Pal et al. 2016), the human angiotensinogen gene (Date et al. 2004), the ζ-globin gene (RL Chen et al. 2010), *HLA-G* (Flajollet et al. 2009), *PSA* whose expression is co-regulated by the androgen receptor (Mukhopadhyay et al. 2007), and *SLC39A1* which encodes the zinc transporter ZIP1 in prostate cancer cells (Milon et al. 2010; Zou et al. 2011). In addition, *RREB1* has been implicated as a potential oncogene in humans (Uren et al. 2008; Starr et al. 2009); overexpression of *RREB1* has been linked to thyroid, pancreatic and colorectal cancer involving the GTPase KRAS and the miR-143/145 cluster (Thiagalingam et al. 1996; Kent et al. 2010; Costello et al. 2012; Kent et al. 2013; Kent et al. 2016; Raphael et al. 2017).

RREB1 is activated by the MAPK/ERK (mitogen-activated protein kinase/ extracellular signal-regulated kinase) pathway (Thiagalingam et al. 1996), a cascade that couples signals from cell surface receptors to TFs in the nucleus. The central component of this pathway is the small GTP-binding protein Ras, an important regulator of cell proliferation and differentiation. Activated Ras-GTP phosphorylates Raf leading to activation of MEK1/2 (MAPK/ERK kinase) which then phosphorylates ERK1/2. Activated ERK dimers can translocate into the nucleus where they phosphorylate TFs, such as RREB1, which in turn stimulate

or repress gene expression (Chang et al. 2001; Kolch 2000). The MAPK/ERK pathway has been implicated in rodent beta cell proliferation (Beith et al. 2008; Friedrichsen et al. 2006; Stewart et al. 2015) and early human beta cell differentiation, where it was shown to be required for Activitin A-induced formation of DE from hESCs (Sui et al. 2012). In addition, phosphorylated ERK1/2 has been detected in human islets from young, but not from adult human donors (H Chen et al. 2011). RREB1 protein expression has been reported to be up-regulated in PGT cells, which resemble an early stage of human beta cell development (DH Lee et al. 2012). Though the role RREB1 plays in glucose homeostasis is not well studied, it is likely that RREB1, as a downstream target of the MAPK/ERK pathway, is important in islet cell development.

Genome-wide meta-analysis of ectopic fat distribution has identified an association at the *RREB1* locus with visceral adipose tissue adjusted for body mass index (BMI) (rs2842895) (Chu et al. 2016) (Figure 1.5). While adipogenic differentiation from adipocyte progenitors isolated from either subcutaneous or visceral depots in mice was not characterised by significant transcriptional regulation of *Rreb1* (Chu et al. 2016), RREB1 was shown to play a role in brown adipogenesis (D Pan et al. 2015; Brunmeir et al. 2016). Using epigenomic annotations in combination with transcriptomic profiling during five stages of brown adipocyte (BA) development, Brunmeier et al. identified RREB1 as a novel activator of BA differentiation in mice (Brunmeir et al. 2016). *Rreb1* expression was markedly up-regulated during brown adipogenesis and comparison of the chromatin landscapes during brown and white adipogenesis highlighted RREB1 as being associated with a BA-specific super enhancer (Brunmeir et al. 2016). Overexpression of *Rreb1* in multipotent mesenchymal stem cells led to an up-regulation of BA and mitochondrial maker genes including *Cidea* and *Cox8b*, as well as enhanced differentiation efficiency and thermogenesis. LOF studies on the other hand revealed reduced differentiation efficiency and down-regulation of mitochondrial as well as BA-specific regulators (*Cidea*, *Ucp1*, *Cox8b*) as a consequence of KD of *Rreb1*, supporting RREB1 as positive regulator of BA differentiation (Brunmeir et al. 2016). In line with this,

another study reported a central role for RREB1 in regulating BA-specific gene expression in mice (D Pan et al. 2015). Pan et al. showed that BA-selective genes including *Ucp1* and *Cidea*, but not white adipocyte (WA)-selective or common fat genes, were characterised by repressive H3K27me3 marks. Removal of these marks by the demethylase JMJD3 was crucial for BA gene expression and browning of WA tissue. H3K27me3 ChIP-Seq analysis identified enrichment of the RREB1 binding motif in promoter regions of BA-selective genes. RREB1 ChIP and co-immunoprecipitation assays confirmed association of RREB1 with the *Ucp1* and *Cidea* promoters as well as with JMJD3 in BAs (D Pan et al. 2015). As KD of *Rreb1* led to decreased *Ucp1* expression and increased H3K27me3 marks at the *Ucp1* and *Cidea* promoters, it was suggested that RREB1 recruits JMJD3 to *Ucp1* and *Cidea* promoters, where JMJD3 demethylates H3K27me3 marks, required for BA-selective gene expression (D Pan et al. 2015).

1.6 Thesis outline

The role of RREB1 in glucose homeostasis is largely unexplored. Associations of *RREB1* p.D1171N with T2D and FG levels suggest that the variant likely affects beta cell function, putting the human beta cell as first choice of disease model. In addition, as the TF RREB1 has been shown to be a downstream target of the MAPK/ERK pathway, which has been implicated in beta cell differentiation, it might also be involved in beta cell development.

The overall objective of my thesis was to investigate whether RREB1 plays a role in human beta cell development and/or function (Chapters 3 and 4). Furthermore, I was interested whether the T2D-associated *RREB1* variant p.D1171N affects RREB1 transcriptional activity and whether it has an impact on beta cell development and/or function (Chapter 5).

Figure 1.8 outlines the background and aims of my thesis, presents the cell models and approaches used, and highlights which chapters address the individual objectives.

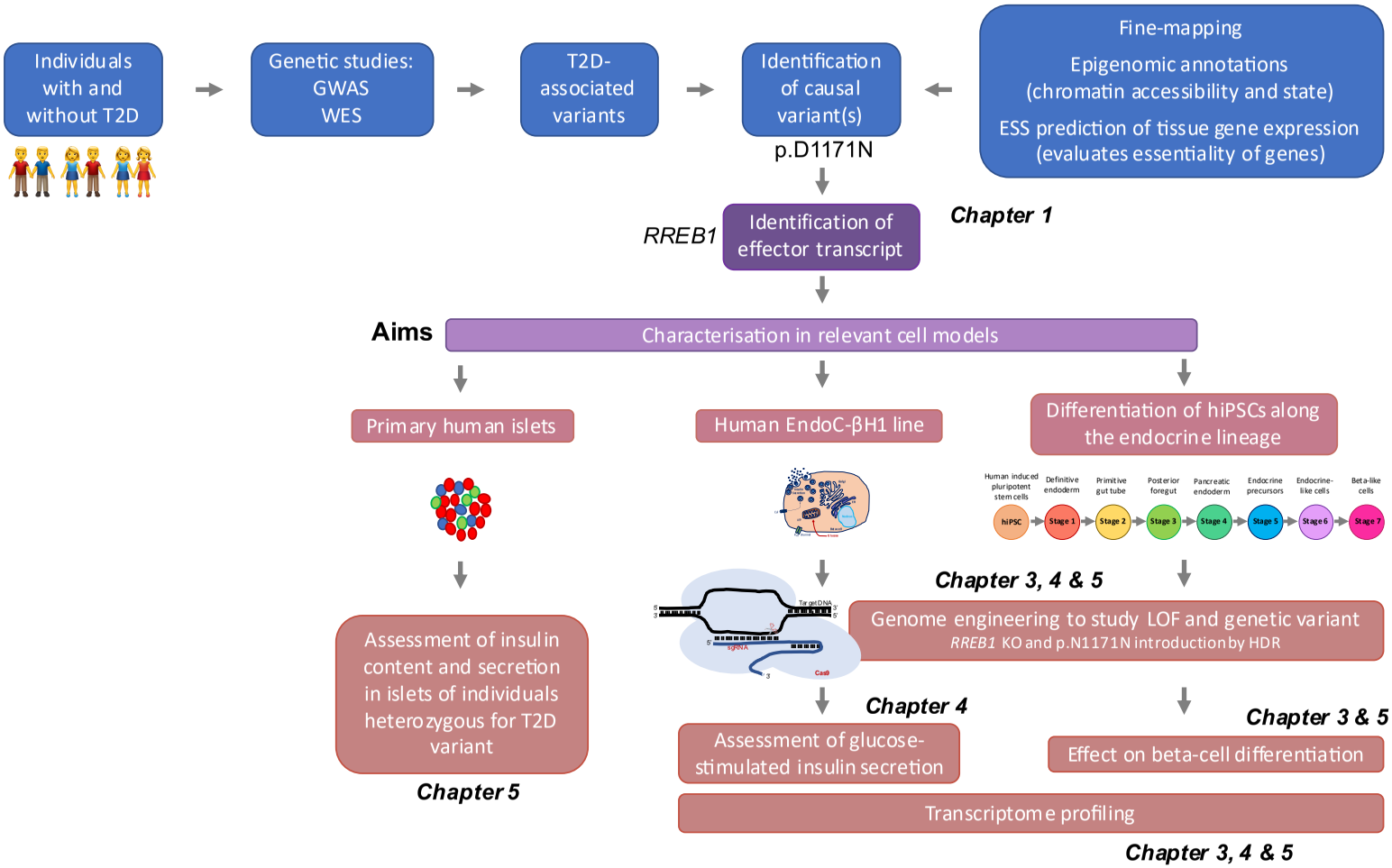


Figure 1.8: Overview of thesis aims, approaches and chapters Outline of thesis including background, aims, cell models and techniques used. Individual chapters are listed according to the results they show. GWAS, genome-wide association study; WES, whole-exome sequencing; ESS, essentiality score simulator; HDR, homology directed repair; LOF, loss-of-function.

Capitalising on recent advances in the field of genome editing I generated *RREB1* KO models in two different cellular systems. KO models and appropriate controls were characterised via cellular phenotyping and transcriptomic profiling. Deletion of *RREB1* in hiPSCs and subsequent differentiation along the endocrine lineage into BLCs enabled me to investigate, whether RREB1 plays a role in human beta cell development (Chapter 3), while KO of *RREB1* in the human EndoC-βH1 line allowed me to assess the impact of RREB1 loss on GSIS (Chapter 4).

To determine the effect of the *RREB1* T2D-associated variant on beta cell function, insulin content and GSIS were evaluated in human islets from donors heterozygous for p.D1171N. Additionally, p.N1171N was introduced into hiPSCs, edited cells were differentiated into BLCs and distinct stages of beta cell development were characterised via transcriptomic profiling (Chapter 5).

Presented results will be discussed briefly at the end of each individual chapter and more extensively in Chapter 6, in which I summarise my findings and discuss how they contribute to our understanding of the role of RREB1 in glucose homeostasis.

2

Materials and Methods

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

2.1 General methods

2.1.1 Cloning of *RREB1* expression plasmids

pSPORT-*RREB1*, which contains an in-frame N-terminal FLAG tag and encodes RREB1 protein isoform 2 (NP_001003698.1), was purchased from Addgene (41145) (Kent et al. 2013). Human *RREB1* isoform 1 cDNA (NM_001003699.3) within a pCMV6 expression vector (with an in-frame C-terminal Myc-FLAG-tag) was purchased from Origene. Short cDNA fragments specific for RREB1 protein isoforms 2 (NM_001168344.1 and NM_001003698.3) and 3 (NM_001003700.1) were subcloned into the equivalent pCMV6-*RREB1*-isoform 1 expression vector via the *SbfI* and *BstEII* restriction enzyme sites. A mutation corresponding to p.D1171N 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 D1171N_F (5'-AACAGCGGCGGGGTGAACCTGGACTCCAGCGGG-3') and D1171N_R (5'-CCCGCTGGAGTCCAGGTTCACCCCGCCGCTGTT-3'). pCMV6-Empty Vector (pCMV6-EV) served as control plasmid. pCMV6 plasmid DNA was transformed and amplified in DH5α competent cells (Thermo Fisher Scientific) and DNA was extracted using PureYield^{TM} Plasmid Mini- or Midiprep kits (Promega) following the manufacturer's instructions. Correct insertion of each *RREB1* fragment and SDM was verified by Sanger sequencing (Eurofins Genomics, Germany). Primers used for sequencing are listed in Table A.1.

2.1.1.1 DNA quantification

Plasmid DNA was quantified using the NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific), which measures absorbance at 260nm to calculate nucleic acid concentrations (dsDNA, ssDNA and RNA). Sample purity was assessed by the A260/280 and A260/230 ratios. Sample values greater than 1.8 were considered as suitable for downstream applications. Lower A260/280 values indicated protein contamination, whilst lower A260/230 values suggested salt or solvent contamination.

2.1.2 Cell culture

2.1.2.1 General maintenance of cell lines (HEK293, EndoC-βH1, hiPSC)

HEK293 and lenti-X HEK293T cells were grown in Dulbecco's modified Eagle's medium (DMEM, Sigma Aldrich, D6429) supplemented with 10% FBS (fetal bovine serum) and 100U/ml penicillin-streptomyocin (P/S, Gibco^{TM}, 15140122). Cells were dissociated using TrypLE^{TM} Select (Thermo Fisher Scientific, 12563011) and passaged twice per week in a 1:20 ratio.

EndoC-βH1 cells were maintained in cell culture flasks coated with DMEM (high glucose, Gibco^{TM}, 41965) supplemented with 1% Extracellular Matrix (Sigma Aldrich, E1270), 2μg/ml fibronectin (Sigma Aldrich, F1141) and 100U/ml P/S

and grown in DMEM (low glucose, Gibco^{TM}, 31885) supplemented with 2% bovine serum albumin (BSA, Fraction V Fatty acid free, ROCHE 10775835001), 50μM 2-mercaptoethanol (Gibco^{TM}, 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^{TM}, 25030081). Cells were split weekly (1.2 x 10^6 cells/T25 flask) using Trypsin-EDTA solution (Gibco^{TM}, 15400054).

The human induced pluripotent stem cell (hiPSC) line SB Ad3.1 was obtained from the StemBancc consortium (www.stembancc.org) via the Human Biomaterials Resource Centre, University of Birmingham (http://www.birmingham.ac.uk/facilities/hbrc). There, hiPSC were generated from human skin fibroblasts, which were obtained from a commercial source (Lonza CC-2511, tissue acquisition number 23447). The donor was a 36 year old female of European ancestry 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 previously described (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 (hESC qualified, Corning, 354277) and grown in mTeSR1 basal medium (Stemcell Technologies, 05851) supplemented with mTeSR1 5x supplement (Stemcell Technologies, 05852) and 100U/ml P/S. Cells were passaged as aggregates every 3-4 days or when confluent using ReLeSR^{TM} (Stemcell Technologies, 05872). When single cells were required, cells were released via TrypLE^{TM} Select (Thermo Fisher Scientific, 12563011) or ACCUTASE^{TM} (Stemcell Technologies, 07920).

All lines were maintained at 37 °C and 5% CO_2 and were free of mycoplasma, as proven by regular testing using the MycoAlert^{TM} Mycoplasma Detection Kit (Lonza).

2.1.2.2 Transient transfections

For transfection experiments, cells were either plated one day prior (HEK293, 400,000 cells/well on 6-well plates) or in parallel to DNA/transfection reagent preparation (EndoC-βH1 and hiPSCs) and cultured in antibiotic-free media. Briefly, FuGENE® 6 Transfection Reagent (Promega) was added to Opti-MEM reduced serum-free medium (Gibco™, 31985062), mixed and incubated for 5min before plasmid DNA was added, incubated for further 15-20min and transferred to cells (transfection reagent:DNA mixture ratios 1:3 for HEK293, 1:5 for hiPSC and 1:6 for EndoC-βH1 cells). Transfected cells were assessed for protein expression via Western blot analysis 24-96h post-transfection.

2.1.3 Protein expression studies

2.1.3.1 Cell lysis

For protein expression studies, cell pellets were collected using Trypsin-EDTA solution or TrypLE™ Select. All steps were carried out on ice or at 4 °C. For cell lysis and protein extraction cells were re-suspended in pre-chilled whole cell extraction buffer supplemented with 1mM DTT (dithiothreitol, Thermo Fisher Scientific) and 1x proteinase inhibitor cocktail (PIC, Sigma Aldrich, P8340) (Table 2.1), vortexed briefly and incubated on an end-over-end rotator for 45min. Samples were spun down at 14,000rpm for 30min in a pre-chilled benchtop centrifuge. Supernatants were transferred into 1.5ml microcentrifuges and stored at −80 °C.

2.1.3.2 Protein quantification

Protein concentrations of cell lysates were quantified using Bradford Assay Reagent (Bio-Rad Laboratories, Inc) which detects changes in the absorbance of the Coomassie Brilliant Blue G-250 dye. When stably bound to proteins this dye turns blue. Absorbance was measured at 595nm on a Versamax microplate reader using

Table 2.1: Cell lysis buffer composition for western blot analysis

Reagent	Final concentration
HEPES pH7.8	20mM
NaCl	0.42M
NP40 detergent	0.5%
Glycerol	25%
EDTA pH8	0.2mM
$MgCl_2$	1.5mM

Whole cell extraction buffer was stored at 4 °C. 1mM DTT and 1x protease inhibitor cocktail (PIC) were added fresh before cell lysis. Reagents were obtained from Sigma Aldrich.

SoftMax® Pro v5.4 software (Molecular Devices). Protein concentrations were calculated based on a BSA (Sigma Aldrich) standard curve.

2.1.3.3 Western blotting (WB)

Protein lysates were separated by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) on a 4-20% Criterion TGX Stain-Free Precast Gel (Bio-Rad Laboratories, Inc). 1x Tris/Glycine/SDS (TGS) running buffer consisted of 25mM Tris, 192mM glycine and 0.1% SDS pH 8.3 (Bio-Rad Laboratories, Inc). Samples containing 10-25μg whole-cell extract were mixed with 4x Leammli Buffer (Bio-Rad Laboratories, Inc), supplemented with 10% β-mercaptoethanol (Sigma Aldrich), and boiled at 80 °C for 10min. Denatured samples were spun down and loaded on a gel together wit 5μl of the molecular weight marker Precision Plus All Blue Standard (Bio-Rad Laboratoreis, Inc). Gels were run at 200V for 30min before activation and imaging using the ChemiDoc MP Imaging System and ImageLab 5 software (Bio-Rad Laboratories, Inc). Proteins were transferred to 0.2μm PVDF (polyvinylidene difluoride) membranes (Bio-Rad Laboratories, Inc) at 2.5A for 15min using the Trans-Blot Turbo transfer system (Bio-Rad Laboratories, Inc).

Transfer was confirmed by imaging of the membrane. Membranes were blocked in 5% milk (non-fat milk powder dissolved in phosphate buffered saline (PBS, Sigma Aldrich), supplemented with 0.1% Tween-20 (Fisher Scientific) (PBST)) at room temperature (RT) for 1h before primary antibody incubation. Primary antibodies and recommended dilutions are listed in Table A.3. Dilutions were prepared in 5% milk. Membranes were incubated in primary antibody dilutions for either 1h at RT or overnight at 4 °C with gentle shaking (Stuart see-saw rocker, Bibby Scientific Ltd). Subsequently, membranes were washed 4x for 5min each in PBST before incubation with secondary antibody with gentle agitation. Antibody details and dilutions are indicated in Table A.4. Membranes were washed again 4x for 5min each in PBST. Protein bands were detected using Clarity Western Enhanced Chemiluminescence Substrate (Bio-Rad Laboratories, Inc) and visualised on a ChemiDoc MP. Signal intensities of bands were quantified by densitometry analysis using Image Lab™ (Bio-Rad Laboratories, Inc). The intensity of protein of interest was normalised to the intensity of loading controls (tubulin or GAPDH) on the same blot. Differences in normalised protein expression were evaluated using a two-tailed unpaired t-test for two unmatched groups or a two-way analysis of variance (ANOVA) followed by Tukey's or Sidak's multiple comparisons test for the comparison of multiple groups.

2.1.4 Immunofluorescence staining

For immunostaining EndoC-βH1 or hiPSC cells were plated in pre-coated 4-chamber culture slides (BD Biosciences) and fixed with 4% paraformaldehyde (Thermo Fisher Scientific) in PBS for at least 15min at RT or overnight at 4 °C. Cells were washed with PBS twice for 5min each before permeabilisation in 0.001% Triton X-100 (Sigma Aldrich) in PBS for 5min. Cells were washed again with PBS twice for 5min each. Non-specific binding was blocked by incubating cells in swine-serum (1/20 dilution in PBS) for 30min at RT. Primary antibody incubations followed at 4 °C overnight. Antibodies and dilutions used are shown in Table A.5. The following day, cells were first washed 3x with PBS for 5min each and then incubated with Alexa Fluor-conjugated secondary antibodies for 30min at RT in the dark.

Secondary antibody details are listed in Table A.6. Cells were washed again in PBS 3x for 5min each. Finally, slides were mounted in Vectashield mounting medium (Vector Laboratories LTd), covered with a glass coverslip (VWR) and sealed using commercially available nail polish. Slides were kept at 4 °C.

2.1.4.1 Confocal microscopy

Immunostained cells were visualised on a Bio-Rad Radiance 2100 confocal microscope with a 60x 1.0 N.A. water immersion objective. The LaserSharp 2000 software was used to acquire images for three channels (green, red and far-red) consecutively. Laser settings were optimised for each channel and were kept the same between *RREB1* EV and KO cell samples. Image files were exported using the LSM Image Browser 4.2 (Carl Zeiss).

2.1.5 Flow cytometry

Human iPSCs and differentiated cells were released into a single-cell suspension, fixed with BD Cytofix™ fixation buffer or 4% paraformaldehyde (Thermo Fisher Scientific), permeabilised using BD Perm/Wash™ buffer or BD Phosflow Perm Buffer III (Biosciences) and stained for various cell surface or intracellular markers. Dead cells were excluded during FACS analysis using the LIVE/DEAD™ Fixable Violet Dead Cell Stain Kit for 405nm excitation (Thermo Fisher Scientific, L34955) according to the manufacturer's instructions. Gating was determined using isotype antibodies. Antibody details including those for isotype controls are listed in Table A.7. Samples were either run on the SH800 Cell Sorter (Sony) or the FACSCanto™ II (Biosciences).

Genome-engineered hiPSC were characterised for expression of pluripotency markers using the BD Human Pluripotent Stem Cell Transcription Factor Analysis Kit (BD Biosciences, 560589) following the manufacturer's recommendations. Markers included OCT4, NANOG, SOX2 and stage-specific embryonic antigen (SSEA-4).

In vitro differentiation efficiency was evaluated by measurement of stage-specific expression markers indicative of endocrine pancreas development via flow cytometry. Key developmental markers included: SYR-box 17 (SOX17) and C-X-C chemokine receptor type 4 (CXCR4) for DE; the TFs PDX1 and NKX6.1 for PE; NEUROD1 for EP; NKX6.1 and C-peptide for EN and BLCs.

2.1.6 RNA interference

Gene silencing was performed in EndoC-βH1 cells using small interfering RNA (siRNA) in combination with lipid-based reverse transfection. ON-TARGETplus siRNAs were purchased from Dharmacon (GE Healthcare) as pool of four oligonucleotides (SMART pool). Table A.8 indicates the details of the siRNA used. siRNAs were diluted in Opti-MEM reduced serum-free medium (Gibco^{TM}, 31985062) to a final concentration of 25nM, mixed with 0.4% Lipofectamine RNAiMAX (Thermo Fisher Scientific, 13778150) and incubated for 15-20min at RT. Transfection complexes were aliquoted into cell culture plates before gentle addition of cells, re-suspended in P/S-free growth medium. Silencing efficiency was assessed 96h later by RT-qPCR and/or WB analysis.

2.1.7 Gene expression analysis

2.1.7.1 RNA extraction

RNA was extracted from cells using the TRIzol® Reagent (Thermo Fisher Scientific) either following the manufacturer's instructions or according to the Direct-zol^{TM} RNA Mini Prep kit (Zymo Research, R2050) manual. For the former, isopropanol-mediated RNA precipitation was performed at $-20\,^{\circ}C$ overnight. Glycogen (Life Technologies, R0551) was added to increase RNA recovery efficiency. Dried RNA pellets were re-suspended in RNAse-free H_2O (Ambion, Thermo Fisher Scientific) and quantified using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific). RNA samples were stored at $-80\,^{\circ}C$.

2.1.7.2 cDNA synthesis

Complementary DNA (cDNA) was synthesised from RNA using GoScript™ Reverse Transcriptase plus a mix of oligo(dT) and random primers (GoScript™ Reverse Transcription System, Promega, A5000) according to the manufacturer's guidelines. Reactions were carried out in MicroAmp™ Optical 8-Cap Strips PCR tubes (Applied Biosystems) in a Tetrad 2 thermal cycler (Bio-Rad Laboratories, Inc). 400ng to 1μg of input RNA was used. cDNA was stored at $-20\,^{\circ}$C.

2.1.7.3 Real-time TaqMan® quantitative PCR

Gene expression was measured with TaqMan® real-time PCR assays run on the 7900HT Fast Real-Time PCR System (Thermo Fisher Scientific) using the SDS v2.3 software (Applied Biosystems). Details of the TaqMan® assays used are listed in Table A.9 and cycling conditions are summarised in Table A.10. Samples were assayed in duplicates or triplicates on 384-well plates (4titude Ltd). Non-template controls were included for each assay. Amplification curves were checked visually and threshold lines were adjusted manually where required. C_t values were converted to gene copy numbers and normalised to the geometric mean of the housekeeping genes (HKG) *TBP* (TATA-box binding protein), *PPIA* (peptidylprolyl isomerase A) and *GAPDH* (glyceraldehyde 3-phosphate dehydrogenase) or *TBP* alone. Fold change of HKG- or *TBP*-normalised gene copy number values between siRNA and siNT samples were calculated and transformed into $\log_2$ values for statistical analysis. Two-tailed unpaired t-test was used for two unmatched groups, comparison of multiple groups was done using a two-way ANOVA followed by Tukey's multiple comparisons test.

2.1.8 *TP53* genotyping

Genome-edited hiPSC lines were screened for common coding mutations in *TP53* via PCR amplification of cDNA synthesised from extracted RNA as described in

Section 2.1.7. Primer pairs and conditions used for PCR amplification and Sanger sequencing of amplified fragments are listed in Tables A.1 and A.2.

2.1.9 Insulin secretion assays in EndoC-βH1

For insulin secretion assays all incubation steps were carried out at 37 °C and 5% CO_2. First, EndoC-βH1 cells were starved overnight in cell culture medium supplemented with glucose to a final concentration of 2.8mM. The next day, cells were further incubated in 0mM glucose cell culture medium for 30min. To initiate secretion in static incubation, starvation medium was replaced with pre-warmed fresh medium containing either 1mM or 20mM glucose. For forskolin (FRSK)-mediated insulin-depletion assays, *RREB1* KO and EV EndoC-βH1 cells were incubated in cell culture media supplemented with 20mM glucose and 10μM FRSK for 30min and allowed to recover in 2.8mM glucose for a further 30min prior to assessment of GSIS. Each condition was measured in three to six wells per independent biological experiment. After 1h, supernatants were collected, spun down in a pre-chilled centrifuge at 300g for 5min and a small volume from the top was transferred to a new plate for storage at −20 °C until analysis. Adherent cells were assayed for viable cell count. This was done using the CyQUANT Direct Cell Proliferation Assay (Thermo Fisher Scientific, C35012) according to the manufacturer's instructions. Briefly, direct nucleic acid dye (0.2%) and background suppressor dye (1%) were diluted in cell culture medium and added to the cells in the dark. Cells were incubated at 37 °C for 1h before fluorescence was measured at 480nm/535nm (excitation/emission) (fluorescein isothiocyanate filter setting) on an EnSpire Multimode Plate Reader (PerkinElmer). Cell count values were expressed as fluorescent units. Subsequently, cells were washed first with cell culture medium, then with PBS, before ice-cold acid-ethanol (1.5% concentrated HCl, 75% ethanol and 23.5% deionized water) was added to extract the intracellular insulin content of the cells. Plates were sealed and covered with parafilm to stop evaporation and stored at −20 °C until analysis. The amount of insulin secreted (supernatants) and the cellular insulin contents were measured using the Insulin (human) AlphaLISA Detection Kit (PerkinElmer,

AL204C), which detects proinsulin as well as insulin. Samples were diluted in 1x AlphaLISA immunoassay buffer (supernatant 1:10, content 1:50). Analysis was performed in white 96-well 1/2 AreaPlates (PerkinElmer) on the EnSpire according to the manufacturer's recommendations. A standard curve using insulin analyte was included on every plate run. Unknown sample values were interpolated from the standard curve using a four-parameter non-linear regression of log-transformed insulin count data in Prism 8 (GraphPad Software). Insulin secretion was expressed as raw insulin released or as ratio of total secreted insulin to either total intracellular insulin content or to cell counts on a per-well basis. Replicate wells were averaged. Values were normalised to averaged basal insulin secretion of *RREB1* EV (NT) control cells for each experiment to eliminate variability in basal secretion rates across independent experiments. Statistical significance between *RREB1* KO (KD) and EV (NT) beta cells was tested using a two-tailed unpaired t-test or two-way ANOVA followed by Sidak's multiple comparisons test.

2.1.10 Insulin secretion assays in primary human islets

Human pancreatic islets were isolated, prepared and assayed at the Alberta Diabetes Institute IsletCore as previously described (Lyon et al. 2015), following the protocol published at dx.doi.org/10.17504/protocols.io.xgsfjwe.

The GSIS protocol for human islets can be found at dx.doi.org/10.17504/protocols.io.wy4ffyw. Briefly, secretion studies were performed at 37 °C in Krebs-Ringer bicarbonate-HEPES (KRBH) buffer (115 mM NaCl, 5mM KCl, 24 mM $NaHCO_3$, 2.5mM $CaCl_2$, 1mM $MgCl_2$, 10mM HEPES, and 0.1% BSA, pH7.4). Islets were washed with KRBH supplemented with 1mM glucose and pre-incubated in 1mM glucose KRBH for 1h. For each donor, triplicate groups of 15 human islets of similar size and shape were picked and incubated in 1mM glucose KRBH for 1h, after which supernatants were collected and media was replaced with 16.7mM glucose KRBH. After 1h, supernatants were collected again and total islet insulin content was extracted via acid ethanol (150ml 95% ethanol, 47ml acetic acid,

and 3ml concentrated HCl). Supernatant and content samples were stored at −20 °C until assayed for insulin using the Stellux® Chemi Human Insulin ELISA kit (ALPCO, 80-INSHU-CH10).

2.1.11 Statistical Analysis

Statistical analysis was performed in Prism 8 for macOS (v8.1.2, GraphPad Software). Results from multiple experiments are expressed as mean ±SD. Statistical tests used to determine p-values for two unmatched groups comprised either the two-tailed unpaired t-test, if values followed a Gaussian distribution, or the nonparametric Mann-Whitney test, if values did not follow a Gaussian distribution. Comparison of multiple groups was done using a two-way ANOVA followed by Sidak's or Tukey's multiple comparisons test. Results were considered statistically significant at $p<0.05$. n denotes the number of biologically independent experiments. More detailed information regarding statistical test used are provided in the individual experimental sections and figure legends.

2.2 Generation of a *RREB1* KO EndoC-βH1 cell line

Figure 2.1 provides an overview of the pipeline used to generate *RREB1* KO EndoC-βH1 cells utilising a pooled lentiviral CRISPR/Cas9 approach. Stable KO line generation was performed with the help of Antje Grotz, a fellow DPhil student. A control cell line, referred to as EV, was engineered in parallel by transducing EndoC-βH1 cells with plentiCRISPRv2 vector, encoding Cas9, but without containing a sgRNA sequence.

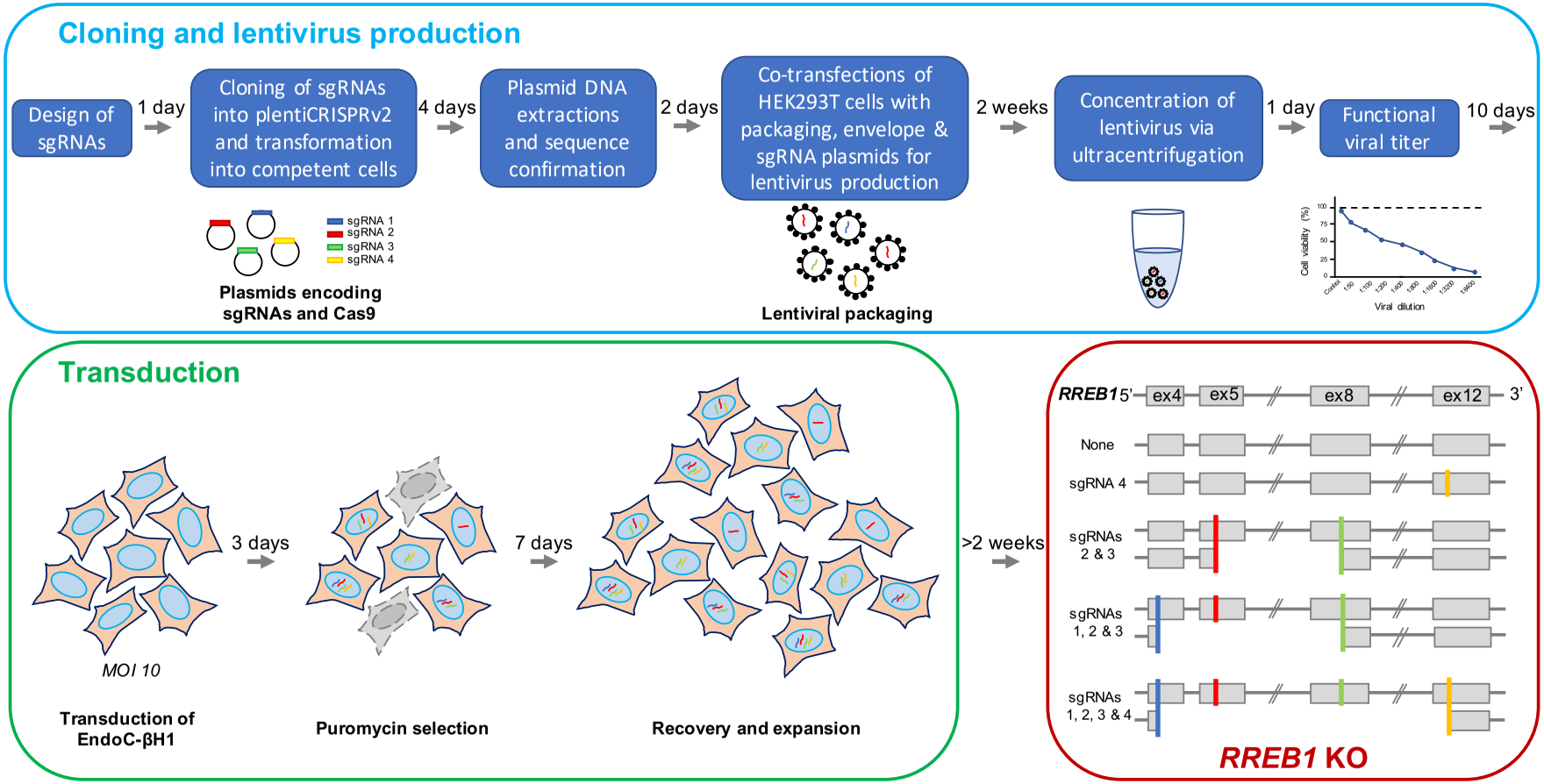


Figure 2.1: Generation of a *RREB1* KO EndoC-βH1 cell line Generation of stable *RREB1* KO and EV control EndoC-βH1 lines comprised three major steps. The first involved cloning of sgRNAs targeting *RREB1* into plentiCRISPRv2, lentivirus production in HEK293T cells and determination of the functional viral titer (lightblue box). Next, EndoC-βH1 cells were transduced and selected for successful sgRNA genome integration using puromycin (green box). Drug-resistant cells were expanded and assessed for indels in the *RREB1* gene (red box, exemplifies possible genetic deletion events). MOI, multiplicity of infection.

2.2.1 Cloning of individual sgRNAs into plentiCRISPRv2

Single gRNAs (sgRNAs) were taken from the TKO Library v3 (T Hart et al. 2017) or designed using the CRISPOR online design tool (http://crispor.tefor.net) (Haeussler et al. 2016). sgRNA sequences targeting *RREB1* exon 4 (ATGACGTCAAGTTCGCCCGC), *RREB1* exon 5 (AGTGCAAATCTTCTCACACA), *RREB1* exon 8 (GTATGGACTGGAGACCCACA) and *RREB1* exon 12 (GACAGACTCCCCCAAAAGCG) were purchased as oligonucleotides from Eurofins Genomics, annealed and sub-cloned into the *BsmBI* restriction enzyme sites in the lentiviral vector plentiCRISPRv2 (Addgene, 52961), as described previously (Shalem et al. 2014). plentiCRISPRv2 was purchased from Addgene (Sanjana et al. 2014). It encodes Cas9 and contains a puromycin selection cassette as well as a sgRNA scaffold driven by a U6 promoter. Plasmid DNA was transformed and amplified by Stbl3 competent cells (Thermo Fisher Scientific) and extracted using PureYield^{TM} Plasmid Mini- or Midiprep kits (Promega) following the manufacturer's instructions. Successful sgRNA integration was confirmed by Sanger sequencing. Primers used for plasmid DNA sequencing are listed in Table A.1.

2.2.2 Lentivirus production

For lentiviral production, lenti-X HEK293T cells were co-transfected with the lentiviral VSV-G envelope expressing plasmid pMD2.G (Addgene, 12259), the lentiviral packaging plasmid psPAX2 (Addgene, 12260) and individual cloned plentiCRISPRv2 sgRNA/Cas9-expression plasmids or a Cas9-only expression vector (empty control vector, EV) as previously described (Shalem et al. 2014). 16h after transfection media was replaced with complete culture media and 72h later supernatant containing viral particles was collected, spun down at 2,000rpm for 5min and filtered (0.45μm filter). For virus concentration, supernatant was ultracentrifuged in a swinging-bucket rotor at 29,000rpm at 4 °C for 2h and the resulting pellet was re-suspended in 1.5% BSA in PBS and aliquots were stored at −80 °C. Transfection and virus collection was performed by Antje

Grotz, ultracentrifugation was done by Elena Navarro Guerrero at the Target Discovery Institute.

2.2.3 Functional viral titer

EndoC-βH1 cells were seeded at a density of 20,000 cells per well on a pre-coated 96-well plate. Lentiviral pellets were diluted in P/S-free media (dilutions ranging from 1:50 to 1:6400) and cells were infected for 6h. For determination of transduced cells, half of the wells of each viral dilution were maintained in culture media containing 3μg/μl puromycin for seven days. Cell viability was assessed via CyQUANT Direct Cell Proliferation Assay (Thermo Fisher Scientific, C35012) according to the manufacturer's instructions. To determine the percentage of cell survival (surrogate marker of transduced cells), cell counts of EndoC-βH1 maintained in puromycin were normalised to their respective controls, cultured in normal growth media. To determine the volume of lentiviral particles needed for transductions of EndoC-βH1 cells, the correct multiplicity of infection (MOI, integrations per cell) for that particular cell line had to be defined. In general, the MOI can be estimated based on a simplified Poisson distribution (2.1).

$$P\ (n > 0) = 1 - e^{-\mathrm{MOI}} \tag{2.1}$$

- $P\ (n > 0)$ = probability that a cell gets infected by at least one viral particle
- MOI = multiplicity of infection, ratio of viral particles to cells in a well

The functional titer is defined as transduction units per μl (TU/μl) with one TU meaning one integration event in target cells. It can be calculated as

$$\mathrm{TU}/\mu\mathrm{l} = \frac{\#\text{ of cells at transduction} * \mathrm{MOI}}{\text{virus } (\mu\mathrm{l}) \text{ used for transduction}} \tag{2.2}$$

- # of cells at transduction = total number of EndoC-βH1 cells when viral particles were added
- MOI = multiplicity of infection
- virus (μl) used for transduction = volume of lentiviral stock (including dilutions) added to cells

2.2.4 Transduction of EndoC-βH1 cells

To generate a stable *RREB1* KO line, EndoC-βH1 cells were transduced at a MOI of 10. The amount of viral particle stock needed was calculated based on the functional titer (2.2). Transduced cells were selected in 4-6μg/μl puromycin for seven days. After selection cells were grown in EndoC-βH1 culture medium and passaged weekly.

2.3 CRISPR/Cas9 genome editing in hiPSCs

Figure 2.2 gives an overview of the CRISPR/Cas9 genome editing pipeline used to generate *RREB1* hiPSC models. In brief, the genome editing workflow comprised three major tasks, which can be broken down into (I) design, preparation and evaluation of sgRNAs, (II) the *RREB1* gene targeting experiment, and (III) quality control (QC) checks of correctly edited hiPSC lines. Only those that passed QC were subsequently differentiated into BLCs (see Section 2.4) and characterised via cellular phenotyping and transcriptome profiling at distinct stages of endocrine development (IV) (see Section 2.5).

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

Single gRNAs targeting *RREB1* exon 4 (GTCAAGTTCGCCCGCTGGCT) or *RREB1* exon 10 (ACCCCGCGCCAACAGCGGCG) were designed using the MIT CRISPR online design tool (http://crispr.mit.edu), purchased as oligonucleotides from Eurofins Genomics, annealed and sub-cloned into the *BsbI* restriction enzyme sites in plasmid pX330-Puro-Cas9, which encodes Cas9 and had previously been modified to also contain a puromycin selection cassette (Cong et al. 2013). Correct insertion of sgRNAs was verified by Sanger sequencing of plasmids pX330-Puro-*RREB1*-exon 4 and pX330-Puro-*RREB1*-exon 10 (Eurofins Genomics, Germany). Primers used for plasmid DNA sequencing are listed in Table A.1. Cutting efficiency for each sgRNA was evaluated in hiPSCs using TIDE (Tracking of Indels by

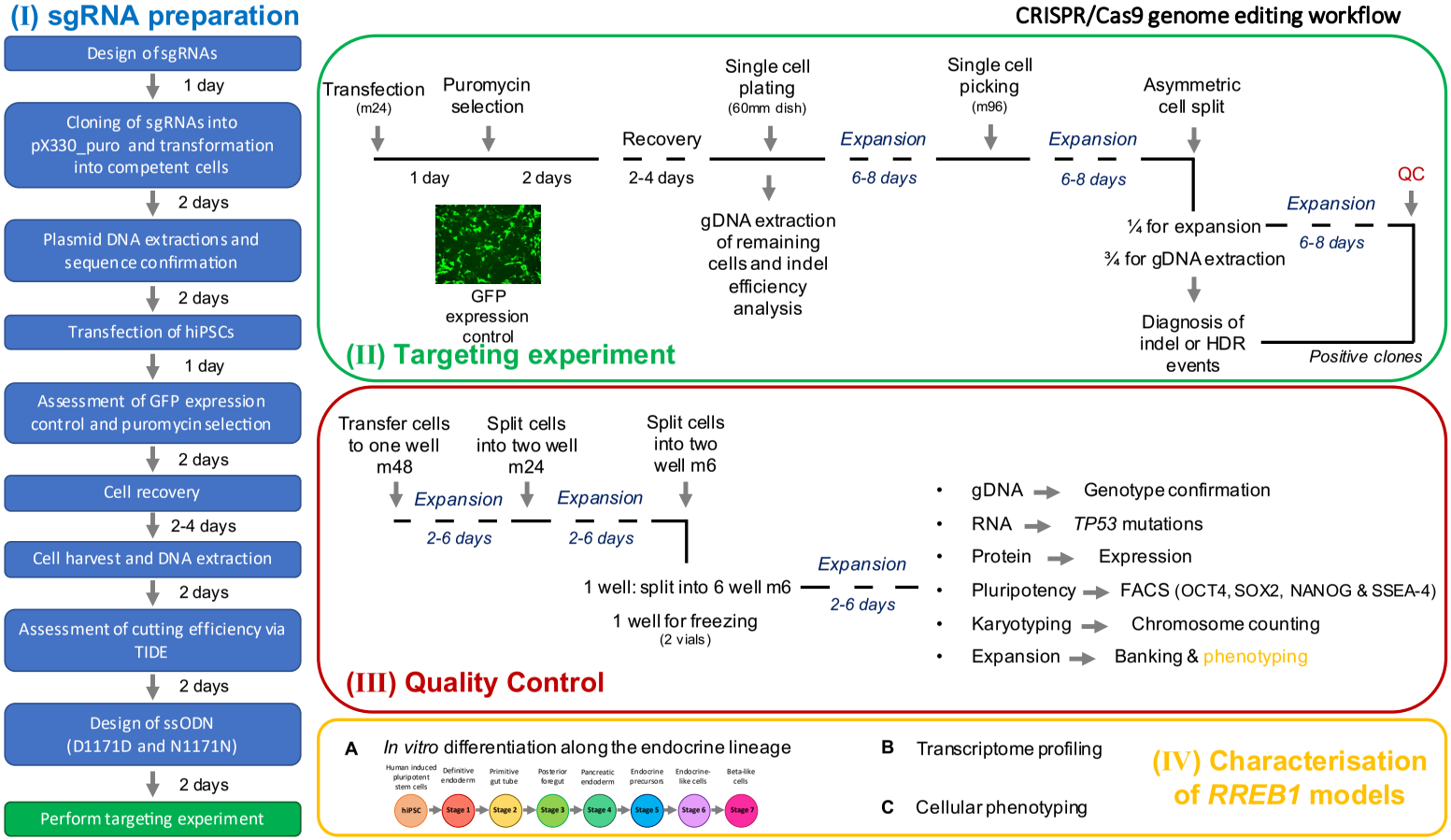


Figure 2.2: Pipeline: Generation of *RREB1* hiPSC lines Left: Workflow outlining each individual step performed to prepare and select sgRNAs targeting *RREB1*. For gene KO two sgRNAs were chosen, for gene knock-in one sgRNA and a ssODN repair template were designed and manufactured. The *RREB1* gene targeting experiment steps included lipid-based transfection of hiPSCs with Cas9 and sgRNA expression plasmids, selection of transfected cells in puromycin and genotyping of microscopy-guided picked transiently drug-resistant colonies (green box). Cell clones of the desired genotype were checked for hPSC quality control (QC) markers listed in the red box. Only *RREB1* hiPSC lines that passed QC were subsequently characterised (yellow box). The time required to complete each step is indicated between blue boxes (left) or grey arrows (right). GFP, green fluorescent protein; ssODN, single-stranded oligodeoxynucleotides; TIDE, Tracking of Indels by DEcomposition; DE, definitive endoderm; PGT, primitive gut tube; PFG, posterior foregut; PE, pancreatic endoderm; EP, endocrine precursors; EN, endocrine-like cells; BLC, beta-like cells.

DEcomposition) analysis (https://tide.deskgen.com) (Brinkman et al. 2014) which is a bioinformatics tool that quantifies the frequency of NHEJ DNA repair and identifies the most common types of indels in the DNA from a pool of cells transfected with the relevant pX330-Puro-Cas9 vector. The TIDE web tool requires two Sanger sequence traces (obtained from targeted and control PCR-amplified DNA) as well as the sequence of the sgRNA as input. 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 A.1 and A.2. Single gRNAs resulting in the highest Cas9 nuclease cutting efficiency were selected for following gene editing experiments.

2.3.2 Design of ssODN repair templates

Two 141nt single-stranded oligodeoxynucleotides (ssODN) repair templates for *RREB1* exon 10 were designed with homologous genomic arms flanking both sites of the target region containing either the T2D risk (c.3511G, p.D1171) or T2D protective (c.3511A, p.N1171) allele sequence and two silent mutations to introduce a new *Hinc*II restriction enzyme site at codon 1170 (c.3510G>C) as well as a CRISPR/Cas9 blocking mutation in the PAM sequence (c.3507G>C) (Figure 2.3). The PAGE-purified ssODNs were synthesised by Eurogentec.

2.3.3 Transfection of hiPSCs

Human iPSC were seeded at a density of 200,000 cells per well on pre-coated 24-well plates. To generate *RREB1* WT (D1171D) and N1171N lines, hiPSCs were co-transfected with pX330-Puro-*RREB1*-exon 10 plasmids and ssODNs using FuGENE® 6 (Promega). To generate *RREB1* KO lines, hiPSCs were co-transfected with pX330-Puro-*RREB1*-exon 4 and pX330-Puro-*RREB1*-exon 10 aiming to delete the genomic region between both sgRNAs. Transfected cells were grown in mTeSR1 containing 10μM ROCK inhibitor (ROCKi, Y-27632 dihyrochloride, Abcam ab120129) at 37 °C and 5% CO_2 for 24h. Media was then changed to

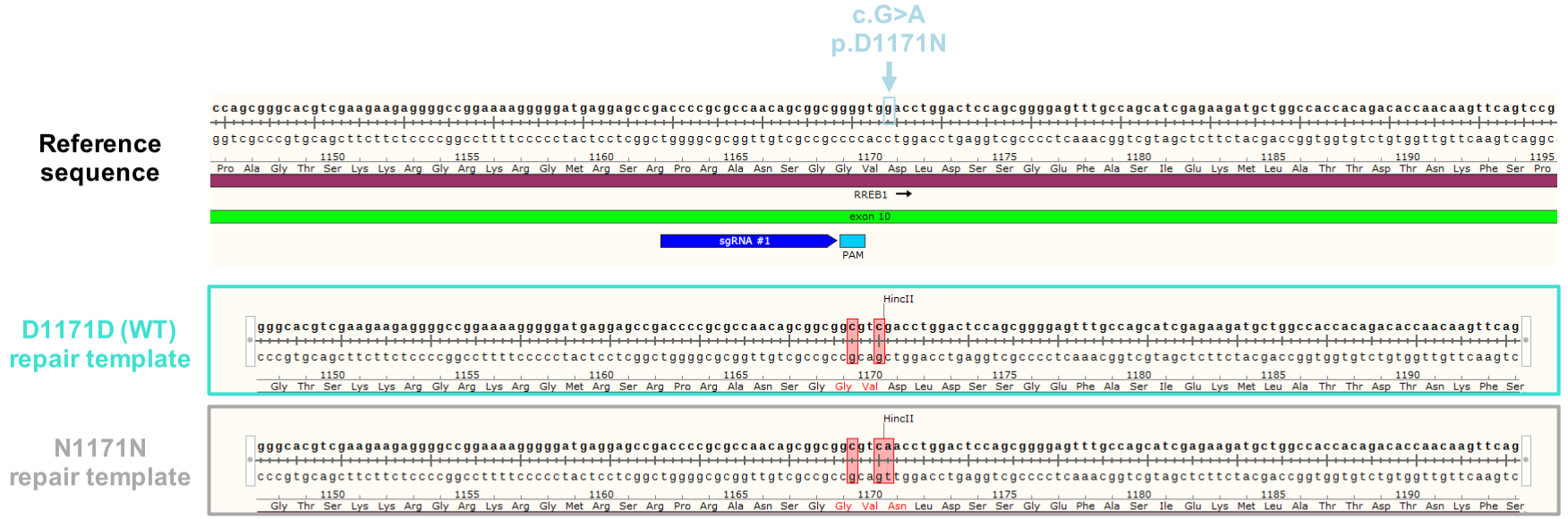


Figure 2.3: HDR design for generation of *RREB1* WT (D1171D) and N1171N hiPSC lines The 141nt single-stranded oligodeoxynucleotides (ssODN) repair templates for *RREB1* exon 10 were designed with homologous genomic arms flanking both sites of the target region containing either the T2D risk (c.3511G, p.D1171D) or protective (c.3511A, p.N1171N) allele sequence and two silent mutations to introduce a new *Hinc*II restriction enzyme site at codon 1170 (c.3510G>C) as well as a CRISPR/Cas9 blocking mutation in the PAM sequence (c.3507G>C).

mTeSR containing 400ng/ml puromycin for selection of successfully transfected cells. After 48h in selection media, cells were grown in antibiotic-free mTeSR until they reached ∼90% confluency and were re-plated at low density (2,000 cells/60mm dish). Approximately six days after expansion, single clones were picked using a microscope-mediated pipetting approach and maintained in mTeSR in individual wells of a 96-well plate for approximately seven days. They then were split into two new pre-coated 96-well plates following an asymmetrical approach, i.e. 1/4 of cells was kept for expansion while the other 3/4 of cells were used for genomic DNA (gDNA) extraction and subsequent genotyping.

2.3.4 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 at 37 °C overnight. The next day, 1/10$^{\text{th}}$ volume of 8M LiCl (Sigma Aldrich) and 1x volume of isopropanol were added per well and plates were kept on a shaker for 20min. After centrifugation at 2,500rpm for 20min at RT, supernatants were tipped off and 150μl 70% EtOH was added per well. Plates were spun again at 2,500rpm for 20min at RT, supernatants tipped off and pellets dried for ∼15min before resuspension in 40μl 0.1% TE buffer.

2.3.5 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 (*RREB1* KO) or outside of the HDR template sequence (*RREB1* WT (D1171D) and N1171N). PCR primer pairs and conditions used are listed in Tables A.1 and A.2. Agarose gel electrophoresis was used to visualise DNA fragments amplified via PCR. For HDR evaluation, PCR products were digested with *Hinc*II at 37 °C for 1h. PCR products of positive *Hinc*II-digested samples were purified using *Exo1* and rSAP and sent for sequencing. Sanger sequencing was performed with the primer pairs used

for PCR amplification. In case of uncertain sequence traces, PCR products were subcloned into pGEM®-T (Promega) according to the manufacturer's instructions. Plasmid DNA was transformed into XL-1 Blue Competent cells (Agilent, 200130), extracted using the PureYield™ Plasmid Miniprep kit (Promega) and sent for Sanger sequencing to establish the sequence of the individual alleles.

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

Genome-edited hiPSC lines were differentiated along the endocrine lineage into BLCs. For differentiation, 1.3 x 10^6 (0.8 x 10^6 for *RREB1* WT clone #3) hiPSC were seeded as single cell suspension in mTeSR1 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 replaced with fresh mTeSR1. *In vitro* differentiation was started ∼24h after plating following the protocol described in Perez-Alcantara et al. 2018. In brief, timely addition of small molecules and recombinant growth factors allowed sequential generation of cells representing seven key developmental stages of the endocrine pancreas: DE, PGT, PFG, PE, EP, EN and BLC. Tables 2.2 and 2.3 contain details on basal differentiation media composition as well as stage-specific growth factor information.

Table 2.2: Basal beta cell 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 2.3: Formulation of beta cell differentiation media including specific growth factors and small molecules added each day during differentiation

Stage	Day	Medium	Factors	Concentration	Manufacturer/Supplier	Cat #
Definitive endoderm (DE)	1	MCDB131 (Stage 1-2)	Activin A	100ng/ml	PeproTech	120-14
			CHIR 99021	3μM	Axon Medchem	1368
	1	MCDB131 (Stage 1-2)	Activin A	100ng/ml	PeproTech	120-14
			CHIR 99021	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
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
			γ-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

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.

2.5 Transcriptome profiling

2.5.1 RNA extraction, sequencing and quantification

At each distinct stage during beta cell development, cells were harvested using either TrypLE™ Select or pre-warmed ACCUTASE™, pelleted, re-suspended in TRIzol® Reagent and stored at −80 °C. RNA was extracted as described in Section 2.1.7. Library preparation and sequencing was performed at the Oxford Genomic Centre (WCHG). Polyadenylated transcripts were isolated using the NEBNext PolyA mRNA Magnetic Isolation Module (New England Biolabs, E7490). The NEBNext Ultra Directional RNA Library Kit with 12 cycles of PCR and custom 8bp indexes (New England Biolabs, E7420) was used for library preparation. All libraries were multiplexed and sequenced over several lanes of the Illumina HiSeq4000 as 75-nucleotide paired-end reads. FASTQ files were obtained from the genomics core and further processed using an in-house RNA-Seq pipeline. Briefly, reads were mapped to the human genome build hg19 using STAR v.2.5 (Dobin et al. 2013) generating BAM (Binary Alignment Map) output files. GENCODE v19 (https://www.gencodegenes.org/releases/19.html) served as transcriptome reference (Harrow et al. 2012). Gene-level quantification was performed with featureCounts from the Subread package v.1.5 (http://subread.sourceforge.net/) (Liao et al. 2014). This generated a count table with each row representing an Ensembl gene (57,820 genes) and each column a sequenced RNA library. Count values represented the unnormalised raw numbers of sequencing reads which were mapped to the respective gene in each library.

For visualisation of the beta cell differentiation RNA-Seq datasets via principal component analysis (PCA) only genes expressed at >1 count per million and detected

in at least one stage in at least one cell line were included. Counts were transformed to the logarithmic scale to base 2 ($\log_2$) and normalised in regard to library size using the concept of variance stabilising transformations (VST) of the DESeq2 package, which removes the dependencies of the variance on the mean. Correlation between VST-normalised samples was visualised via the heatmap function in R.

2.5.2 Differential gene expression analysis

Software packages used for differential gene expression analysis are described in more detail in the following paragraphs. Custom scripts were written in R version 3.3.3 (https://www.r-project.org) or using macOS bash.

2.5.2.1 RUVSeq

RUV (removal of unwanted variation) is a between-sample normalisation strategy that adjusts for technical effects (Risso et al. 2014). It is based on a generalised linear model, where unnormalised counts are regressed on known covariates (e.g. genotype and differentiation stage for the beta cell differentiation RNA-Seq datasets) and unknown variables, i.e. factors of unwanted variation, such as library preparation and batch. Factors of unwanted variation are estimated on a subset of the data. These are then adjusted for in the model for differential expression analysis. The factors of unwanted variation were estimated using centred replicate control samples for which the covariates were constant, i.e. centred counts of replicate samples (RUVs). In the case of the beta cell differentiation RNA-Seq dataset, samples of the same genotype and developmental stage, e.g. three WT DE samples, four KO DE samples, three WT PGT samples, etc. were considered as replicate groups. RUV was conducted in R using the Bioconductor package RUVSeq according to instructions in the manual compiled on May 2, 2019 (Risso 2019). Before normalisation, counts were filtered to include only genes that reached one transcript per million (TPM) in at least one cell line and in at least one stage.

2.5.2.2 DESeq2

Differential analysis of RNA-Seq count data was conducted in R using the Bioconductor package DESeq2 (Love et al. 2014) following the vignette compiled on November 30, 2016 (Love et al. 2016). A DESeqDataSet was build from the unnormalised count table (see Section 2.5.1) and a table of sample information, including genotype and developmental stage. The DESeq function first performs an estimation of size factors, which controls for differences in library sizes of individual samples, then estimates the dispersion for each gene and finally fits a negative binominal generalised linear model. Results include (I) the average of the normalised count values, divided by size factors, taken over all samples, (II) the effect size estimate ($\log_2$FC), which describes 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), (III) the standard error estimation for the $\log_2$FC estimate, (IV) a p-value, calculated via the Wald test, and (V) an adjusted p-value (q-value), which represents the p-value corrected for multiple testing using the Benjamini-Hochberg adjustment method with a false discovery rate (FDR) of $<1\%$ as significant threshold.

The beta cell differentiation RNA-Seq datasets were tested for differences between samples of different *RREB1* genotypes at each developmental stage separately, controlling for the (estimated) factors of unwanted variation (calculated via the RUVs function, see Section 2.5.2.1). The EndoC-βH1 RNA-Seq datasets were interrogated for differences between *RREB1* KO (KD) and EV (NT) samples, controlling for batch effects (pairwise preparation of individual samples).

2.5.2.3 DEXSeq

To test for differential exon usage the Bioconductor package DEXSeq was applied (S Anders et al. 2012). Changes in the relative usage of exons of the *RREB1* gene between *RREB1* KO (KD) and WT/EV (NT) samples were calculated as

$$\frac{\text{number of transcripts from } \textit{RREB1} \text{ that contain a particular exon}}{\text{number of all transcripts from } \textit{RREB1}}$$

according to the instructions in the DEXSeq vignette (Reyes et al. 2016). DEXSeq requires count reads aligned to a reference genome in the SAM (Sequence Alignment Map) format. Therefore, BAM files were first sorted by read name and then converted to SAM files using the SAMtools software package (v1.9) (H Li et al. 2009). To align the read counts to a reference genome, the DEXSeq analysis uses the Python package HTSeq and comes with two Python scripts. One of these scripts was run to translate the human gencode v19 (https://www.encodeproject.org/files/gencode.v19.annotation/) GTF (Gene Transfer Format) file into a GFF (General Feature Format) file with collapsed exon counting bins, i.e. a list of intervals where each interval corresponds to one exon. If an exonic region has different exon boundaries in different transcripts, an exon counting bin corresponds to a part of an exon, thus one exon is described by several exon counting bins. The second Python script counts the number of reads of the SAM file which overlap with each of the exon counting bins (count files). Similar to the DESeq2 package, the DEXSeq analysis includes (I) building of a DEXSeqDataSet object which consists of the count files and a table of sample information, including sample name and genotype, (II) normalisation of samples via an estimation of size factors which controls for different sequencing depths, (III) an estimate of the variability of the data in order to distinguish noise (technical and biological variation) from real effects on exon usage by calculating the dispersion estimates, and (IV) fitting a generalised linear model for each exon in each gene. Results of the analysis were visualised by plotting either the fitted expression for each exon of *RREB1* or the fitted splicing, which represents the fitted expression after subtraction of the overall changes in *RREB1* gene expression.

2.5.3 Gene ontology and transcription factor binding motif enrichment

Differentially expressed genes between *RREB1* KO (KD) and WT/EV (NT) as well as *RREB1* N1171N, D1171N and D1171D samples were tested for enrichment in gene ontology (GO) terms using g:Profiler (https://biit.cs.ut.ee/gprofiler/gost). For the beta cell differentiation datasets this was done at each individual differentiation

stage. g:Profiler's g:GOSt tool performs an over-representation analysis based on a custom input gene list (Raudvere et al. 2019). Genes are mapped to functional information in order to detect statistically significantly enriched biological terms, including GO biological processes (Ensembl) and pathways (KEGG and Reactome). All human genes annotated in the Ensembl database were used as background. Significance threshold was set to $p<0.01$ using the tailor-made g:SCS algorithm for multiple testing.

To predict upstream regulators (TFs) of differentially expressed genes the iRegulon (v1.3) Cytoscape (v3.7.0) plugin (Janky et al. 2014) was used. It maps gene regulatory networks based on enriched motifs in a set of co-expressed genes. Motif enrichment is discovered via a ranking-and-recovery method. Ranking was done for 9,713 position weight matrices (PWMs, i.e. TF motifs) and 1,120 ENCODE ChIP-Seq tracks (centred 10kb around the transcription start site (TSS)) resulting in a gene-ranking for each PWM and ChIP-Seq track. The recovery step identifies motifs/tracks which are enriched in a set of co-expressed input genes by calculating the Area Under the Curve (AUC) of the cumulative recovery curve. Motifs with an AUC higher than the average AUC of all PWMs and ChIP-Seq tracks are considered to be enriched. Enrichment is measured by a normalised enrichment score (NES). The default NES cutoff is ≥ 3, corresponding to a FDR between 3 and 9%. Enriched motifs and tracks are linked to candidate TFs that could bind to the enriched motifs via the motif2TF function. iRegulon's final output consists of a list of enriched motifs and tracks, candidate TFs and a set of direct targets for each motif/track. Gene regulatory networks can be proposed based on the predicted TF-target gene interactions.

Gene enrichment analysis was performed using a hypergeometric test.

2.5.4 ISMARA - Integrated System for Motif Activity Response Analysis

Motif Activity Response Analysis (MARA) was used to identify key TFs mediating the gene expression variation across various *RREB1* samples (Balwierz et al. 2014). FASTQ files were uploaded to a web server via the ISMARA (Integrated System for MARA) online tool (https://ismara.unibas.ch/mara/), processed for MARA, and results including identified key TFs as well as predicted details regarding their regulatory roles were downloaded.

For RNA-Seq datasets, MARA estimates and models the expression profiles of individual promoters. For this purpose a reference set of human promoters was created and a set of regulatory motifs (TF sequence specificities) was curated using the JASPAR (Mathelier et al. 2014) and TRANSFAC (Matys et al. 2006) databases. Regulatory motifs are represented as PWMs which are named after TFs annotated to bind to their site. In total, 340 human TFs were associated with 189 PWMs and are included in MARA. Next, TFBSs in the proximal promoter region, defined as the region 500bp upstream and 500bp downstream of a TSS, were predicted using the MotEvo algorithm (Arnold et al. 2012). In order to process expression data, MARA first maps raw reads to the human genome build hg19 and then associates a read with a specific transcript, if it maps entirely to an exon of the transcript. To determine the expression level for a particular promoter, a weighted average over all reads which map to transcripts associated with that promoter are calculated. As a read can map to multiple genomic loci and a single genomic locus might contain several transcripts associated with more than one promoter, a promoter's expression is based on the total number of transcripts deriving from this promoter. For motif activity fitting a promoter's expression is modelled as a linear function of the total number of predicted sites for all motifs associated with that promoter, i.e TFBSs in the proximal promoter region. ISMARA results are presented within an interactive web-interface. Output consists of a list of regulatory motifs sorted by their significance (Z-score) and their associated TFs, genome-wide

targets of the identified TFs, enriched GO categories among the targets and direct interactions between the regulators.

2.5.5 Weighted gene co-expression network analysis

For the *RREB1* KO and WT as well as the *RREB1* N1171N, D1171N and D1171D beta cell differentiation RNA-Seq datasets correlation of gene expression patterns across all developmental stages was calculated using the weighted gene co-expression network analysis (WGCNA) R software package (v.1.51) (Langfelder et al. 2008; B Zhang et al. 2005).

Raw counts were filtered to only include genes that reached one TPM in at least one cell line and in at least one stage. In addition, protein-coding genes and long intergenic non-coding RNAs (lincRNAs) were identified in the BioMart Ensembl database using the Bioconductor biomaRt package in R and retained for downstream analysis. Counts were normalised using the VST-function of the DESeq2 package.

The WGCNA package executes several steps, including (I) construction of a gene co-expression network (defined as an undirected, weighted gene network), (II) detection of modules (clusters of highly correlated genes), and (III) identification of biologically interesting modules, which can then be further interrogated for example for upstream key regulators using external software packages. In brief, first, an adjacency matrix was constructed from the gene count matrix using a soft thresholding method (power β=5 and 3, respectively) and a scale-free topology model fit. To define gene modules, the weighted topological overlap measure was applied, resulting in a topological overlap matrix, which was then used as input for unsupervised hierarchical cluster analysis. Module allocation was determined by the Dynamic Tree Cut package for R (Figure 2.4). A unique colour identifier was assigned to each module and clusters were summarised by using the module eigengene (ME). ME equals the first principal component of each module, hence represents the gene expression profiles in a module. MEs were calculated and tested for statistically significant differences between *RREB1* KO and WT or *RREB1*

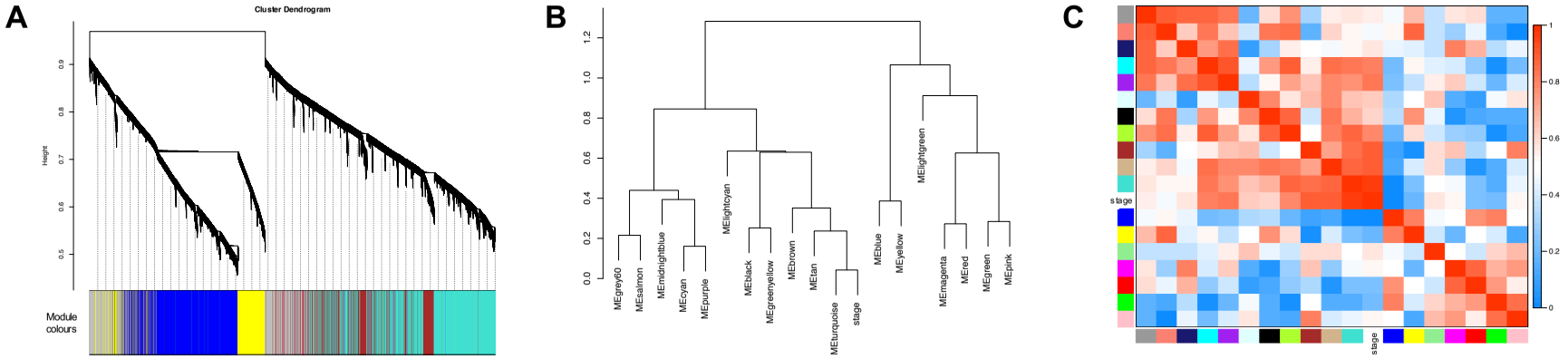


Figure 2.4: Example of modules and eigengene networks identified by WGCNA (A) Gene dendrogram received through average linkage hierarchical clustering. The module assignment was determined by the Dynamic Tree Cut method and is displayed as colour row underneath. (B) Hierarchical clustering of module eigengenes (ME) summarises the modules detected in the cluster analysis. ME are labelled by unique colours. Branches of the dendrogram (meta-modules) group together positively correlated eigengenes. (C) Heatmap plot of the adjacencies in the eigengene network including the trait 'stage'. Each row and column corresponds to one ME, labelled by the same colour as in (B) or the trait ´stage'. Blue colour in the heatmap represents low adjacency (negative correlation), red depicts high adjacency (positive correlation). The meta-modules are shown as red squares along the diagonal.

N1171N, D1171N and D1171D samples using unpaired t-tests followed by the Holm-Sidak correction method for multiple comparisons or two-way ANOVA followed by Tukey's multiple comparisons test, respectively.

Applying hierarchical clustering of adjacency-based dissimilarity to the *RREB1* KO and WT (N1171N, D1171N and D1171D) beta cell differentiation RNA-Seq datasets divided 16,876 (17,189) protein-coding genes and lincRNAs into 19 (18) co-expressed modules.

2.6 Transcription factor ChIP-sequencing

For ChIP sequencing experiments, 6 x 10^6 EndoC-βH1 *RREB1* KO cells were seeded per pre-coated 10cm dish. Transfections were carried out using either 8μg FLAG-tagged pSPORT-*RREB1* (FLAG) or pCMV6-EV (CTRL) on five individual days. A total of 10 samples (5x FLAG and 5x CTRL) were first processed individually and then pooled at the DNA purification step using the ChIP-IT High Sensitivity® kit (Active Motif, 53040) according to the manufacturer's manual with some modifications. Pooling was necessary due to a low DNA yield per individual sample. Reagents used were components of the kit unless stated otherwise.

I. Cell fixation

48h after transfection, cells were washed twice with PBS prior to fixation with 1% formaldehyde (Sigma Aldrich) in PBS for 10min at RT. The reaction was quenched with 1.25mM glycine for 5min. Fixed cells were washed once with ice-cold PBS, scrapped off and spun down at 900g for 5min at 4 °C. Cell pellets were snap frozen and stored at −80 °C.

II. Chromatin sonication and input preparation

Cell pellets were defrosted on ice, washed once in ice-cold PBS Wash Buffer, centrifuged at 1,250g for 4min at 4 °C before re-suspension in Chromatin Prep Buffer, supplemented with protease inhibitor cocktail (PIC) and PMSF. Samples were individually transferred to a chilled dounce homogeniser (Active Motif, 40401).

A tight fitting pestle (Active Motif, 40415) was used to homogenise each sample for 50 strokes. Samples were transferred back into microcentrifuge tubes and centrifuged at 1,250g for 3min at 4 °C. After removal of supernatant, cell pellets were re-suspended in 350μl ChIP Buffer supplemented with 3.5μl PIC and 3.5μl 100mM PMSF. Chromatin was fragmented using the EpiShear^{TM} Cooled Sonication Platform (Active Motif, 53080) and the following sonication conditions:

- total sonication 'on' time: 2min 30sec
- 30sec on, 30sec off
- amplitude 30%
- total pulses per sample: 20

After sonication samples were spun at maximum speed for 2min to pellet cellular debris. For analysis of shearing efficiency and chromatin quantification, 20μl of each chromatin preparation was transferred to and pooled in a new microcentrifuge tube for FLAG and CTRL, respectively (input samples).

Pooled chromatin preparations were treated with RNAse A for 30min at 4 °C. Proteinase K was added and samples were further incubated for 30min at 55 °C and then at 80 °C for an additional 2h. DNA precipitation was achieved by incubation with Precipitation Buffer, Carrier (glycogen) and absolute ethanol at −80 °C overnight. The next day, samples were spun in a microcentrifuge at maximum speed for 15min. Cell pellets were washed with 70% ethanol. Residual ethanol was removed carefully and the cell pellets were air dried for 15min at RT before re-suspension in DNA Purification Elution Buffer. Input DNA was used to analyse the shearing efficiency of the chromatin preparations using the Tape Station D1000 (Agilent) before being stored at −20 °C.

III. Chromatin immunoprecipitation

The remaining sonicated chromatin preparations were used to set up the ChIP reactions. 5μl of PIC and 5μl of blocker were mixed and incubated for 1min at RT before being added to each chromatin preparation. Finally, aFLAG polyclonal antibody (1:50, Cell Signalling, 2368) was added and samples were incubated on an

end-over-end rotator at 4 °C overnight. The next morning, MagReSyn® Protein A beads (ReSyn Biosciences, MR-PRA002) were washed twice in ChIP Buffer, added to the sample-aFLAG mix (10μl beads/chromatin preparation) and incubated end-over-end for another 3h at 4 °C. Afterwards, each sample was washed 5x with Wash Buffer AM1 on a magnetic rack. After the last wash 100μl pre-warmed (37 °C) Elution Buffer AM4 was added to each sample, which then contained the ChIP DNA.

IV. Reversal of cross-links and DNA purification

For reversal of cross-links, samples were first treated with Proteinase K for 30min at 55 °C and then at 80 °C for 2h. DNA Purification Binding Buffer and 3M Sodium Acetate (for pH adjustment) were added. All five FLAG and all five CTRL samples were pooled in the next step, respectively. pH-adjusted samples were added to DNA purification columns and spun at 14,000rpm for 1min. Columns were washed with DNA Purification Wash Buffer. DNA was eluted by adding 36μl of 37 °C DNA Purification Elution Buffer. ChIP samples were stored at −20 °C.

Input and ChIP DNA concentrations were determined on a Qubit® v2 using the Qubit® ds DNA HS Assay Kit (ThermoFisher Scientific, Q32851) according to the manufacturer's recommendations.

2.6.1 Sequencing and anlaysis of ChIP samples

Preparation of ChIP-Seq libraries and sequencing was done by the Oxford Genomics Centre at the WCHG. Input and ChIP sample libraries were multiplexed and sequenced over one lane of the Illumina NovaSeq6000 as 150-nucleotide paired-end reads to a mean read depth of 126.5±19.6 million reads per sample.

ChIP-Seq data analysis was performed by Dr Toryn Poolman (OCDEM). Peaks were called using MACS 2.1.2 (Y Zhang et al. 2008; Feng et al. 2011) with default parameters (q<0.01). The Bioconductor package ChIPseeker in R was used for ChIP peak annotation and functional enrichment analysis (Yu et al. 2015). RREB1 binding site enrichment analysis in islet regulatory regions was done by Dr Jason Torres (WCHG) using a permutation-based enrichment test.

3

The role of RREB1 in human beta cell development

3.1 Introduction

For the majority of T2D-associated genetic variants identified in GWAS, whole-genome and whole-exome sequencing studies little is known about the mechanisms underlying disease association (Thomsen et al. 2014; Morris 2018). The current challenge is to translate these genetic discoveries into disease mechanisms using disease-relevant cell types, which is particularly complex for GWAS loci that have been proposed to affect beta cell development.

While animal models have been valuable in modelling many monogenic causes of diabetes, they do not always recapitulate developmental or adult human phenotypes likely due to species differences in endocrine cell differentiation, highlighting the need for novel human cellular model systems (Sellick et al. 2004; Watanabe et al. 2009; Flanagan et al. 2014; Ryffel 2001; Allen et al. 2011; Xuan et al. 2012; Jennings et al. 2013; Nair et al. 2015). Human PSCs represent a readily available platform to model human endocrine pancreas development and mimic developmental disorders (Teo et al. 2013; Stepniewski et al. 2015). A directed differentiation protocol initially

established by D'Amour et al. and refined by several research groups, based on timely addition of recombinant growth factors and small molecules allows the formation of DE ($SOX17^+/CXCR4^+$) and subsequent differentiation into PGT ($FOXA2^+/HNF1B^+/HNF4A^+$), PFG (PDX^+) and PE ($PDX^+/NKX6.1^+/SOX9^+$), EP ($NEUROG3^+/NKX2.2^+/NEUROD1^+$) and hormone-expressing endocrine cells ($NKX6.1^+$/insulin$^+$) (D'Amour et al. 2006; Rezania et al. 2012; Bruin et al. 2013; Rezania et al. 2013; Pagliuca et al. 2014; Bruin et al. 2014; Rezania et al. 2014). Isogenic cell lines can be generated using CRISPR/Cas9, i.e. patient-specific mutations or gene KO can be introduced in healthy control hPSC lines or candidate variants can be corrected in patient-derived hiPSCs, differentiated into BLCs and interrogated for T2D disease mechanisms (Balboa et al. 2019).

The first attempt of hESC-based T2D-associated GWAS candidate gene disease modelling was carried out by Zeng at al. focusing on three of the first identified and confirmed T2D GWAS loci (Zeng et al. 2016). *CDKAL1*, *KCNJ11* and *KCNQ1*-specific KO hESC lines were generated, differentiated into BLCs and transplanted under the kidney capsule of immunodeficient mice (Zeng et al. 2016). Beta-like KO cells were characterised by defective insulin secretion *in vitro* and *in vivo*. Further interrogation of *CDKAL1*-KO BLCs revealed hypersensitivity to gluco- and lipotoxicity. Treatment of cells and mice with a newly identified compound T5224 rescued CDKAL1-specific defects through inhibition of the FOS/JUN pathway (Zeng et al. 2016). A mechanistic follow-up study using the same hESC-based platform revealed down-regulation of metallothionein genes (*MT*) and increased ER stress in *CDKAL1* KO models (M Guo et al. 2017). Forced expression of *MT1E* alleviated hypersensitivity to gluco- and lipotoxicity, exemplifying the power of hPSCs in defining the precise roles of genes associated with T2D.

3.1.1 Experimental aims

GWAS and exome sequencing efforts have identified T2D-associated genetic variants in *RREB1* (Mahajan et al. 2015; Fuchsberger et al. 2016; Mahajan et al. 2018a;

Mahajan et al. 2018b). So far, a role for RREB1 in glucose homeostasis has not been described. *RREB1* is expressed in diabetes-relevant tissues and during pancreatic endocrine cell development. In particular, RREB1 protein expression has been shown to be up-regulated in PGT cells, thus it has been suggested that RREB1 might play a transcriptional regulating role during beta cell development (DH Lee et al. 2012). In addition, RREB1 has been implicated as a downstream target of the MAPK/ERK pathway, a signalling cascade active in early stages of beta cell differentiation (Thiagalingam et al. 1996; Sui et al. 2012).

To investigate, whether RREB1 plays a role in human pancreatic endocrine cell development, isogenic *RREB1* KO and WT hiPSC lines were generated, differentiated along the endocrine lineage and characterised by cellular phenotyping and transcriptome profiling at seven distinct stages of beta cell development.

3.2 Methods

3.2.1 Genome editing

Single gRNAs targeting *RREB1* were designed using the MIT CRISPR online design tool (http://crispr.mit.edu). Oligonucleotides encoding the target sites were purchased from Eurofins Genomics and subcloned into pX330-Puro-Cas9 to produce pX330-Puro-*RREB1*-exon 4 and pX330-Puro-*RREB1*-exon 10 as described in Section 2.3. To generate *RREB1* WT hiPSC lines a ssODN repair template was designed and co-transfected with pX330-Puro-*RREB1*-exon 10. For *RREB1* KO, hiPSCs were co-transfected with pX330-Puro-*RREB1*-exon 4 and pX330-Puro-*RREB1*-exon 10. Transiently puromycin-resistant cells were genotyped via PCR amplification and Sanger sequencing (see Tables A.1 and A.2). Deletion of RREB1 was confirmed by RT-qPCR, WB analysis and IF staining as described in Sections 2.1.7, 2.1.3 and 2.1.4. Correctly edited cells were evaluated for expression of pluripotency markers via FACS (see Section 2.1.5), genomic integrity via chromosome counting, and screened for common mutations in *TP53* (see Tables A.1 and A.2).

3.2.2 Differentiation experiments

To model pancreas development, a directed seven step *in vitro* differentiation protocol was followed as described in Section 2.4.

3.2.3 Transcriptome profiling

For transcriptome profiling RNA was extracted and processed from differentiated *RREB1* KO and WT cells at seven distinct stages of *in vitro* beta cell development as described in Section 2.5.1. Software packages for RNA-Seq data analysis included the Bioconductor packages RUVSeq (Risso et al. 2014), DEXSeq (S Anders et al. 2012) and DESeq2 (Love et al. 2014) in R v3.3.3. Gene co-expression network analysis was performed using the R software package WGCNA (v.1.51) (Langfelder et al. 2008; B Zhang et al. 2005). Gene ontology and TF binding motif enrichment

analysis were done using g:Profiler (Raudvere et al. 2019) and the iRegulon (v1.3) Cytoscape (v3.7.0) plugin (Janky et al. 2014), respectively. In addition, motif activity response analysis (MARA) was applied to identify key TFs mediating gene expression variation across *RREB1* KO and WT samples using the online tool ISMARA (Integrated System for MARA) (Balwierz et al. 2014). Individual analysis methods are described in more detail in Section 2.5.

3.3 Results

3.3.1 Generation of hiPSC *RREB1* WT and KO models

To generate *RREB1* KO hiPSC lines I used two sgRNAs targeting sequences either shortly after the start codon in exon 4 or in exon 10, both common to all protein-coding *RREB1* transcripts (Figure 3.1 A). Binding of and cleavage by Cas9 at both genomic regions caused a deletion of approximately 50kb in the *RREB1* gene.

Sequencing of the SB Ad3.1 hiPSC line revealed that it was heterozygous for the T2D-associated variant rs9379084 (c.3511G>A, p.D1171N) in *RREB1*. Thus, in parallel with generation of the *RREB1* KO lines, the SB Ad3.1 line was also genetically edited to be homozygous for the T2D risk allele at rs9379084 (c.3511G, p.D1171D), referred to as *RREB1* WT or *RREB1* D1171D (in Chapter 5) lines. To activate the HDR pathway, allowing precise replacement mutations, a ssODN repair template was provided with the sgRNA targeting exon 10 (see Figure 2.3). Besides the desired allele (c.3511G), the repair template contained two 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 (*Hinc*II) for future evaluation of CRISPR clones (Figure 3.1 B).

3.3.1.1 Genotyping of edited *RREB1* hiPSC lines

RREB1 WT hiPSC lines

Clones showing ssODN-directed repair were identified via PCR amplification of a 447bp fragment (using primer pair exon 10 shown in Figure 3.2 B) and subsequent restriction enzyme digest with *Hinc*II of amplified fragments. Unedited alleles did not contain a recognition site for *Hinc*II and were identified as an uncut 447bp fragment ($HDR^{-/-}$). Clones characterised by ssODN-directed repair on one allele revealed a pattern of three bands (447bp, 341bp and 106bp) in agarose gel electrophoresis ($HDR^{+/-}$), and clones showing ssODN-directed repair on both alleles displayed fragments of 341bp and 106bp ($HDR^{+/+}$) (Figure 3.2 A). PCR-amplified

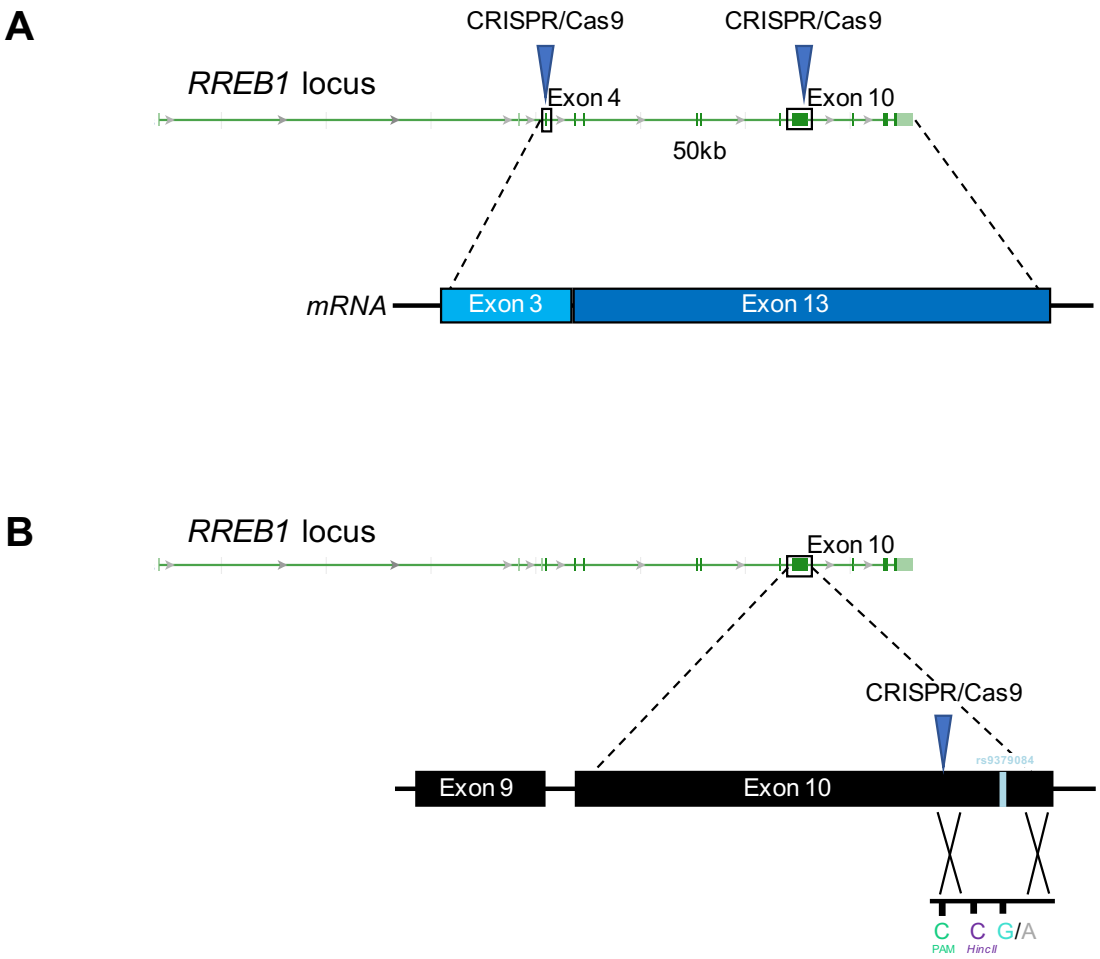


Figure 3.1: Genome editing strategies to generate *RREB1* KO and WT hiPSC lines (A) Schema of *RREB1* highlighting the locations of the two sgRNAs used to generate *RREB1* KO hiPSC lines. Cleavage of Cas9 at both locations led to a deletion of around 50kb of the genomic region in the *RREB1* gene. (B) Schema of *RREB1* highlighting rs9379084 (lightblue). The SB Ad3.1 hiPSC line, heterozygous for rs9379084, was genetically engineered to be either homozygous for the T2D risk (p.D1171) or the T2D protective (p.N1171) allele. PAM, protospacer adjacent motif.

fragments of $HDR^{+/-}$ and $HDR^{+/+}$ clones were Sanger sequenced to verify their genotype (c.3511G). In summary, 12 of 192 hiPSC clones picked revealed correct ssODN-directed repair, accounting for a gene editing efficiency of 6.3% (Table 3.2).

RREB1 KO hiPSC lines

To verify the generation of KO alleles, edited lines were genotyped by PCR and assayed for NHEJ events (indels) by Sanger sequencing. Small indels and non-edited alleles were confirmed by PCR amplification and sequencing of the regions around

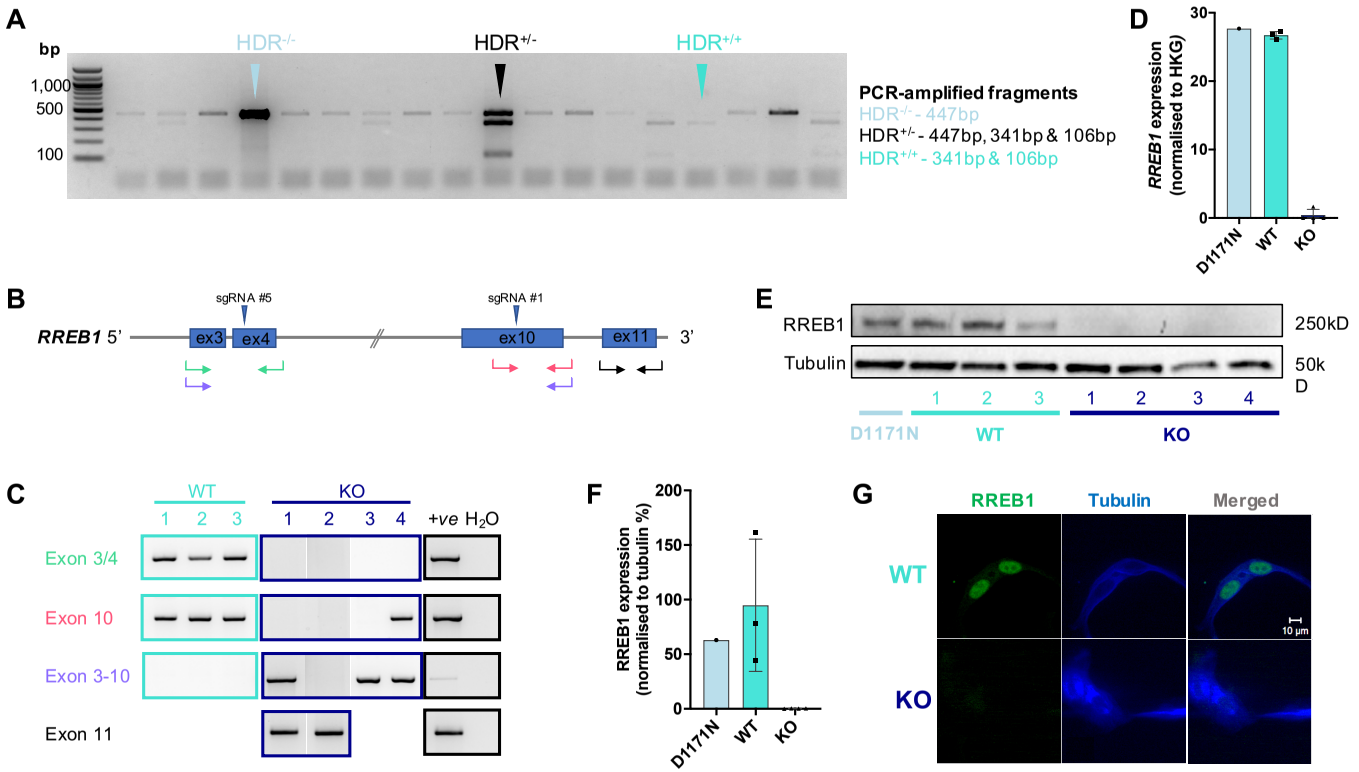


Figure 3.2: Genotyping and determination of RREB1 expression in *RREB1* WT and KO hiPSC lines (A) Example agarose gel after PCR amplification and *Hinc*II restriction enzyme digest of amplified fragments to assess ssODN-directed repair in *RREB1* WT hiPSC clones. (B) Genotyping strategy used to confirm correctly targeted *RREB1* WT and KO clones. Arrows represent primers and indicate their location in the genome. Primer pairs are colour-coded. (C) Gel electrophoresis of PCR-amplified fragments. The exon 3/4 and exon 10 primer pairs were used in independent reactions to detect small deletions, unedited and WT alleles. The purple primer pair was used to confirm deletion of a ~50kb genomic region in the KO alleles. Alleles were Sanger sequenced with the respective primer pairs used for genotyping. (D) *RREB1* mRNA expression in unedited SB Ad3.1 (D1171N), *RREB1* WT and KO hiPSC clones. (E) Western blot analysis for validation of loss of RREB1 protein (~250kD) in hiPSC KO lines. Tubulin (50kD) was used as protein loading control. (F) Quantification of RREB1 protein showed no difference between expression levels in the unedited (D1171N) and edited *RREB1* WT hiPSC clones. (G) Immunofluorescence staining of RREB1 (green) and tubulin (blue, to outline the cell shape). HDR, homology directed repair; +ve, positive control; D1171N n=1, WT n=3, KO n=4; values are displayed as mean±SD; D1171N (unedited SB Ad3.1 clone), lightblue ; WT, turquoise; KO, darkblue.

the binding sites of the sgRNAs targeting exon 4 (green primer pair) and exon 10 (pink primer pair), respectively. KO alleles containing large deletions were identified using a primer pair, with the forward primer binding upstream of the exon 4 sgRNA binding region and the reverse primer binding downstream of the exon 10 sgRNA location (purple primer pair) (Figure 3.2 B). In total, six of 192 picked colonies (3.1%) were characterised by biallelic deletions in *RREB1* (Table 3.2).

Below, the genotyping strategy is described in detail for four selected *RREB1* KO hiPSC lines. Their individual genotypes are summarised in Table 3.1. The expected ∼50kb deletion in correctly targeted *RREB1* KO lines was observed in three of the four KO clones, displaying an amplification product using the purple PCR primer pair (exon 3-10) (Figure 3.2 C). In comparison, none of three selected *RREB1* WT hiPSC clones were positive for exon 3-10 genotyping PCR reactions. Sanger sequencing confirmed a genomic deletion of ∼50kb, but only for one allele in each of the three *RREB1* KO hiPSC clones.

Gel electrophoresis revealed that none of the KO clones showed an amplification product for exon 3/4 PCR (green), however, KO clone #4 displayed a band when using exon 10 primer pair (pink), whose size was similar to the amplification products detected in the three selected *RREB1* WT clones (Figure 3.2 C). Amplification and Sanger sequencing of a larger fragment comprising exon 4 revealed a frame-shift deletion of 724bp around the translation start site and a single nucleotide deletion in exon 10 for one allele of *RREB1* KO clone #4.

Interestingly, KO clone #2 did not show a PCR-amplified fragment for any of the primer pair combinations used. To test whether the lack of amplification product was caused by insufficient or poor quality DNA, a control region (exon 11 of *RREB1*) was PCR amplified. Gel electrophoresis and subsequent Sanger sequencing of the amplified fragments excluded lack of DNA in the sample (Figure 3.2 C). The deletion caused by CRISPR/Cas9 highly likely affected one or both of the primer binding sites preventing PCR amplification. Indeed, amplification and sequencing of KO clone #2 cDNA revealed complete deletion of exons 4 to 12 (data not shown).

Similar to KO clone #2, the deletion on the second allele of clones #1 and 3 probably compromised PCR primer binding site impeding amplification.

Table 3.1: *RREB1* KO hiPSC clones genotyping

Clone #	Allele 1	Allele 2
1	unknown	∼50kb deletion
2	unknown	unknown
3	∼50kb deletion	unknown
4	∼50kb deletion	724bp deletion around start codon, 1nt deletion in exon 10

Allele 1, c.3511G (T2D risk allele, p.D1171); Allele 2, c.3511A (T2D protective allele, p.N1171)

3.3.1.2 Evaluation of *RREB1* mRNA and RREB1 protein expression in edited *RREB1* hiPSC lines

RREB1 mRNA expression was determined using a Taqman® probe spanning the boundary between exon 9 and 10, a genomic region expected to be deleted in KO alleles. *RREB1* expression levels in WT clones were similar to those observed in a SB Ad3.1 hiPSC clone, which had passed through the CRISPR/Cas9 pipeline, but remained unedited (referred to as *RREB1* D1171N) (Figure 3.2 D). Whilst in three of the four *RREB1* KO lines, *RREB1* transcripts were undetectable, KO clone #4 showed only a marked reduction in *RREB1* expression (Figure 3.2 D). Detectable *RREB1* mRNA expression in *RREB1* KO clone #4 was likely attributable to allele 2 (Table 3.1). This suggests that despite deletion of the translation start site resulting in a frame-shift and premature stop codon, *RREB1* mRNA did not get degraded by the cell's nonsense-mediated decay pathway. Complete loss of protein expression, however, was confirmed for all four *RREB1* KO hiPSC lines by WB analysis (Figure 3.2 E and F) and IF staining (Figure 3.2 G). RREB1 protein expression levels were quite variable in the WT hiPSC lines, but did not seem to be affected by genome engineering, when compared to those observed in the

RREB1 D1171N hiPSC clone (Figure 3.2 F). WT RREB1 was found in the nucleus of hiPSCs, which is in concordance with previous subcellular localisation studies in different cellular systems (Fujimoto-Nishiyama et al. 1997) (Figure 3.2 G).

3.3.2 Characterisation of *RREB1* WT and KO hiPSC models for typical hPSC properties

To assess whether genome editing had an impact on integrity of hiPSCs, correctly targeted *RREB1* WT and KO clones were characterised for typical hPSC properties (see Section 1.4.3.1) (Table 3.2).

Table 3.2: QC overview of edited *RREB1* WT and KO hiPSC clones

	# Clones				
	Picked	**Correct genotype**	**Without *TP53* mutations**	**Pluripotent (OCT4, NANOG, SOX2 & SSEA-4)**	**Diploid (46 chr/cell)**
WT	192	12 (6.3%)	12	11	4 (2.1%)
KO	192	KO/KO: 6 (3.1%) KO/WT: 0	6	6	4 (2.1%)

In the undifferentiated state, all *RREB1* WT and KO clonal lines showed a typical hiPSC colony morphology. Expression of pluripotency markers OCT4, SOX2, NANOG and SSEA-4 was confirmed via FACS (Figure 3.3 A-C). One *RREB1* WT line displayed reduced pluripotency marker expression (only 50% of cells were positive for all four markers), indicating spontaneous differentiation and hence was excluded from further characterisation. On average 94.9±4.8% cells of the three selected WT lines and 87.6±11.3% cells of the four selected KO clones were quadruple-positive for OCT4, SOX2, NANOG and SSEA-4 (Figure 3.3 A).

To monitor genomic stability of edited hiPSCs, chromosome counting of DAPI-stained metaphases was performed by Dr Daniela Moralli at the Chromosome Dynamics Core (WCHG). This method can detect chromosomal rearrangements including aneuploidy (absence or addition of a chromosome), but also more subtle

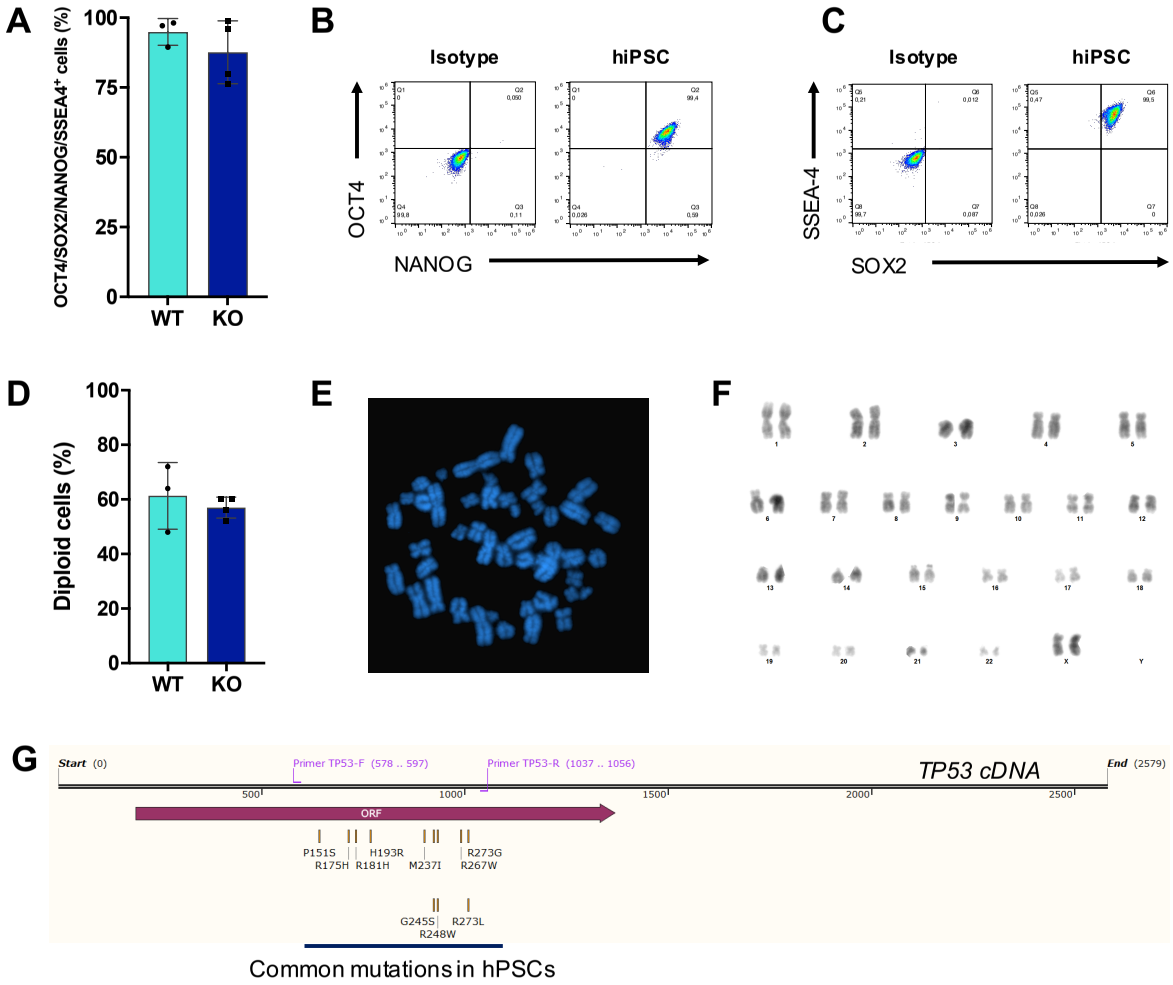


Figure 3.3: Characterisation of *RREB1* hiPSC lines for typical stem cell properties (A) Edited hiPSC lines were assessed for expression of common pluripotency markers including OCT4 and NANOG (B) as well as SSEA-4 and SOX2 (C). (D) Karyotyping was done via chromosome counting of DAPI-stained metaphases (blue, E). An example tentative karyogram in inverted DAPI (black and white) is shown (F). (G) Schema showing the locations of common coding mutations in the *TP53* gene in hPSC lines. WT n=3, KO n=4; values are displayed as mean±SD; WT, turquoise; KO, darkblue.

deletions or insertions of a few megabases and translocations (Trask 2002). A cell line was defined as having passed karyotyping if the median count was 46 chromosomes per cell and no metaphase contained more than 46 chromosomes. Lower chromosome counts were highly likely to reflect technical issues, for example loss of a chromosome during wash steps. Based on this definition, only four of 12 *RREB1* WT and four of six KO clones were characterised by a diploid karyotype (Figure 3.3 D-F). For the three selected WT clones 61.3±12.2% of the counted metaphases were characterised by 46 chromosomes/cell. DAPI-stained chromosome counting of the four selected KO lines revealed a mean of 57±3.8% diploid cells (Figure 3.3 D).

Screening of genome-edited *RREB1* KO and WT lines for the most common coding mutations in *TP53* revealed that none of the cell lines had acquired such a mutation (Figure 3.3 G).

In summary, for both the *RREB1* WT and the *RREB1* KO hiPSC clones, only four of the originally picked 192 clones (2.1%) were finally characterised by the desired genotype and passed hPSC QC checks. While none of the clones had acquired any of the 10 analysed common coding mutations in *TP53* frequently detected in hPSCs, and only one of the newly engineered lines displayed spontaneous differentiation, the majority of the WT (66.7%) and KO (33.3%) clones were unusable due to detection of aneuploidy, i.e. duplication of a chromosome, in a small percentage of cells in individual lines. Thus, obtaining cells post-editing with a normal karyotype represented the biggest bottleneck in generating *RREB1* hiPSC lines.

To investigate a potential role for RREB1 in beta cell development the previously selected three WT and four *RREB1* KO hiPSC clones were taken forward for *in vitro* beta cell differentiations.

3.3.3 Differentiation of *RREB1* WT and KO hiPSC models into pancreatic beta-like cells and cellular phenotyping at seven discrete stages of development

3.3.3.1 Evaluation of *RREB1* mRNA and RREB1 protein expression in differentiated *RREB1* WT and KO lines

To model pancreatic endocrine development a directed step-wise protocol was followed (see Sections 1.4.4.1 and 2.4). RNA and protein lysates were collected at all seven distinct stages of beta cell development. *RREB1* mRNA and RREB1 protein were detected in WT cells, but not in compound heterozygous KO cells, as assessed by RT-qPCR (KO clone #4 displayed low levels of *RREB1*, as previously observed, see Section 3.3.1.2) and WB analysis, respectively (Figure 3.4 A and B). Interestingly, *RREB1* mRNA expression levels were significantly higher in differentiated cells (at all seven stages of development) compared to those observed in hiPSCs for WT clones ($p<0.05$).

3.3.3.2 Cell morphology differed between *RREB1* WT and KO clones at stages 5, 6 and 7

Cell morphology of differentiated *RREB1* WT and KO clones was similar for the first four stages of differentiation (data not shown). However, from stage 5 (EP) onwards, all four *RREB1* KO clones began to form three-dimensional clusters that were connected with each other via cellular projections and seemed to become bigger and darker along subsequent differentiation stages. Nothing similar was observed in any of the differentiating WT clones (Figure 3.5).

3.3.3.3 Assessment of key developmental marker expression in differentiated *RREB1* WT and KO lines

As described in Section 1.4.4.2 distinct stages of endocrine development express specific TFs and proteins important for beta cell development, functional maturation and/or beta cell identity. To evaluate whether differentiated *RREB1* WT and KO

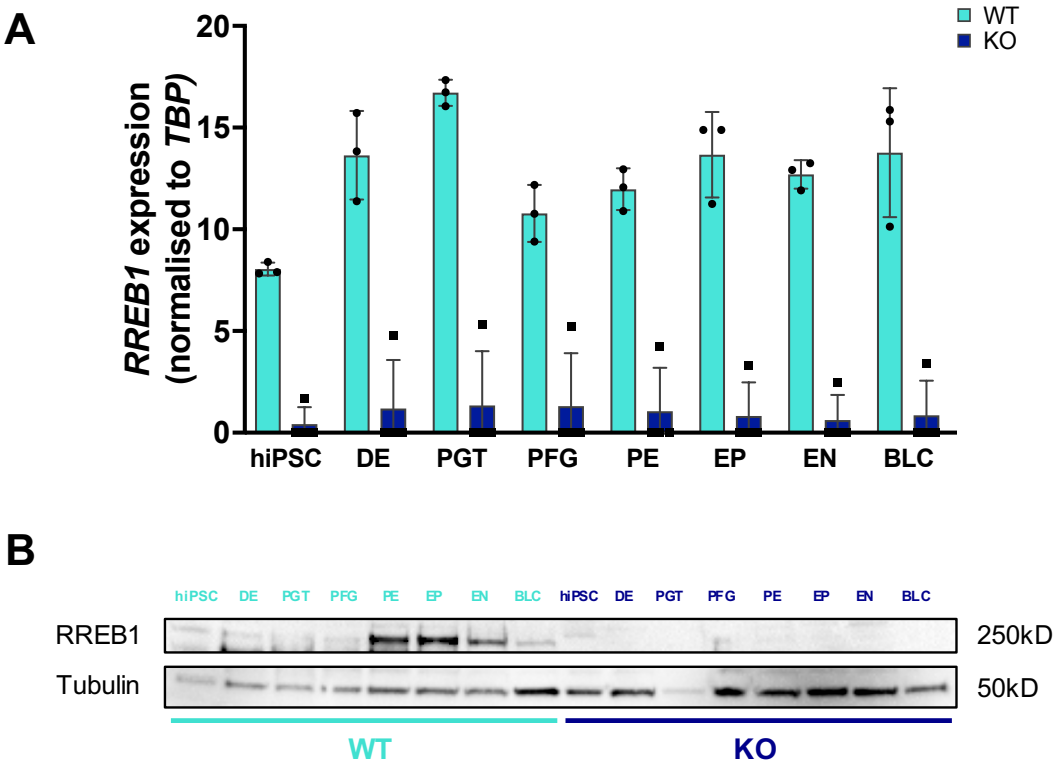


Figure 3.4: *RREB1* mRNA and RREB1 protein expression profiles during beta cell development RREB1 was expressed throughout all seven distinct stages of beta cell development in *RREB1* WT lines at the transcript (A) and protein level (example WB blot of WT clone #3 and KO clone #2) (B). WT n=3, KO n=4; values are displayed as mean±SD; $p<0.01$ (KO vs WT for hiPSCs and differentiated cells), $p<0.05$ (differentiated cells vs hiPSC) for unpaired t-test (A); DE, definitive endoderm; PGT, primitive gut tube; PFG, posterior foregut; PE, pancreatic endoderm; EP, endocrine precursors; EN, endocrine-like cells; BLC, beta-like cells; WT, turquoise; KO, darkblue.

cells followed endocrine lineage trajectories, they were characterised for mRNA and/or protein expression of stage-specific key developmental markers.

Stages 1 & 2 - definitive endoderm and posterior foregut

Levels of *OCT4* were comparable between *RREB1* WT and KO lines measured in hiPSCs, DE and PGT cells, respectively (Figure 3.6 A). While hiPSC lines expressed high levels of the pluripotency marker, expression was markedly reduced in DE cells (p=3.64x10^{-4} for WT, p=6.31x10^{-5} for KO lines) and stayed low in PGT cells, as expected when hiPSCs differentiate.

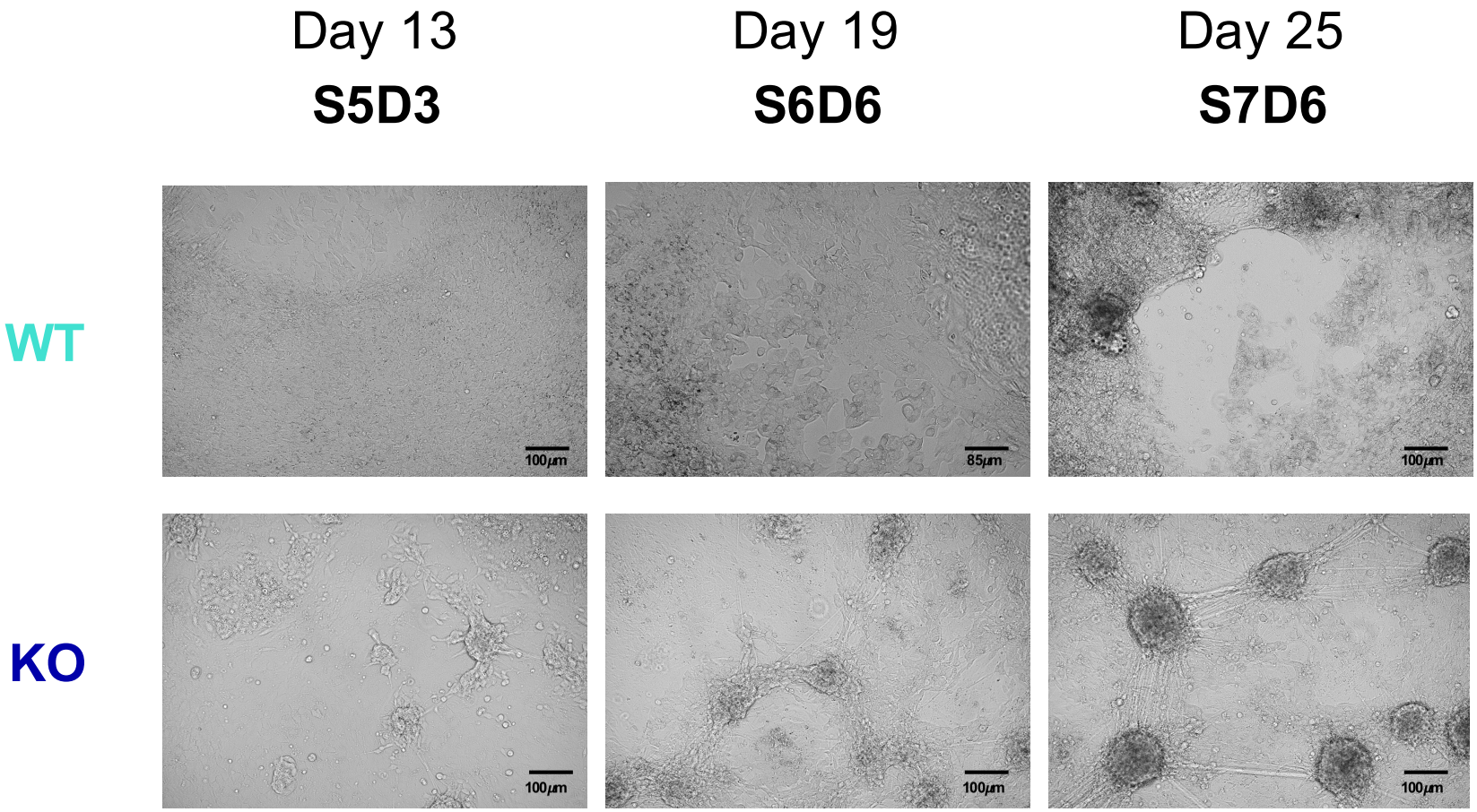


Figure 3.5: Cell morphology of differentiated *RREB1* KO and WT clones Morphology of *RREB1* KO and WT endocrine precursors (EP, S5D3), endocrine-like (EN, S6D6) and beta-like cells (BLC, S7D6). S, stage; D, day.

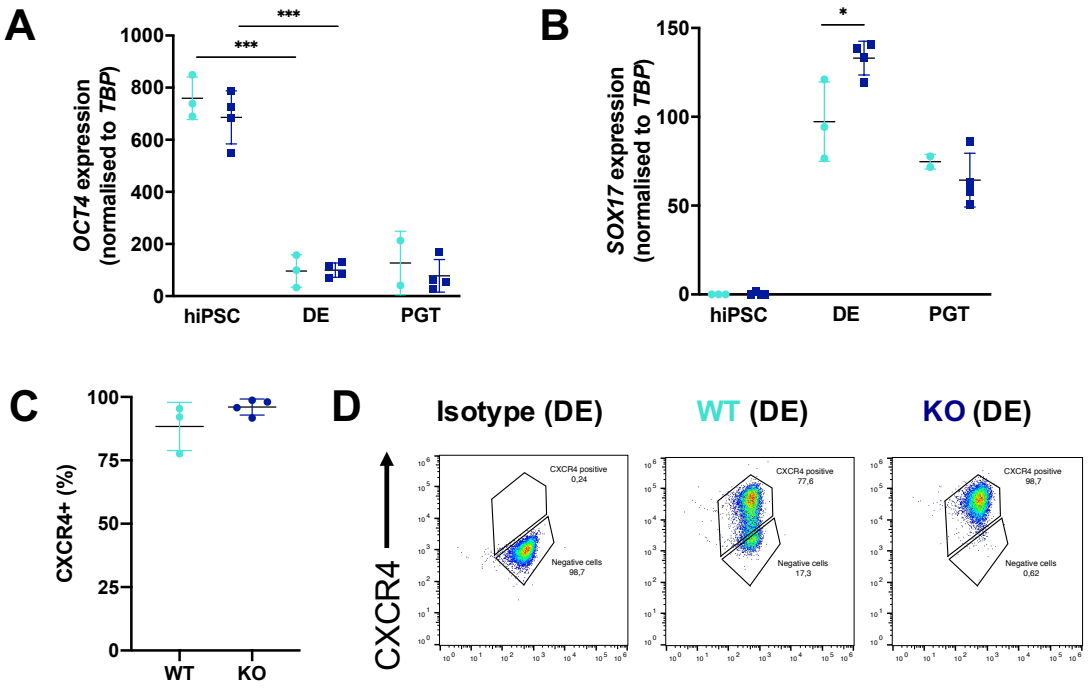


Figure 3.6: Stages 1 & 2 - Assessment of key developmental stage marker expression in differentiated *RREB1* WT and KO cells Differentiated *RREB1* KO and WT cells were evaluated for expression of *OCT4* (A) and *SOX17* (B). (C) The chemokine cell receptor CXCR4 is a marker of definitive endoderm (stage 1). (D) *RREB1* WT clone #2 revealed a lower percentage of CXCR4-positive cells. WT n=3, KO n=4; scatter dot plots display mean±SD; *p*-values *<0.05, ***<0.001 for unpaired t-test; DE, definitive endoderm; PGT, primitive gut tube; WT, turquoise; KO, darkblue.

The formation of DE in mouse and human ESCs is initiated by the TF SOX17 and marked by the expression of the chemokine receptor CXCR4 (Yasunaga et al. 2005; Qu et al. 2008; Séguin et al. 2008; D'Amour et al. 2005). *SOX17* expression was almost undetectable in hiPSC lines, peaked in DE cells and dropped again in PGT cells, suggesting successful induction of DE formation in both *RREB1* WT and KO lines (Figure 3.6 B). Expression of the cell surface marker CXCR4 assessed via FACS revealed that *RREB1* WT and KO hiPSCs had generated 88.4±9.5% and 96.1±3.2% DE cells, respectively (p=0.1806) (Figure 3.6 C), which is similar to hPSC-derived DE formation reported by other published studies (Bruin et al. 2014). While formation of DE was very homogenous from *RREB1* KO hiPSC lines, much more variability was observed in the differentiation efficiency of WT clones. For example, *SOX17* transcript levels were significantly higher in *RREB1* KO DE cells compared to WT lines, which was mainly driven by lower *SOX17*

expression in WT clones #1 and #2 (-36.7%, p=0.0325) (Figure 3.6 B). In line with this, *RREB1* WT clone #2 displayed a lower percentage of CXCR4-positive cells (77.6%) compared to all other DE samples assessed (>90% CXCR4-positive cells), matching the lower *SOX17* levels (Figure 3.6 D).

Stages 3, 4 & 5 - posterior foregut, pancreatic endoderm and endocrine precursors

RREB1 KO and WT lines first expressed *PDX1* at stage 3 (PFG), which stably persisted until stage 7 (BLC), with *RREB1* WT clone #3 steadily displaying higher levels of *PDX1* compared to the other *RREB1* WT as well as KO lines (Figure 3.7 A).

NKX6.1 expression was undetectable in most of the *RREB1* clones at stage 4 (PE), which was in contrast to robust *NKX6.1* expression observed by others in hESC-based models of human beta cell development following a similar differentiation protocol (Rezania et al. 2014). From stage 5 through to stage 7 *NKX6.1* transcript levels gradually increased in all clones, with *RREB1* KO lines showing higher *NKX6.1* levels compared to WT clones #1 and #2 and lower expression compared to WT clone #3 (Figure 3.7 B).

NEUROG3 was transiently expressed during stages 4 (PE) and 5 (EP) in *RREB1* WT and KO lines, consistent with previous observations (Rezania et al. 2014). Expression was significantly higher in *RREB1* KO clones compared to *RREB1* WT lines at stage 5 (102.3%, p=0.0069), pointing to a potential role for RREB1 in endocrine lineage specification (Figure 3.7 C).

NEUROD1 was first detected at stage 4 (PE). Whilst for *RREB1* WT clones, transcript levels gradually increased during subsequent differentiation stages with WT clone #3 consistently displaying higher mRNA levels than WT clones #1 and #2, in *RREB1* KO lines *NEUROD1* expression peaked at stage 5 (EP) before progressively decreasing (Figure 3.7 D).

Expression levels of *PAX4* and *HNF4A* were not affected by loss of RREB1 (Figure 3.7 E and F).

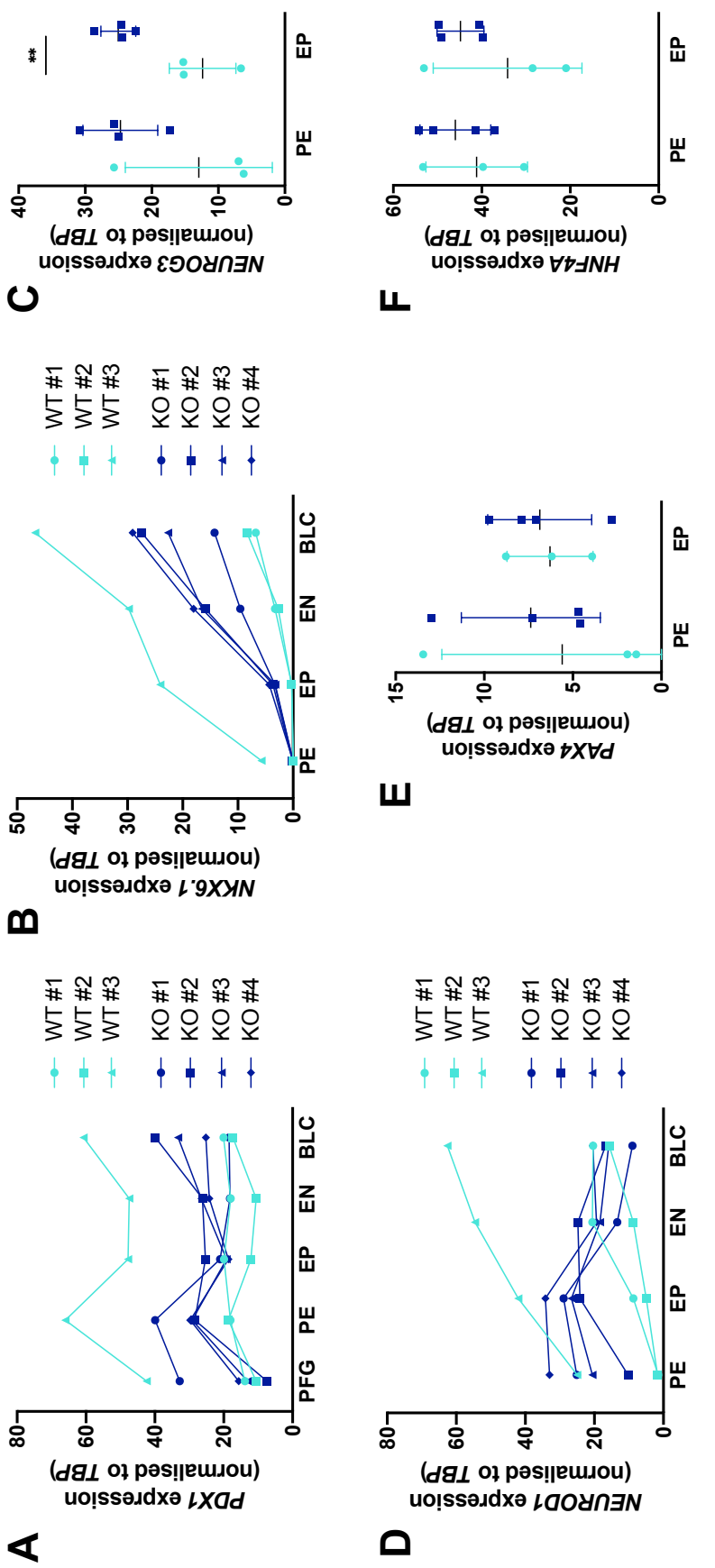


Figure 3.7: Stages 3, 4 & 5 - Assessment of key developmental stage marker expression in differentiated *RREB1* WT and KO cells PFG, PE and EP cells were evaluated for the expression of genes encoding the TFs *PDX1* (A), *NKX6.1* (B), *NEUROG3* (C), *NEUROD1* (D), *PAX4* (E) and *HNF4A* (F). WT n=3, KO n=4; scatter dot plots display mean±SD; *p*-values ** <0.01 for unpaired t-test; PFG, posterior foregut; PE, pancreatic endoderm; EP, endocrine precursors; EN, endocrine-like cells; BLC, beta-like cells; WT, turquoise; KO, darkblue.

Stages 6 & 7 - endocrine-like and beta-like cells

Endocrine-like and beta-like cells were characterised for the expression of the hormones insulin (*INS*) and glucagon (*GCG*) as well as genes implicated in insulin secretion (*ABCC8*, *KCNJ11* and *GJD2*). As observed for prior developmental marker expressions, *RREB1* WT clone #3 displayed higher levels of *INS*, *GCG*, *ABCC8* and *KCNJ11* compared to the other *RREB1* clones at stages 6 and 7 (Figure 3.8 A-D). *RREB1* KO clones revealed similar expression levels of *INS* and *GCG*, but higher levels of *ABCC8* (∼4.7x), *KCNJ11* (∼2.6x) and *GJD2* (∼4.1-13.2x) compared to *RREB1* WT clones #1 and #2 in the last two differentiation stages (Figure 3.8 A-E).

Co-expression of NKX6.1 and C-peptide was used as a surrogate to evaluate the percentage of BLCs in differentiated cultures at stage 7. Differentiation culture media contained ITS-X (Insulin-Transferrin-Selenium-Ethanolamine), therefore, to avoid confounding effects of potential uptake of insulin into cells (Rajagopal et al. 2003; Hansson et al. 2004), expression of C-peptide, a by-product of insulin synthesis and co-secreted with insulin, was used for the quantification of insulin positive cells in FACS analysis (Figure 3.8 F). On average 19.4±5.7% and 29.8±7.6% cells of *RREB1* WT and KO clones were characterised by co-expression of both NKX6.1 and C-peptide, respectively (p=0.1374) (Figure 3.8 G). Consistent with the observations of *INS* and *NKX6.1* transcript expression levels at stage 7, *RREB1* KO clones #1-4 as well as WT clone #3 displayed similar percentages of cells double positive for NKX6.1 and C-peptide (20.5%, 33.5%, 36.8%, 24.2% and 26.0%, respectively), while *RREB1* WT clones #1 and #2 were characterised by a lower proportion of these cells (16.2% and 16.0%).

Taken together, despite, sometimes high variability in stage-specific TF expression, characterisation of differentiated *RREB1* KO and WT cells revealed that each *RREB1* hiPSC line had followed established islet cell development and generated BLCs co-expressing NKX6.1 and insulin at stage 7, albeit to varying

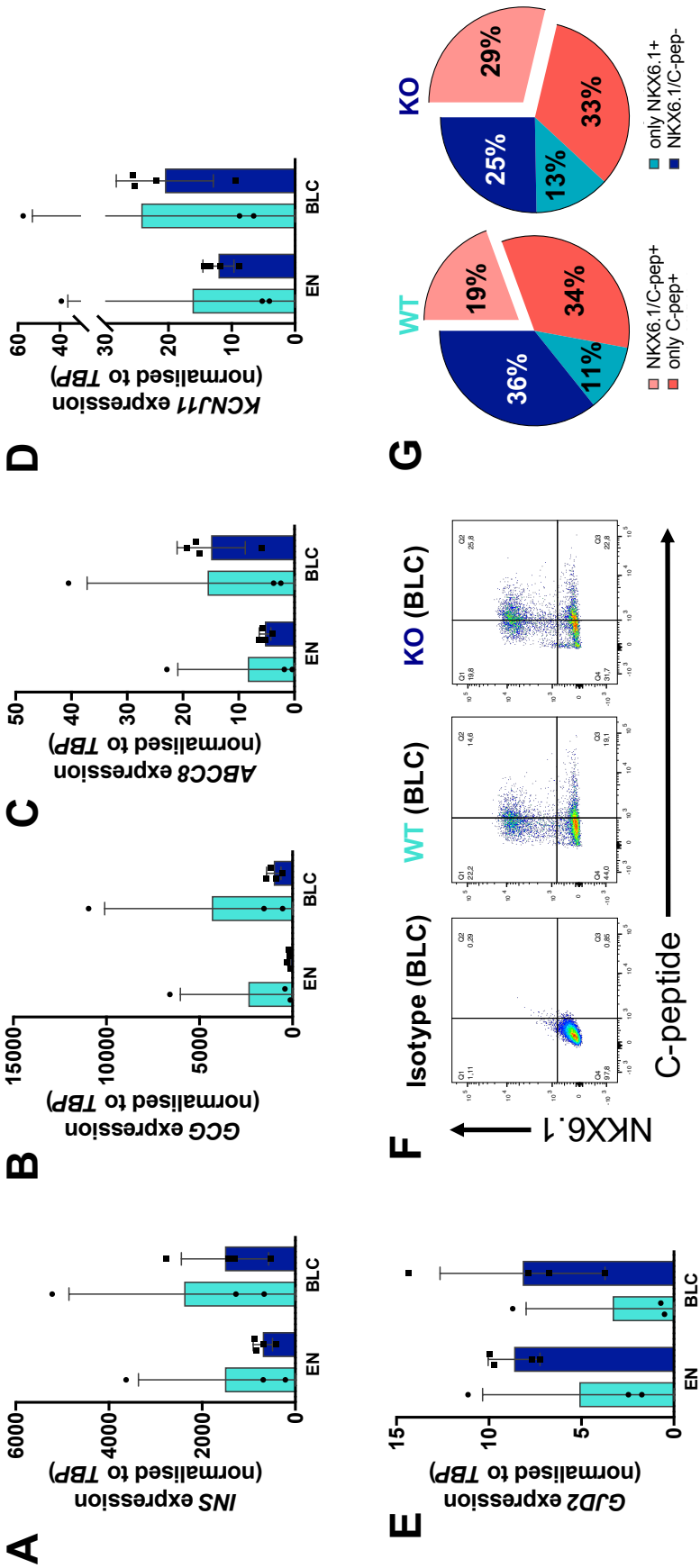


Figure 3.8: Stages 6 & 7 - Assessment of key developmental stage marker expression in differentiated *RREB1* WT and KO cells Endocrine-like and beta-like cells were evaluated for expression of the hormones *INS* (A) and *GCG* (B) as well as the K_{ATP} channel subunits *ABCC8* (C) and *KCNJ11* (D) and *GJD2*, encoding connexin 36 (E). (F) BLCs were assessed for the co-expression of NKX6.1 and C-peptide (surrogate marker for insulin) at stage 7. (G) Proportions of cells positive for NKX6.1 and C-peptide, just NKX6.1, just C-peptide or negative for both. WT n=3, KO n=4; scatter dot plots display mean±SD; EN, endocrine-like cells; BLC, beta-like cells; GCG, glucagon; WT, turquoise; KO, darkblue.

degrees. Significantly higher levels of *NEUROG3* in *RREB1* KO clones at stage 5 (EP) pointed to a potential role for RREB1 in endocrine differentiation.

3.3.4 Transcriptome profiling of *RREB1* KO and WT hiPSC lines differentiated along the endocrine lineage

Having confirmed successful differentiation of *RREB1* KO and WT hiPSCs into BLCs, I next sought to investigate how loss of RREB1 had affected the transcriptomes including transcriptional networks of differentiated cells. Thus, *RREB1* WT and *RREB1* KO clones were characterised via transcriptomic profiling at all seven distinct stages of *in vitro* beta cell differentiation.

3.3.4.1 *RREB1* transcript levels were up-regulated in differentiated *RREB1* KO lines

In total, 49 RNA-Seq libraries were sequenced to a mean read depth of 15.5±1.5 million reads per sample. *RREB1* was expressed throughout all seven stages of beta cell development in the WT lines (CV = 10.7%) (Figure 3.9 A) consistent with RT-qPCR findings (see Figure 3.4 A). Unexpectedly, assessment of *RREB1* expression patterns of individual WT and KO clones also revealed expression of *RREB1* in KO clones, and paradoxically at some stages even higher than expression levels observed in WT clones (Figure 3.9 B).

TFs have been shown to regulate their own expression via a negative feedback loop, i.e. the TF represses its own gene expression (G Pan et al. 2006). A search in the JASPAR (Mathelier et al. 2014) and TRANSFAC (Matys et al. 2006) databases revealed potential RREB1 binding sites in the *RREB1* gene promoter region, suggesting RREB1 might bind to its own promoter to repress further transcription. In *RREB1* KO lines, RREB1 cannot act as its own repressor due to loss of RREB1 protein, which might lead to gene derepression and up-regulation of mutated *RREB1* transcription. To investigate, whether this was the case for *RREB1* KO lines, *RREB1* exon usage in WT and KO lines was evaluated using the Bioconductor DEXSeq package in R (see Section 2.5.2.3). *RREB1* gene and transcript expression revealed

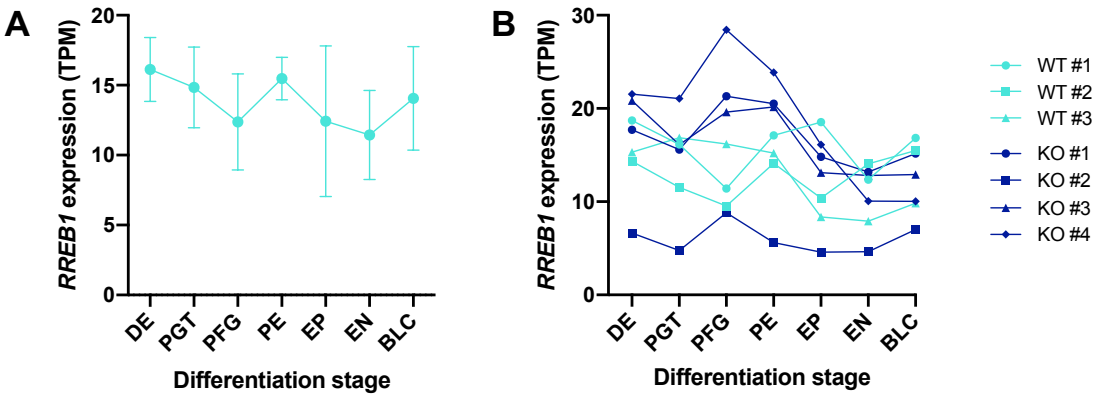


Figure 3.9: *RREB1* expression profiles of *RREB1* WT and KO clones during *in vitro* beta cell differentiation (A) *RREB1* was expressed throughout all seven distinct stages of beta cell development in *RREB1* WT lines. (B) Assessment of *RREB1* expression in individual WT and KO lines revealed unexpected expression of *RREB1* in KO clones. WT n=3, KO n=4; values are displayed as mean±SD (A); DE, definitive endoderm; PGT, primitive gut tube; PFG, posterior foregut; PE, pancreatic endoderm; EP, endocrine precursors; EN, endocrine-like cells; BLC, beta-like cells; TPM, transcripts per million; WT, turquoise; KO, darkblue.

differential exon usage for almost all exons in *RREB1* (Figure 3.10). As expected exons 4 to 10 were hardly expressed in *RREB1* KO lines and significantly lower than in WT clones (q<0.001). The low expression level originated from *RREB1* KO clone #4, highly likely attributable to the allele marked by a 724bp deletion (see Table 3.1). Expression of exons 1 to 3 was significantly higher in *RREB1* KO BLCs compared to WT controls (q<0.01) highly likely responsible for the observed *RREB1* expression in differentiated transcriptome-profiled *RREB1* KO clones.

3.3.4.2 Principal component analysis of differentiated *RREB1* WT and KO clones reveals heterogeneity between WT lines

Principal component analysis (PCA) of the transcriptome of the four differentiated *RREB1* KO and three WT lines showed that for the first three stages samples clustered mostly within their developmental stage (Figure 3.11). At stage 1 (DE) there was already a clear separation between *RREB1* KO and WT lines, however, while all four KO clones clustered tightly together, there was a more prominent separation among the WT lines. WT clones #1 and #2 clustered closer together than with clone #3. This divergence became more apparent from stage 3 (PFG)

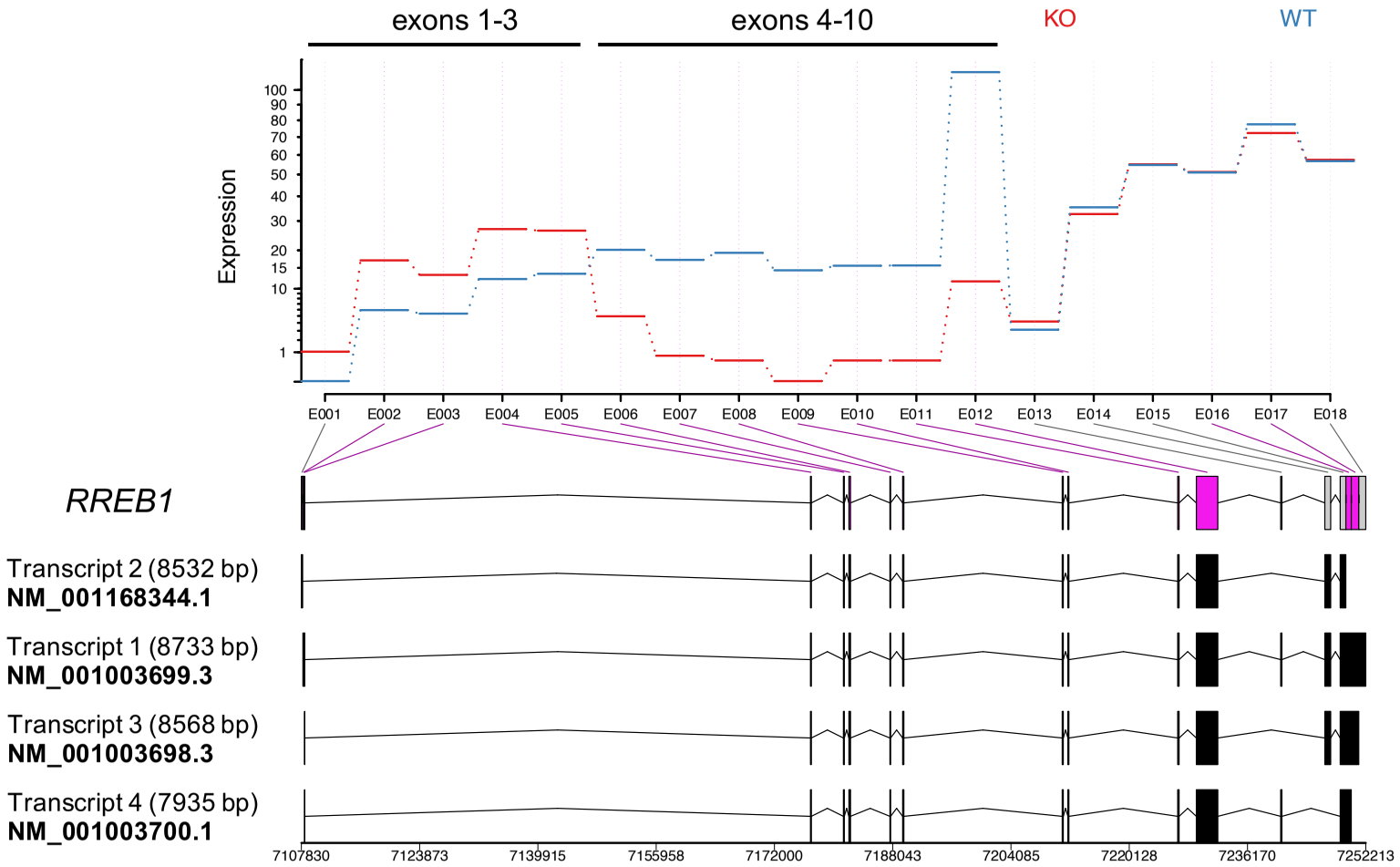


Figure 3.10: *RREB1* gene expression after analysis of differential exon usage using the DEXSeq package The fitted expression estimates for each exon of the *RREB1* gene are plotted for *RREB1* KO (red) and WT (blue) BLC samples. Exons that show significant differential exon usage are shown in pink. The annotated transcript variants for *RREB1* are specified. WT n=3, KO n=4; E001-E018 represent exon counting bins of the *RREB1* gene.

onwards. *RREB1* WT clone #3 clustered closer with *RREB1* KO clones and the separation between them and the two WT lines increased along the endocrine differentiation pathway.

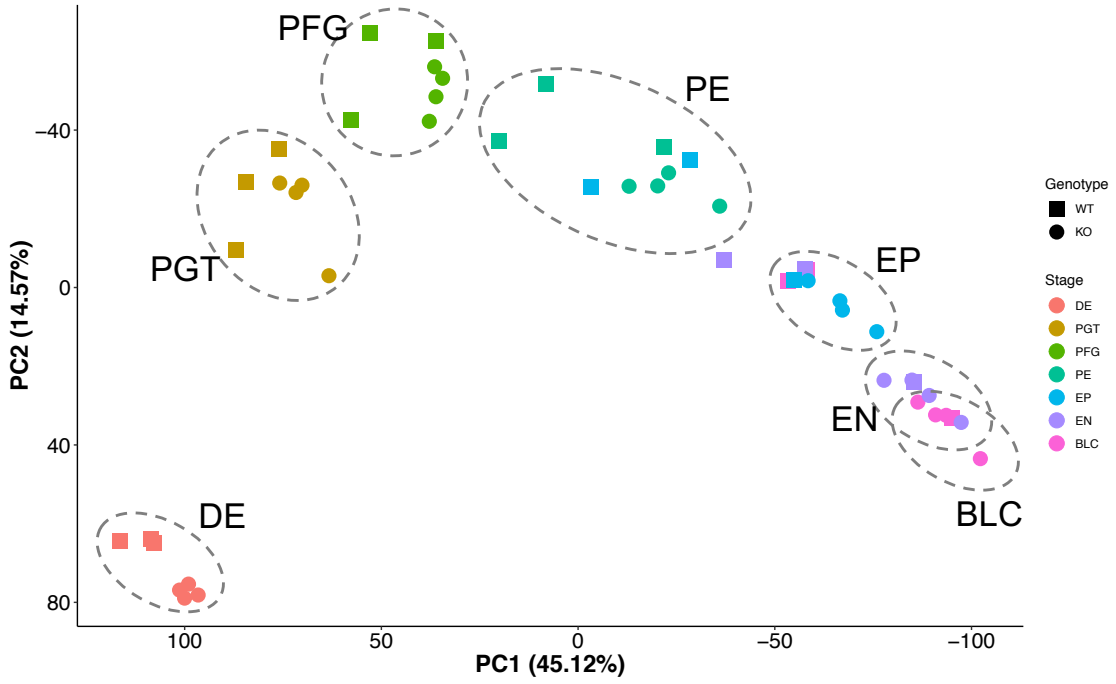


Figure 3.11: Principal component analysis of the transcriptomes of differentiated *RREB1* KO and WT clones The first two principal components (PC1, PC2) were calculated using normalised gene counts of *RREB1* KO and WT clones for all seven stages of *in vitro* beta cell differentiation. WT n=3, KO n=4; DE, definitive endoderm; PGT, primitive gut tube; PFG, posterior foregut; PE, pancreatic endoderm; EP, endocrine precursors; EN, endocrine-like cells; BLC, beta-like cells; WT, squares; KO, circles.

Due to the observed difference in PCA and key developmental marker expression between the individual *RREB1* WT clones, transcriptome data was analysed using a between-sample normalisation strategy which adjusts for technical effects called RUVSeq (removal of unwanted variation) (see Section 2.5.2.1). Figure 3.12 shows the effects of normalisation and RUVs on the individual RNA-Seq libraries as relative log expression (RLE, defined as the log-ratio of a read count to the median read count across a sample, top panel) and PCA plots (bottom panel). The boxplots of unnormalised RLE displayed distributional differences between library preparations of individual samples (Figure 3.12 A). RLE distributions are expected to centre

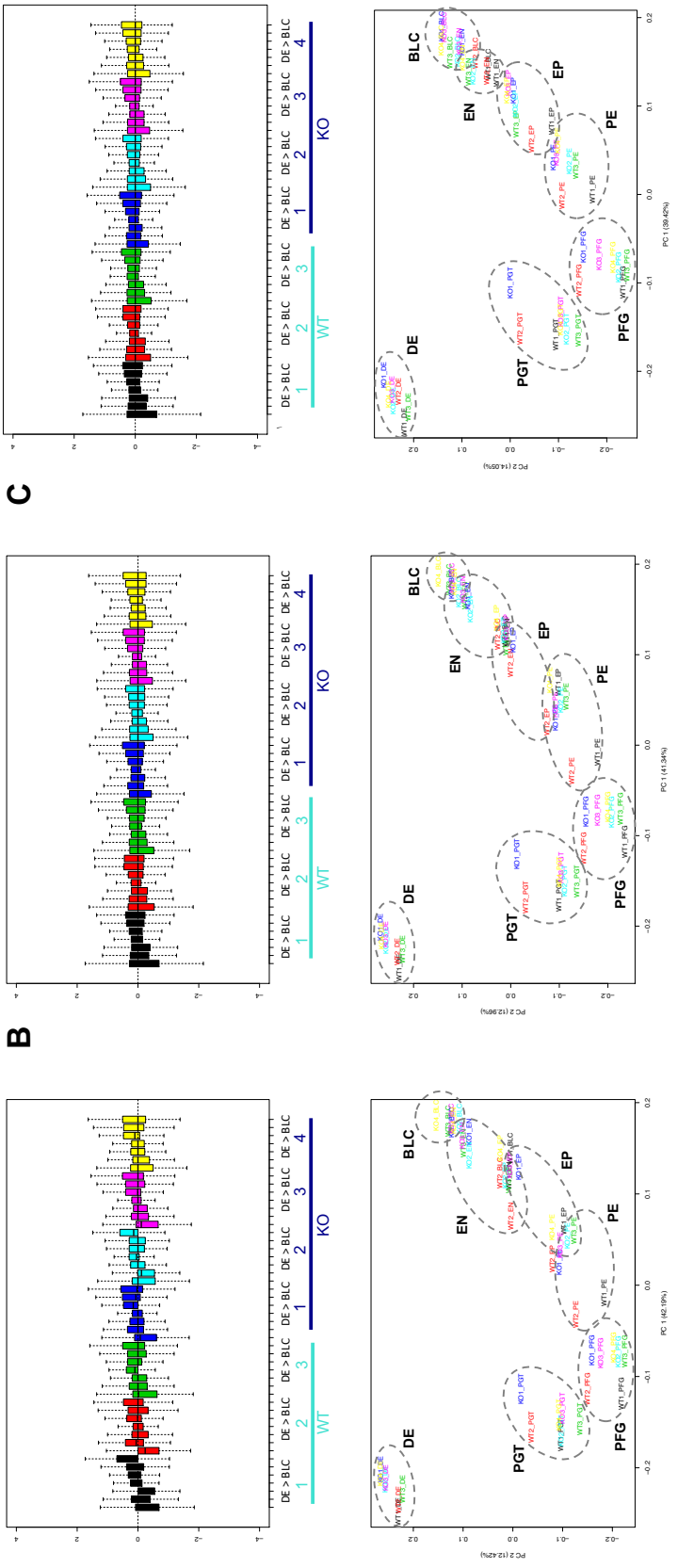


Figure 3.12: Removal of unwanted variation RLE (boxplots, top lane) and PCA (scatter plots of first two principal components, bottom lane) of unnormalised counts (A), counts after upper-quartile normalisation (B) and counts after RUVs (C). WT n=3, KO n=4; RLE, relative log expression; PCA, principal component analysis; boxplots: bottom and top of boxes represent the 1st and 3rd quartiles, respectively, while the black line inside indicates the median; whiskers show at 1.5 of the interquartile range (IQR) above and below the boxes; DE, definitive endoderm; PGT, primitive gut tube; PFG, posterior foregut; PE, pancreatic endoderm; EP, endocrine precursors; EN, endocrine-like cells; BLC, beta-like cells; WT, turquoise; KO, darkblue.

around zero and to look as similar as possible to each other. After upper-quartile normalisation, which adjusted for sequencing depth, RLE was centred around zero. However, scatter plots of the first two principal components showed that WT clone #3 remained one stage ahead compared to the other two WT clones (Figure 3.12 B). RUVs removed further variability between different libraries, visualised by reduced whisker length for some of the RLE boxplots and improved clustering in PCA, i.e. it resulted in a clearer separation between distinct stages of development as well as between *RREB1* WT and KO lines (Figure 3.12 C). Therefore, all further analysis was performed on RUVs normalised counts.

3.3.4.3 Differential gene expression analysis

Genes that were differentially expressed as a consequence of loss of RREB1 were identified at each developmental stage using the Bioconductor DESeq2 package (see Section 2.5.2.2). A total of 5,476 differentially expressed genes (DEGs) were identified across all stages between *RREB1* KO and WT lines, of which 159 were common to all developmental stages. Table 3.3 gives an overview of the number of DEGs detected at each individual stage (q<0.01). The majority of DEGs were up-regulated in the *RREB1* KO lines (62.6±4.5% vs 37.4±4.5%) and on average 47±2% of these matched computationally predicted RREB1 target genes (Matys et al. 2006; Mathelier et al. 2014; Marbach et al. 2016). A subset of enriched biological terms and pathways among up-regulated DEGs for each differentiation stage are listed in Table 3.4. Interestingly, all stages were enriched for the terms 'nervous system development' and 'neural system', which might be due to the fact that neurons and beta cells share various phenotypic traits including the expression of neuronal TFs, electrical excitability and an exocytotic machinery (Atouf et al. 1997; van Arensbergen et al. 2010). In line with this, further biological terms relating to neurons, their development and synaptic processes ('neuron differentiation', 'netrin-1 signalling', 'cAMP signalling' 'glutamatergic synapse', 'transmission across chemical synapses', 'axon guidance') were observed across all seven differentiation stages. Interestingly, DEGs up-regulated in *RREB1* KO PE cells contained a subset of genes

Table 3.3: Differentially expressed genes between *RREB1* WT and KO lines at seven stages of beta cell development

	DE	**PGT**	**PFG**	**PE**	**EP**	**EN**	**BLC**
DEGs	1,759	1,278	1,230	1,408	2,611	1,525	1,948
Up-regulated	1,015 (58%)	853 (67%)	837 (68%)	926 (66%)	1,570 (60%)	947 (62%)	1,108 (57%)
Predicted RREB1 target genes	489 (48%)	423 (49%)	418 (50%)	437 (47%)	698 (44%)	431 (45%)	510 (46%)
Down-regulated	744 (42%)	425 (33%)	393 (32%)	482 (34%)	1,041 (40%)	578 (38%)	840 (43%)
Predicted RREB1 target genes	434 (58%)	237 (56%)	194 (49%)	267 (55%)	602 (58%)	282 (49%)	378 (45%)

27,340 genes tested using DESeq2; 5,476 unique genes differentially expressed at $q<0.01$; 159 genes common to all stages; DEGs, differentially expressed genes; DE, definitive endoderm; PGT, primitive gut tube; PFG, posterior foregut; PE, pancreatic endoderm; EP, endocrine precursors; EN, endocrine-like cells; BLC, beta-like cells.

Table 3.4: Subset of enriched biological terms and pathways among DEGs up-regulated in *RREB1* KO lines at seven distinct stages of beta cell development

Stage		# enriched terms	Examples of enriched terms	padj (q)
DE	GO	181	nervous system development	1.22×10^{-33}
			trans-synaptic signalling	4.54×10^{-27}
			synaptic signalling	8.55×10^{-27}
	REACTOME	14	neuronal system	1.37×10^{-16}
			protein-protein interactions at synapses	2.16×10^{-9}
			transmission across chemical synapses	4.64×10^{-7}
	KEGG	5	glutamatergic synapse	3.44×10^{-8}
			axon guidance	6.41×10^{-6}
			cell adhesion molecules (CAM)	0.0010
PGT	GO	193	nervous system development	6.21×10^{-21}
			neuron projection development	6.79×10^{-16}
			synaptic signalling	1.57×10^{-13}
	REACTOME	16	neuronal system	1.85×10^{-12}
			protein-protein interactions at synapses	7.51×10^{-8}
			transmission across chemical synapses	8.90×10^{-6}
	KEGG	13	cell adhesion molecules (CAM)	2.75×10^{-8}
			cAMP signalling pathway	0.0012
			calcium signalling pathway	0.0049
PFG	GO	182	nervous system development	2.63×10^{-26}
			generation of neurons	2.10×10^{-21}
			synaptic signalling	1.79×10^{-14}
	REACTOME	15	neural system	8.70×10^{-15}
			protein-protein interactions at synapses	2.11×10^{-7}
			netrin-1 signalling	0.0062
	KEGG	6	axon guidance	1.86×10^{-6}
			glutamatergic synapse	5.34×10^{-6}
			cell adhesion molecules (CAM)	4.68×10^{-5}
PE	GO	201	nervous system development	3.40×10^{-34}
			neuron differentiation	4.15×10^{-25}
			synaptic signalling	5.69×10^{-22}
	REACTOME	26	neuronal system	4.98×10^{-21}
			protein-protein interactions at synapses	4.31×10^{-6}
			regulation of gene expression in endocrine-committed ($NEUROG3^{+}$) progenitor cells	0.0062
	KEGG	18	glutamatergic synapse	1.43×10^{-6}
			cAMP signalling pathway	0.0028
			calcium signalling pathway	0.0043

Continued on next page

Stage		# enriched terms	Examples of enriched terms	padj (q)
EP	GO	337	nervous system development	$2.58x10^{-62}$
			synaptic signalling	$4.12x10^{-48}$
			cell-cell signalling	$3.68x10^{-33}$
	REACTOME	50	neuronal system	$6.77x10^{-40}$
			regulation of insulin secretion	$4.99x10^{-5}$
			netrin-1 signalling	0.0042
	KEGG	21	cAMP signalling pathway	$2.05x10^{-8}$
			synaptic vesicle cycle	$1.23x10^{-7}$
			insulin secretion	$2.45x10^{-5}$
EN	GO	236	nervous system development	$4.56x10^{-43}$
			synaptic signalling	$9.77x10^{-41}$
			synaptic vesicle exocytosis	$1.8x10^{-17}$
	REACTOME	40	neuronal system	$9.61x10^{-35}$
			transmission across chemical synapses	$4.68x10^{-24}$
			regulation of insulin secretion	$2.67x10^{-5}$
	KEGG	26	cAMP signalling pathway	$1.27x10^{-5}$
			insulin secretion	$1.29x10^{-5}$
			gap junction	$1.05x10^{-4}$
BLC	GO	269	synaptic signalling	$1.28x10^{-51}$
			nervous system development	$4.60x10^{-43}$
			synaptic vesicle exocytosis	$2.14x10^{-17}$
	REACTOME	42	neuronal system	$1.11x10^{-45}$
			transmission across chemical synapses	$4.66x10^{-31}$
			regulation of insulin secretion	$3.95x10^{-6}$
	KEGG	21	calcium signalling pathway	$8.10x10^{-8}$
			cAMP signalling pathway	$6.43x10^{-6}$
			insulin secretion	$5.58x10x10^{-4}$

837-1,570 up-regulated DEGs were tested for enriched biological terms and pathways using g:profiler (see Section 2.5.3). All human genes annotated in the Ensembl database were used as background. Significance threshold was set to $q<0.01$ using the tailor-made g:SCS algorithm for multiple testing. DEGs, differentially expressed genes; DE, definitive endoderm; PGT, primitive gut tube; PFG, posterior foregut; PE, pancreatic endoderm; EP, endocrine precursors; EN, endocrine-like cells; BLC, beta-like cells.

involved in the 'regulation of gene expression in endocrine committed ($NEUROG3^+$) progenitor cells' and from stage 5 (EP) onwards up-regulated DEGs were enriched for pathways relating to 'insulin secretion' and 'regulation of insulin secretion'.

3.3.4.4 Endocrine progenitor markers were enriched among genes up-regulated in *RREB1* KO lines

To assess whether loss of RREB1 had an impact on differentiation quality, DEGs between *RREB1* KO and WT clones were tested for pancreatic progenitor, endocrine progenitor and endocrine cell-specific marker expression, identified in human fetal pancreata (Ramond et al. 2018), using a hypergeometric enrichment analysis

test. This revealed an enrichment of endocrine progenitor markers among genes up-regulated in the *RREB1* KO lines at stages 4 (PE, q=4.0x10^{-83}), 5 (EP, q=6.3x10^{-104}) and 6 (EN, q=5.5x10^{-43}) (Figure 3.13 A). Up-regulated endocrine progenitor markers included genes encoding TFs with a well-established role in beta cell development such as *NEUROG3* and its targets *NKX2.2*, *NEUROD1* and *RUNX1T1* as well as REST target genes *MFNG*, *SYT4* and *SCG2*.

In parallel there was an enrichment of DEGs down-regulated in early and late pancreatic progenitors at stage 5 (EP, q=3.2x10^{-28} and q=7.2x10^{-24}, respectively). Differentially expressed down-regulated genes comprised acinar cell markers *CPA2* and *NR5A2* as well as the TF *HNF1B*, which has been implicated in the maintenance and expansion of multipotent pancreatic progenitors in mice (Haumaitre et al. 2005; De Vas et al. 2015; Larsen et al. 2017). Additionally, members of the Notch signalling pathway *NOTCH1*, *NOTCH2* and *JAG1* were significantly lower expressed in *RREB1* KO than in WT lines. Notch signalling is active in pancreatic progenitors, ceases during endocrine differentiation and has been suggested to get up-regulated again in mature islet cells playing a role in maintaining beta cell identity (Apelqvist et al. 1999; Murtaugh et al. 2003; Szabat et al. 2012). Furthermore, the lower expression of *ERBB3* and *FGFR2* suggests a down-regulation of the EGF and FGF pathways in *RREB1* KO EP cells (Miettinen et al. 2000; Bhushan et al. 2001).

The opposing enrichment of up-regulated DEGs in endocrine progenitor and endocrine cell populations, as well as down-regulated DEGs in pancreas progenitor populations, suggests that a higher percentage of *RREB1* KO hiPSCs generated endocrine precursors, while *RREB1* WT clones contained a higher proportion of pancreatic progenitors at differentiation stage 5.

To investigate whether the increase of endocrine progenitor marker transcripts was also reflected at the protein level, NEUROD1 expression was evaluated in cells differentiated to stages 4 (PE) or 5 (EP) via WB analysis. *GAPDH* levels were not affected by loss of RREB1 (q=0.5602 for PE and q=0.4232 for EP), thus GAPDH was used as protein loading control (Figure 3.14 A bottom). NEUROD1 expression

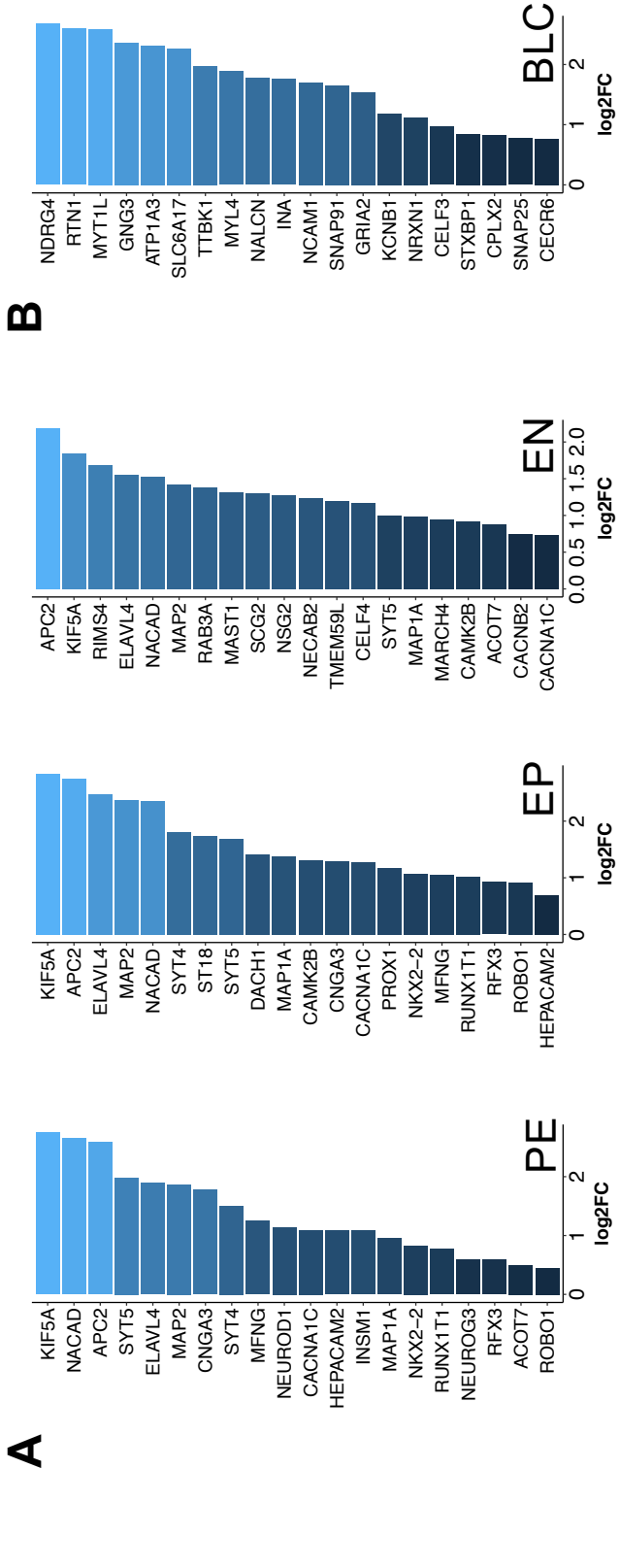


Figure 3.13: Endocrine progenitor markers were enriched among DEGs up-regulated in *RREB1* KO lines at stages 4 to 6 (A) *RREB1* KO lines were characterised by higher expression levels of endocrine progenitor markers at developmental stages 4 to 6 (PE to EN). (B) Endocrine markers were enriched in up-regulated DEGs in *RREB1* KO clones at stage 7 (BLC). PE, pancreatic endoderm; EP, endocrine precursors; EN, endocrine-like cells; BLC, beta-like cells.

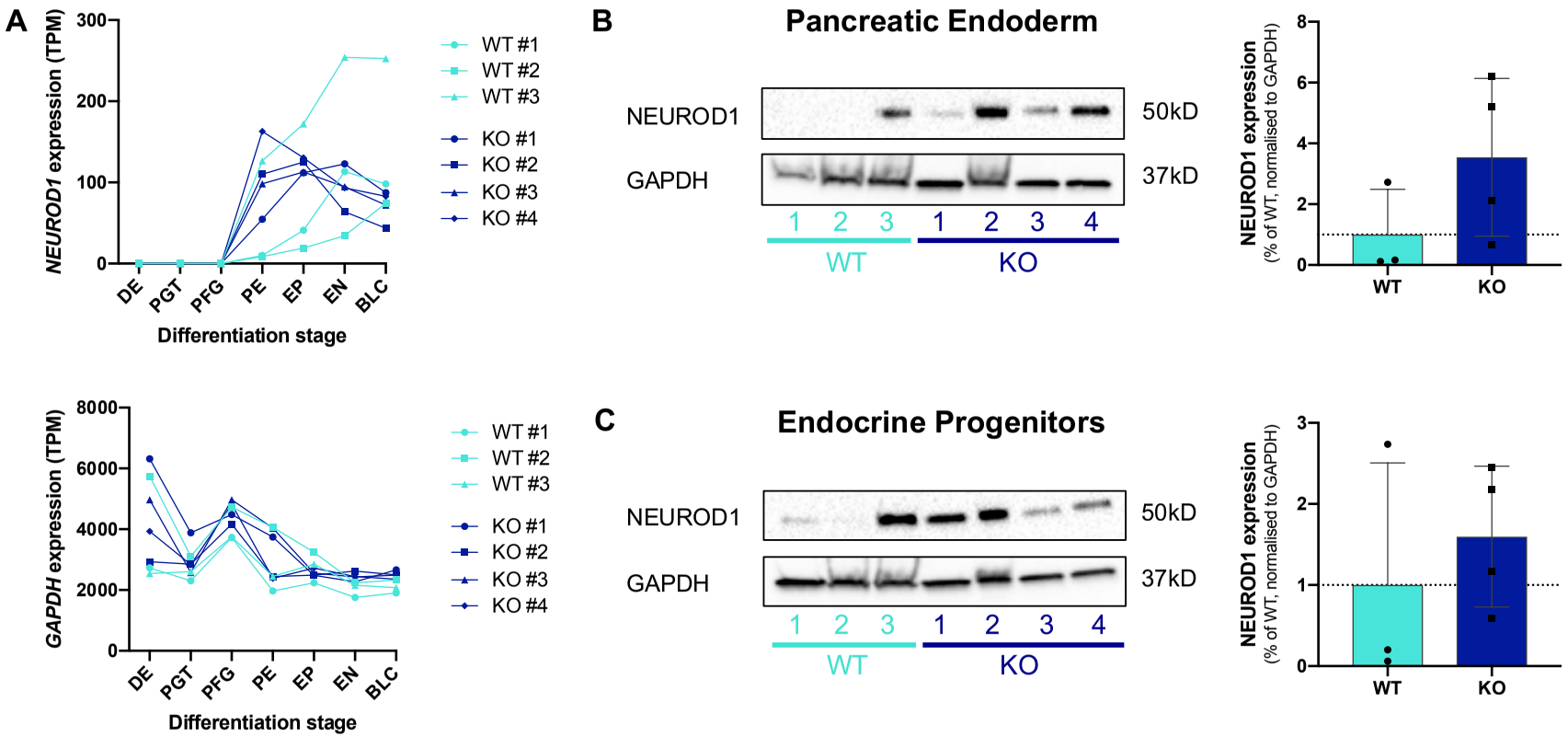


Figure 3.14: Effect of loss of RREB1 on NEUROD1 expression during beta cell differentiation (A) *NEUROD1* and *GAPDH* transcript levels in individual *RREB1* WT and KO models throughout beta cell differentiation. (B) NEUROD1 protein expression in individual *RREB1* WT and KO clones at stage 4 (PE) (B) and stage 5 (EP) (C). GAPDH expression was used as protein loading control. WT n=3, KO n=4; values are displayed as mean±SD; DE, definitive endoderm; PGT, primitive gut tube; PFG, posterior foregut; PE, pancreatic endoderm; EP, endocrine precursors; EN, endocrine-like cells; BLC, beta-like cells; TPM, transcripts per million; WT, turquoise; KO, darkblue.

was highly variable between individual *RREB1* clones (PE: WT CV=149.0%, KO CV=73.2% and EP: WT CV=150.6%, KO CV=54.4%) (Figure 3.14 B and C), reflecting the variability already observed at the transcript level (Figure 3.14 A top). While *RREB1* WT clones #1 and #2 hardly expressed NEUROD1, expression levels of NEUROD1 in WT clone #3 were comparable to those observed in *RREB1* KO clones in PE and EP cells (PE, p=0.1937 and EP, p=0.5328).

3.3.4.5 Genes implicated in insulin exocytosis are enriched among genes up-regulated in *RREB1* KO BLCs

Assessment of endocrine marker expression (see Section 1.4.5.2) in BLCs revealed an enrichment in DEGs up-regulated in *RREB1* KO lines at stage 7 (q=1.6x10^{-35}) (Figure 3.13 B). Genes implicated in insulin exocytosis showed higher expression levels in *RREB1* KO than WT clones. Examples included *SNAP25*, *SNAP91*, *STXBP1* and *CPLX2* which are part of the SNARE complex as well as *NRXN1*, encoding neurexins, and *GRIA2*, encoding a glutamate receptor subunit (Mosedale et al. 2012; ZY Wu et al. 2012).

3.3.4.6 Prediction of upstream transcriptional regulators of DEGs

For assessment of TF enrichment, upstream regulators of DEGs at each individual developmental stage were predicted using the iRegulon Cytoscape plugin, which is based on combined motif and track discovery (9,713 TF motifs and 1,120 ENCODE ChIP-seq tracks, centred 10kb around the TSS) (see Section 2.5.3).

RREB1 was among the top three predicted regulators of DEGs up-regulated in *RREB1* KO lines across all seven differentiation stages (Table 3.5). Interestingly, the most enriched motif across all stages (except EP) belonged to the RFX TF family that comprises eight members, characterised by a highly conserved DNA binding domain (Aftab et al. 2008; Sugiaman-Trapman et al. 2018). *RFX1*, *RFX2*, *RFX3*, *RFX5*, *RFX7* and *RFX8* are predicted RREB1 target genes (Matys et al. 2006; Mathelier et al. 2014; Marbach et al. 2016) and RFX3 and RFX6, have been shown to play a role in beta cell development (Ait-Lounis et al. 2007; Ait-Lounis et al.

2010; Soyer et al. 2010; Smith et al. 2010; Taleb et al. 2011). The third predicted regulator was the transcriptional repressor REST, which has been shown to act as a key repressor of beta cell differentiation (D Martin et al. 2015; D Martin et al. 2017).

Predicted upstream regulators of DEGs down-regulated in *RREB1* KO clones showed an enrichment of hepatocyte nuclear factors including the HNF4 (HNF4A and HNF4G) and the HNF3 (FOXA1, FOXA2 and FOXA3) subfamily at stage 3 (PFG). Down-regulated DEGs at stages 4 (PE) and 7 (BLC) were also predicted to be regulated by hepatocyte nuclear TFs (FOXA2, HNF4A and HNF4G for PE; HNF1B, HNF4A and HNF4G for BLCs). HNF1B, HNF4A and FOXA2 are well-established developmental transcriptional regulators of the pancreas (Haumaitre et al. 2005; N Gao et al. 2008; Ng et al. 2019). In addition, iRegulon suggested that a subset of DEGs down-regulated in *RREB1* KO clones at stages 4 (PE) and 5 (EP) were regulated by TEAD TFs. In particular, TEAD1 and TEAD4, members of the transcriptional enhancer family, are part of the Hippo signalling pathway, a key regulator in the expansion of human pancreatic progenitors (Cebola et al. 2015).

3.3.4.7 Prediction of genome-wide regulatory interactions during *in vitro* beta cell development using motif activity response analysis

Motif activity response analysis (MARA) is another approach to interrogate genome-wide regulatory interactions underlying gene expression variation across a set of samples (Balwierz et al. 2014). To predict key regulators, their interactions and target genes implicated in beta cell development and to assess the impact of loss of RREB1 on identified regulatory networks, *RREB1* WT and KO transcriptome data was evaluated using MARA at seven distinct stages of *in vitro* beta cell differentiation (see Section 2.5.4).

MARA predicted RFX2 and RFX3 as well as REST as the key TFs driving gene expression variation across *RREB1* KO and WT samples (RFX2/3 Z=12.43, REST Z=12.08) in accordance with the predictions made by iRegulon (see Table 3.5).

Table 3.5: Predicted upstream regulators for DEGs up-regulated in *RREB1* KO clones at seven stages of beta cell development using iRegulon

	DE		**PGT**		**PFG**		**PE**		**EP**		**EN**		**BLC**	
TF	NES	# targets	NES	# targets	NES	# targets	NES	# targets	NES	# targets	NES	# targets	NES	# targets
RFX	3.1-5.3	335	3.2-5.4	323	3.1-6.6	410	3.2-6.4	431	3.2-5.7	590	3.5-5.9	314	3.2-5.5	265
RREB1	4.1-4.9	617	4.0-4.2	479	3.7-5.1	485	3.4-5.1	431	4.1	683	3.3-4.8	440	3.5-4.1	331
REST	3.5	321	NA	NA	3.4	284	3.3-5.3	279	3.7-6.1	444	3.0-5.6	524	3.4-5.4	339

Top three TF motifs, most common across all stages of 9,713 position weight matrices (PWMs) and 1,120 ENCODE ChIP-Seq tracks (centred 10kb around TSS) tested. NES, normalised enrichment score with cutoff set to $\geq$3 (corresponding to a false discovery rate (FDR) of 3-9%) (see section 2.5.3); # targets, number of targets for TF motif with highest NES; RFX comprises RFX1-6, 8, RFXANK and RFXAP; NA, TF motif not enriched in DEGs; DEGs, differentially expressed genes; TF, transcription factor; DE, definitive endoderm; PGT, primitive gut tube; PFG, posterior foregut; PE, pancreatic endoderm; EP, endocrine precursors; EN, endocrine-like cells; BLC, beta-like cells.

3.3.4.8 MARA identified differential transcriptional activities for RFX2 and RFX3 between *RREB1* KO and WT lines across all developmental stages

RFX2/3 motif activity resembled a sigmoid (S-shaped) curve with stage 4 (PE) representing the turning point. Motif activity was significantly increased in *RREB1* KO lines across all stages of beta cell development (DE q=1.0x10^{-14}, PGT q=3.14x10^{-13}, PFG q=2.2x10^{-14}, PE q<1.0x10^{-15}, EP q<1.0x10^{-15}, EN q=2.2x10^{-15}, BLC q=8.0x10^{-15}) (Figure 3.15 A). As the binding motifs of RFX2 and RFX3 were very similar (Figure 3.15 B), it was not clear which of the two TFs or whether both of them were accountable for the observed motif activity. *RFX2* expression was significantly increased in *RREB1* KO clones at various stages of development (DE q=3.29x10^{-8}, PGT q=1.02x10^{-5}, PE q=8.91x10^{-5}) (Figure 3.15 C). The same was observed for *RFX3* (DE q=6.33x10^{-6}, PGT q=0.0252, PFG q=0.0002, PE q=5.29x10^{-6}, EP q=3.08x10^{-6}) (Figure 3.15 D). Both *RFX2* and *RFX3* expression significantly correlated with motif activity (r=0.6919 and r=0.8665, respectively) (Figure 3.15 E and F). The positive correlation strongly suggested that RFX2 and RFX3 acted as activators. RFX2/3 target genes *CAMK2A* (Ma et al. 2017) and *GPR56* (Bae et al. 2014) were among the strongest up-regulated DEGs. Gene expression was significantly increased in *RREB1* KO lines across all seven stages and *GPR56* correlated markedly with *RFX3* gene expression (r=0.8709, p=5.02x10^{-5}), confirming RFX2 and RFX3 as transcriptional activators (Figure 3.15 G and H). The promoter regions of both *RFX2* and *RFX3* contain a putative RREB1 binding site (Matys et al. 2006; Marbach et al. 2016), thus RREB1 might act as a direct transcriptional repressor of these genes.

3.3.4.9 Loss of RREB1 did not affect transcriptional activity or expression of RFX6

Originally, the TF RFX6 was not included in MARA due to insufficient information about its binding motif (Balwierz et al. 2014). As RFX6 is the best characterised islet-specific RFX TF with an established role in human beta cell development (Zhu

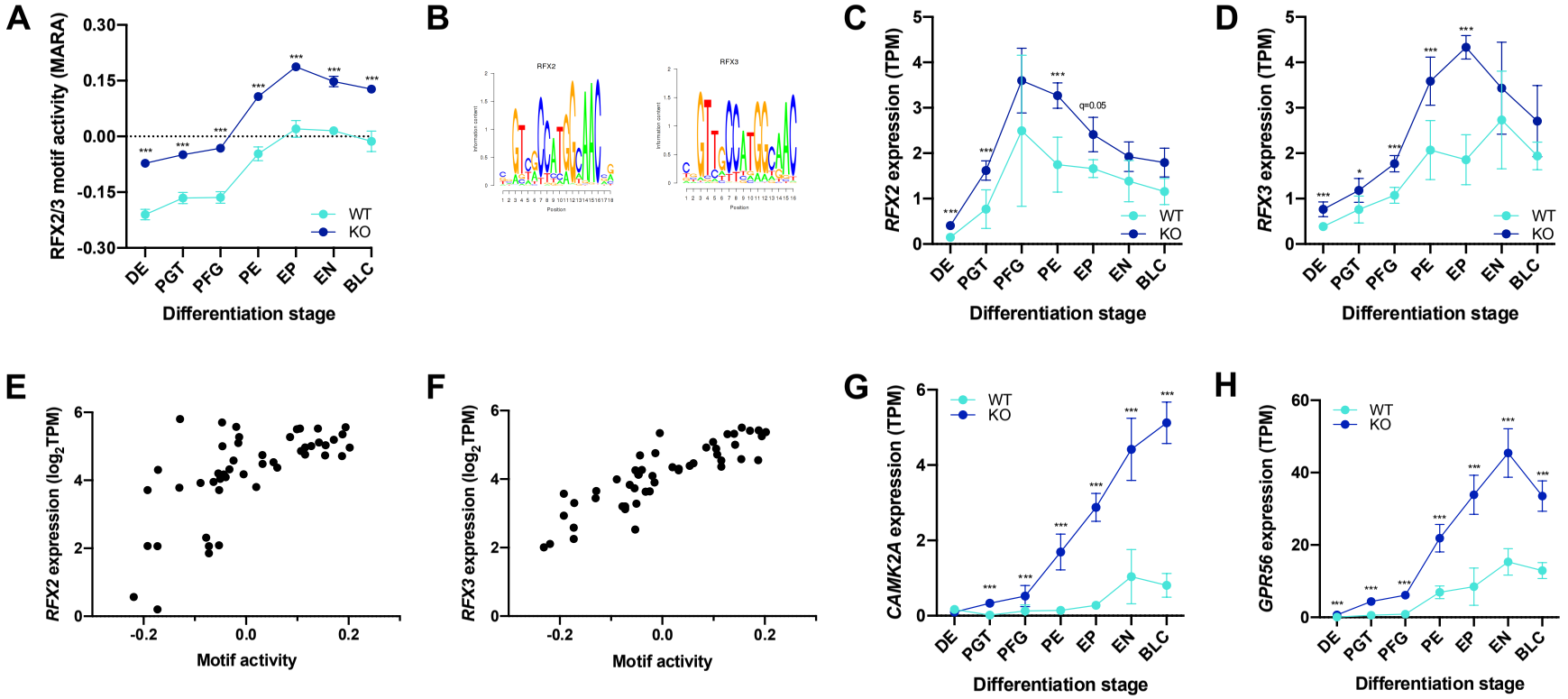


Figure 3.15: RFX2/3 motif activity was significantly increased in *RREB1* KO lines (A) RFX2/3 motif activity profile for *RREB1* WT and KO lines during beta cell development calculated using MARA. (B) RFX2 and RFX3 motifs displayed as PWMs. Expression of *RFX2* (C) and *RFX3* (D) in *RREB1* WT and KO lines during beta cell development. *RFX2* (E) and *RFX3* (F) expression positively correlated with motif activity. Expression of RFX2/3 target genes *CAMK2A* (G) and *GPR56* (H) were significantly increased in *RREB1* KO lines. WT n=3, KO n=4; values are displayed as mean±SD; *p*-values *<0.05, **<0.01, ***<0.001 for unpaired t-test followed by Holm-Sidak correction for multiple comparisons (A) or Wald test followed by Benjamini-Hochberg adjustment (C, D, G and H); Pearson correlation (E and F); DE, definitive endoderm; PGT, primitive gut tube; PFG, posterior foregut; PE, pancreatic endoderm; EP, endocrine precursors; EN, endocrine-like cells; BLC, beta-like cells; PWM, position weight matrix; TPM, transcripts per million; WT, turquoise; KO, darkblue.

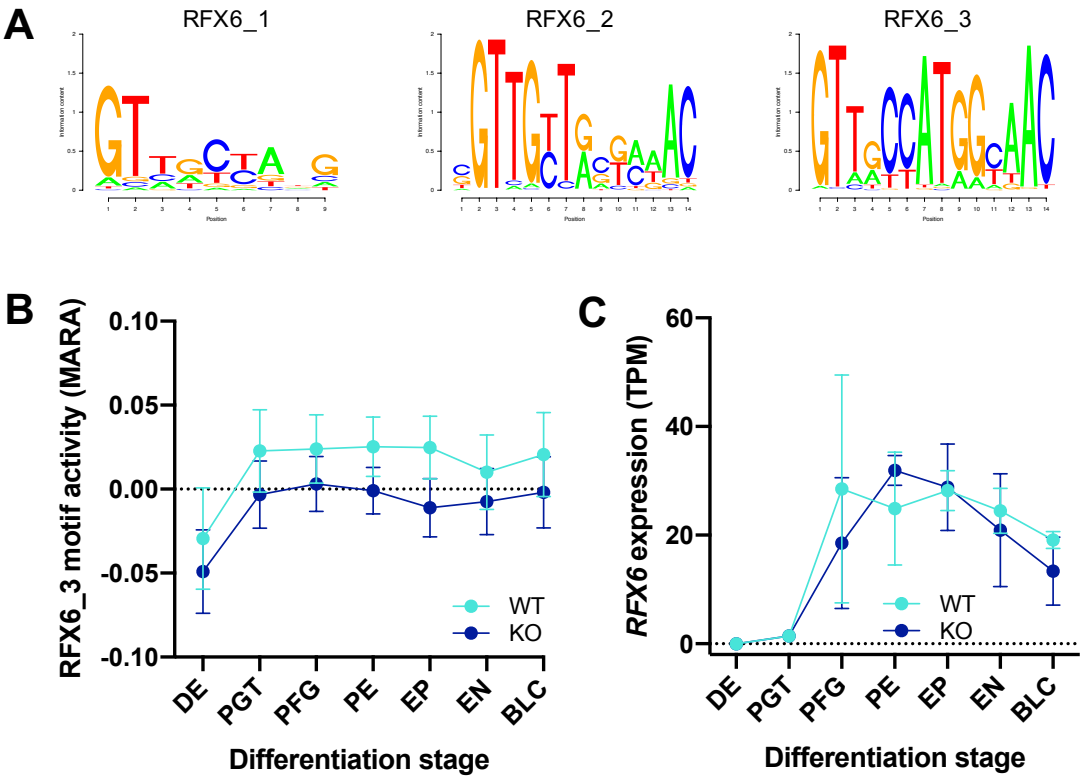


Figure 3.16: Evaluation of RFX6 motif activity in *RREB1* KO and WT lines during beta cell development (A) RFX6 motifs displayed as PWMs. (B) RFX6_3 motif activity profile for *RREB1* WT and KO lines during beta cell development calculated using MARA. (C) Expression of *RFX6* in *RREB1* WT and KO lines during beta cell development. WT n=3, KO n=4; values are displayed as mean±SD; DE, definitive endoderm; PGT, primitive gut tube; PFG, posterior foregut; PE, pancreatic endoderm; EP, endocrine precursors; EN, endocrine-like cells; BLC, beta-like cells; PWM, position weight matrix; TPM, transcripts per million; WT, turquoise; KO, darkblue.

et al. 2016), a fellow DPhil student Marta Pérez Alcántara compiled three potential PWMs for RFX6 using HOMER (Figure 3.16 A) and Mikhail Pachkov (Biozentrum, University of Basel and SBS Swiss Institute of Bioinformatics, Basel, Switzerland) kindly re-ran MARA including the three RFX6 PWMs on my transcriptome dataset. MARA's overall TF prediction did not change, RFX6_1 (Z=0.07) and RFX6_2 (Z=0.14) motifs were ranked lowest, while RFX6_3 (Z=0.94) was found at position 430 of 503 TF motifs tested. Interestingly, the PWM of RFX6_3 was very similar to the RFX2 and RFX3 binding motifs. RFX6_3 motif activity seemed to be slightly lower (though not statistically significant) in *RREB1* KO lines compared to WT clones, opposed to increased motif activity of RFX2/3 (Figure 3.16 B). *RFX6*

expression was not affected by loss of RREB1 (Figure 3.16 C).

3.3.4.10 The repressor REST revealed differential transcriptional activity between *RREB1* KO and WT lines at stages 4 to 7

In addition to RFX2 and RFX3, MARA predicted the TF REST to be one of the key regulators driving gene expression variation across differentiated *RREB1* KO and WT samples. REST motif activity was similar between *RREB1* KO

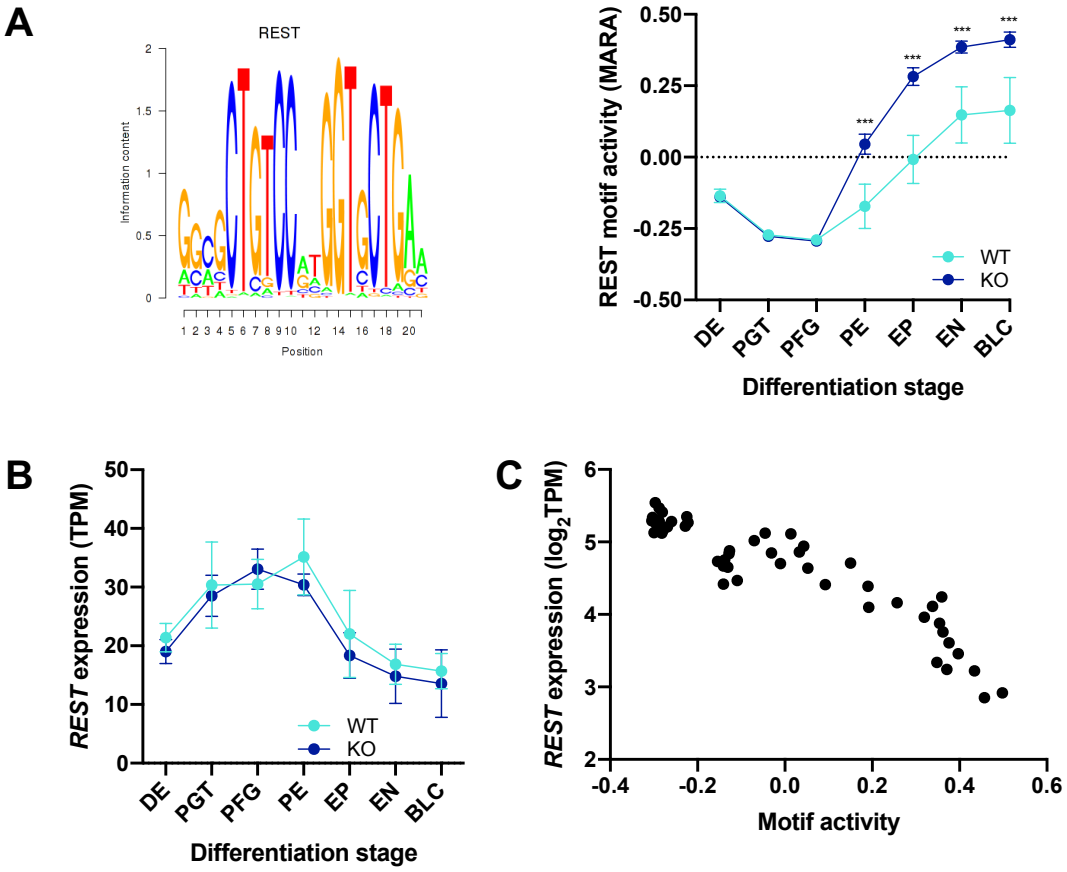


Figure 3.17: ISMARA predicted REST as one of the key TFs driving gene expression variation across *RREB1* KO and WT lines during beta cell differentiation (A) REST motif displayed as PWM and motif activity profile for *RREB1* WT and KO lines during beta cell development calculated using MARA. (B) Expression of *REST* in *RREB1* WT and KO lines during beta cell development. (C) *REST* expression correlated negatively with motif activity confirming REST as transcriptional repressor. WT n=3, KO n=4; values are displayed as mean±SD; *p*-values *<0.05, **<0.01, ***<0.001 for unpaired t-test followed by Holm-Sidak correction for multiple comparisons (A); Pearson correlation (C); DE, definitive endoderm; PGT, primitive gut tube; PFG, posterior foregut; PE, pancreatic endoderm; EP, endocrine precursors; EN, endocrine-like cells; BLC, beta-like cells; PWM, position weight matrix; TPM, transcripts per million; WT, turquoise; KO, darkblue.

and WT lines for the first three differentiation stages, then started to increase more in *RREB1* KO lines, thus it was significantly higher in *RREB1* KO lines compared to WT clones for stages 4 to 7 (PE q=7.71x10^{-6}, EP q=4.64x10^{-8}, EN q=1.96x10^{-6}, BLC q=1.08x10^{-6}) (Figure 3.17 A).

Loss of RREB1 did not affect expression of *REST* during beta cell development (Figure 3.17 B). In both *RREB1* KO and WT lines *REST* expression levels peaked in pancreatic progenitors (PFG and PE) before they gradually decreased in differentiated endocrine cells (EP through to BLC) in concordance with previous reports (D Martin et al. 2015). The negative correlation of *REST* mRNA levels and motif activity supported a transcriptional repressor role for REST (r=-0.9233, p<1.0x10^{-15}) (Figure 3.17 C). Indeed, REST target genes revealed reciprocal expression patterns compared to the *REST* expression profile during beta cell differentiation. Among those several were differentially expressed between *RREB1* KO and WT clones (Table 3.6). These included genes involved in insulin sensing, secretion and exocytosis (*SNAP25* (encoding the synaptosomal-associated protein 25) and *SYT4* (encoding the Ca^{2+}-sensitive synaptotagmin IV), both part of the SNARE complex, *SCG2* (encoding secretogranin 2, which is involved in the formation of secretory vesicles) (Hohl et al. 2005), *GCK* (encoding the glucose sensor glucokinase), *NRXN1* and *NRXN2* (encoding neurexins) (Mosedale et al. 2012) as well as *GRIA1*, *GRIA2* and *GRIK2* (encoding subunits of glutamate receptor channels) (ZY Wu et al. 2012)) (Figure 3.18 A), genes with anti-apoptotic, pro-survival activity in beta cells (*PTPRN* (Mziaut et al. 2008), *CDK5R2* (neuron and beta cell specific activator of kinase CDK5) (Tang et al. 1995; D Martin et al. 2012) and *MAPK8IP1* (scaffold protein interacting with JNK pathway, protecting beta cells against apoptosis) (Bonny et al. 2000; Haefliger et al. 2003) (Figure 3.18 B), a gene important for cell-to-cell communication (*GJD2* (gap junction protein that controls insulin secretion)) (Klee et al. 2011; Hohl et al. 2005; D Martin et al. 2003; Head et al. 2012) (Figure 3.18 C) and genes encoding TFs and proteins involved in endocrine cell differentiation (*NEUROG3*, *NEUROD1*, *RFX3*, *MYT1* (G Gu et al. 2004; S Wang et al. 2007; S Wang et al. 2008), *MFNG* (Svensson et al.

Table 3.6: Subset of REST target genes differentially expressed between *RREB1* WT and KO lines at seven stages of beta cell development

Gene	DE	PGT	PFG	PE	EP	EN	BLC
SNAP25					q= 1.55x10^{-13}	q=3.85x10^{-5}	q=0.0027
SYT4				q= 1.84x10^{-4}	q= 2.52x10^{-17}		
SCG2				q=4.27x10^{-4}	q=8.01x10^{-18}	q=2.54x10^{-4}	q=0.0030
GCK		q=2.80x10^{-12}	q=4.07x10^{-8}	q=2.08x10^{-10}	q=5.99x10^{-11}		
NRXN1	q=1.17x10^{-16}	q=2.39x10^{-8}	q=7.64x10^{-14}	q= 0.0076	q= 2.76x10^{-7}	q=3.16x10^{-6}	q=4.13x10^{-5}
NRXN2	q= 2.04x10^{-80}	q=3.15x10^{-94}	q=1.75x10^{-4}	q=2.86x10^{-43}	q= 1.35x10^{-44}	q=9.29x10^{-7}	q=4.19x10^{-6}
GRIA1				q= 0.0227	q= 1.08x10^{-35}	q=0.0138	
GRIA2					q= 1.85x10^{-17}	q=4.30x10^{-11}	q=4.61x10^{-16}
GRIK2	q=8.48x10^{-11}	q=0.0529		q= 0.0023	q= 3.47x10^{-12}	q=0.0066	
PTPRN				q= 0.0210	q= 1.18x10^{-4}		
CDK5R2				q=5.28x10^{-4}	q= 2.25x10^{-9}		q= 0.0774
MAPK8IP1			q= 0.0299	q= 0.0018	q= 5.23x10^{-7}	q=0,03019	q=2.70x10^{-4}
GJD2				q= 0.0212	q= 6.19x10^{-4}	q=0.0583	q=0.0127
MYT1	q=2.61x10^{-5}			q= 9.66x10^{-5}	q= 1.99x10^{-11}	q= 1.12x10^{-5}	q=4.24x10^{-6}
MFNG				q= 1.36x10^{-5}	q=0.0057		
DNER				q= 1.40x10^{-7}	q= 8.92x10^{-9}	q=0.0110	q=9.74x10^{-4}
CELSR3			q= 1.61x10^{-4}		q= 1.22x10^{-7}		

From top: genes involved in insulin exocytosis, genes with pro-survival activity in beta cells, gene involved in cell-to-cell communication, TFs and proteins involved in endocrine differentiation. DE, definitive endoderm; PGT, primitive gut tube; PFG, posterior foregut; PE, pancreatic endoderm; EP, endocrine precursors; EN, endocrine-like cells; BLC, beta-like cells.

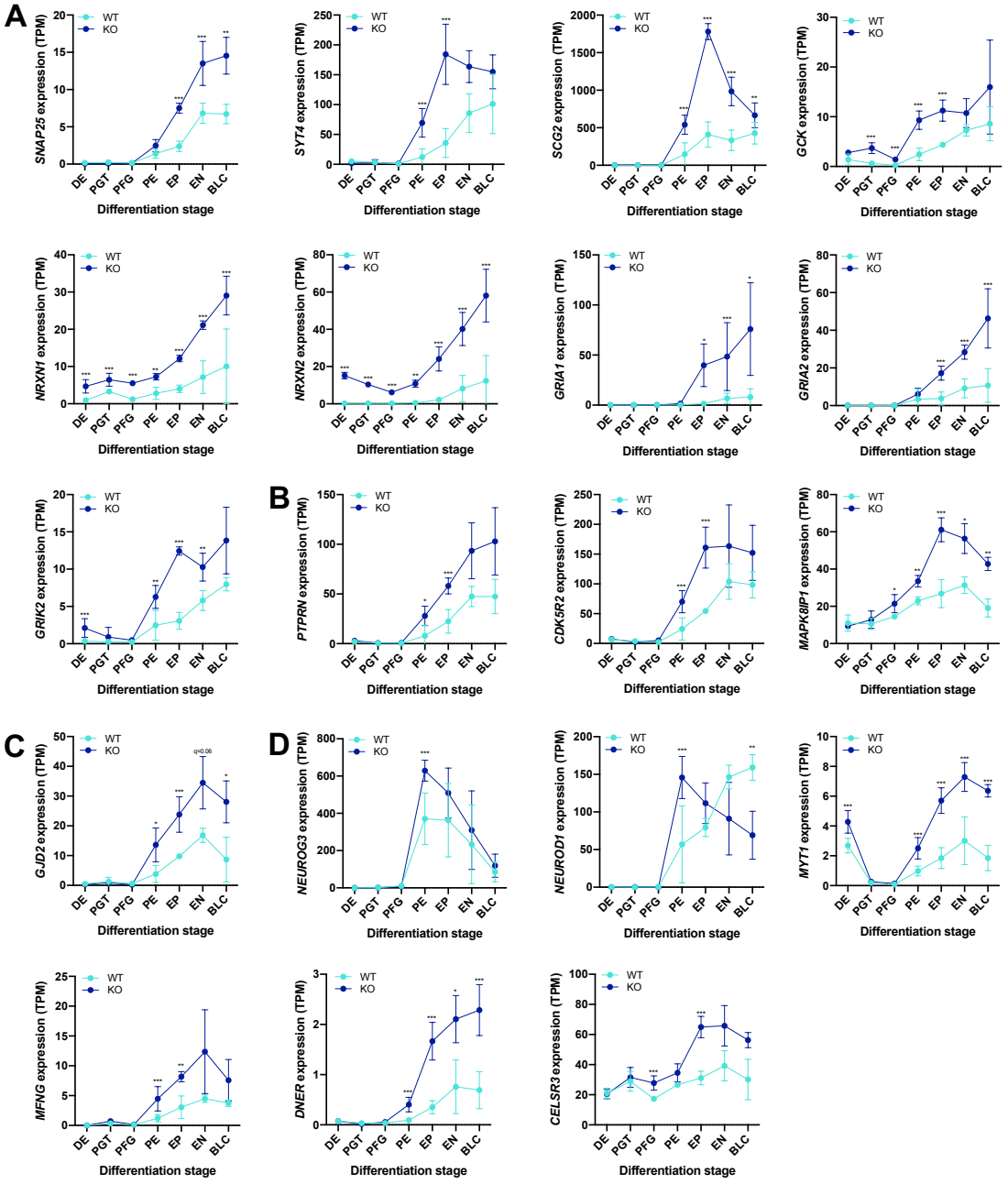


Figure 3.18: REST target genes differentially expressed between differentiated *RREB1* KO and WT lines REST target genes involved in insulin exocytosis (A), beta cell survival (B), cell communication (C), and endocrine differentiation (D) were differentially expressed between *RREB1* KO and WT lines. WT n=3, KO n=4; values are displayed as mean±SD ; p-values *<0.05, **<0.01, ***<0.001 for Wald test followed by Benjamini-Hochberg adjustment; DE, definitive endoderm; PGT, primitive gut tube; PFG, posterior foregut; PE, pancreatic endoderm; EP, endocrine precursors; EN, endocrine-like cells; BLC, beta-like cells; TPM, transcripts per million; WT, turquoise; KO, darkblue.

2009; Y Xu et al. 2006), *DNER* (Hald et al. 2012) and *CELSR3* (Cortijo et al. 2012)) (Figure 3.18 D).

3.3.4.11 RREB1 motif activity negatively correlated with REST motif activity

RREB1 motif activity (Z=1.88) was ranked at position 110 out of 503 TF motifs tested by MARA. While for stages 1 to 3 motif activity was almost identical to those observed in WT lines, diverging motif activity between *RREB1* KO and WT lines was first observed at stage 4 (PE, q=0.0725), which coincided with the first observed difference in REST motif activity. As motif activity slightly increased at stage 7 in *RREB1* WT, but not KO lines, transcriptional activity in *RREB1* KO-derived BLCs was significantly lower compared to WT BLCs (q=0.0026) (Figure 3.19 A). Interestingly, the RREB1 motif activity pattern was almost exactly reciprocal to the REST motif activity profile and Pearson correlation revealed that motif activities were negatively correlated (r=-0.9471, p=2.82x10^{-7}) (Figure 3.19 B).

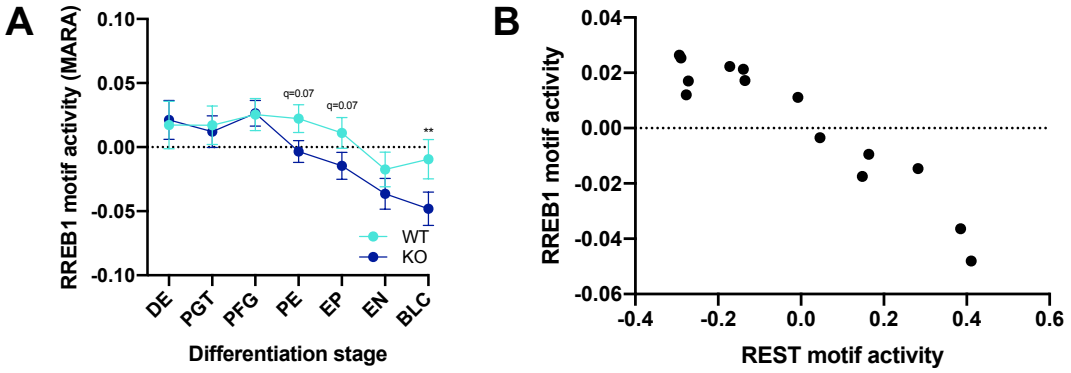


Figure 3.19: Characterisation of RREB1 motif activity identified by MARA (A) RREB1 motif and its activity profile during beta cell development for *RREB1* WT and KO lines calculated using MARA. (B) RREB1 and REST motif activity profiles negatively correlated. WT n=3, KO n=4; values are displayed as mean±SD ; p-values **<0.01 for unpaired t-test followed by Holm-Sidak correction for multiple comparisons (A) and Pearson correlation (B); DE, definitive endoderm; PGT, primitive gut tube; PFG, posterior foregut; PE, pancreatic endoderm; EP, endocrine precursors; EN, endocrine-like cells; BLC, beta-like cells; WT, turquoise; KO, darkblue.

3.3.4.12 Weighted gene co-expression network analysis identified subsets of co-expressed genes differentially expressed between *RREB1* KO and WT lines during *in vitro* beta cell development

To assess whether loss of RREB1 affected transcriptional networks initiating and regulating beta cell development, weighted gene co-expression network analysis (WGCNA) was performed on 16,876 protein-coding genes and lincRNAs (see Section 2.5.5). This approach identifies modules of co-expressed genes, such as genes with a similar expression profile during *in vitro* beta cell differentiation. As functionally related genes are often co-expressed and tend to be regulated by the same upstream TFs, such a module is likely to represent a transcriptional network or pathway (Latchman 1995; Davidson 2001).

A total of 19 co-expressed modules were identified, containing between 48 and 4,000 genes. Evaluation of module eigengene (ME) expression revealed that stage 5 (EP) contained the most modules (11) significantly different between *RREB1* KO and WT samples, shortly followed by stage 4 (PE) with seven differentially expressed ME modules (Table 3.7).

The grey module (2,603) showed the strongest difference in ME expression between *RREB1* KO and WT samples (Table 3.7). It was characterised by opposite ME directions across all developmental stages and contained *RREB1* (Figure 3.20 A).

RREB1 predicted target genes were enriched in the blue module (3,186 genes) (q=8.12x10 $^{-53}$) which showed decreasing ME expression along the seven differentiation stages indicating that this module likely contained genes whose expression was not necessary for the later stages of beta cell development or potentially even disallowed in beta cells (Figure 3.20 B). Indeed, *REST*, a forbidden gene in mature beta cells, was part of the blue module. Furthermore, *SOX17* and *HHEX*, encoding the TF hematopoietically-expressed homeobox protein, which plays a key role in the formation and differentiation of DE (Rankin et al. 2011; Fisher et al. 2017) and later gets restricted to delta cells (J Zhang et al. 2014), were found in the blue module. Interestingly, *HHEX* transcript levels were significantly higher in *RREB1*

Table 3.7: Module eigengenes identified in WGCNA for *RREB1* WT and KO lines at seven stages of beta cell development

MEs	# genes	DE	PGT	PFG	PE	EP	EN	BLC
black	443	q=0.0016			q=0.0454	q=0.0058	q=0.0136	
blue	3,186					q=0.0278		
brown	1,914							
cyan	183	q=0.0380		q=0.0007	q=0.0435	q=0.0051		
green	657							
greenyellow	266	q=0.0385				q=0.0169		
grey	2,603	q=0.0048	q=0.0035	q=0.0007	q=0.0017	q=0.0006	q=0.0291	q=0.0135
grey60	59					q=0.0161		
lightcyan	100			q=0.0018	q=0.0215	q=0.0127	q=0.0021	
lightgreen	48							
magenta	280				q=0.0456	q=0.0278	q=0.0367	q=0.0488
midnightblue	102					q=0.0268	q=0.0207	
pink	414							
purple	279	q=0.0058	q= 0.0377		q=0.0140	q=0.0088	q=0.0109	q=0.0369
red	502				q=0.0111			
salmon	190							
tan	242							
turquoise	4,000					q=0.0062		
yellow	1,408							

16,876 protein-coding genes and lincRNAs tested; unpaired t-test followed by Holm-Sidak correction for multiple comparisons; q<0.05; DE, definitive endoderm; PGT, primitive gut tube; PFG, posterior foregut; PE, pancreatic endoderm; EP, endocrine precursors; EN, endocrine-like cells; BLC, beta-like cells.

KO than WT lines at stage 1 (DE) (q=0.0078, $\log_2$FC=0.6748), consistent with higher *SOX17* expression mentioned previously (see Figure 3.6).

REST target genes implicated in endocrine cell differentiation (*NEUROD1*, *DNER*, *MFNG*), insulin secretion (*SNAP25*, *SYT4*, *VAMP2*, *SCG2*, *PAM*, *GCK*, *NRXN1*, *GRIA2*, *CHGA*, *GPR56*, *SST*), insulin maturation (*PCSK1*), beta cell survival (*MAPK8IP1* and *CDK5R2*), and cell connectivity (*GJD2*) on the other hand were found in the turquoise module which revealed a reciprocal expression profile compared to the blue module (increasing ME expression throughout development) and contained the largest number of genes, namely 4,000 (Figure 3.20 C). Consistent with this, endocrine progenitor and endocrine marker genes (see Section 1.4.5.2) were also enriched in the turquoise module (q=4.56x10^{-18} and q=3.98x10^{-19}, respectively). The top two iRegulon-predicted upstream regulators of genes in the turquoise module were REST (NES=3.0-5.4) and the RFX TF family (NES=3.1-3.7) signifying their role as endocrine TFs. Additionally, this module was enriched for biological terms including 'nervous system development' (GO, q=1.55x10^{-16}) and 'synaptic signalling' (GO, q=1.87x10^{-16}), 'regulation of insulin secretion' (GO, q=0.0021; Reactome, q=5.57x10^{-5}) and 'insulin secretion' (GO, q=1.35x10^{-4}; KEGG, q=3.65x10^{-7}), 'calcium signalling pathway' (KEGG, q=0.0011) and 'cAMP signalling pathway' (KEGG, q=0.0023).

RFX2 and *RFX6* were found in the brown module (Figure 3.20 D). ME expression gradually increased from stage 1 (DE) to 4 (PE) before becoming stationary, suggesting that the brown module likely contained genes important in early stages of development. Consistent with this, the brown module was enriched for pancreatic progenitor signature genes (q=2.80x10^{-13}) (see Section 1.4.5.1). Among those several were differentially expressed between *RREB1* KO and WT clones (q<0.01). While *ONECUT2* (stages 5 and 6) and *TGFB2* (stage 5) levels were higher, expression of *HNF4A* (stage 2), *FOXP4* (stage 2), *CDX2* (stages 3-5) and *IGF2* (stages 3-7) was significantly lower in *RREB1* KO compared to WT cells. *CDX2* encodes a nuclear homeobox TF which is involved in intestinal cell differentiation and present in epithelial enterocytes, therefore considered as an intestinal marker

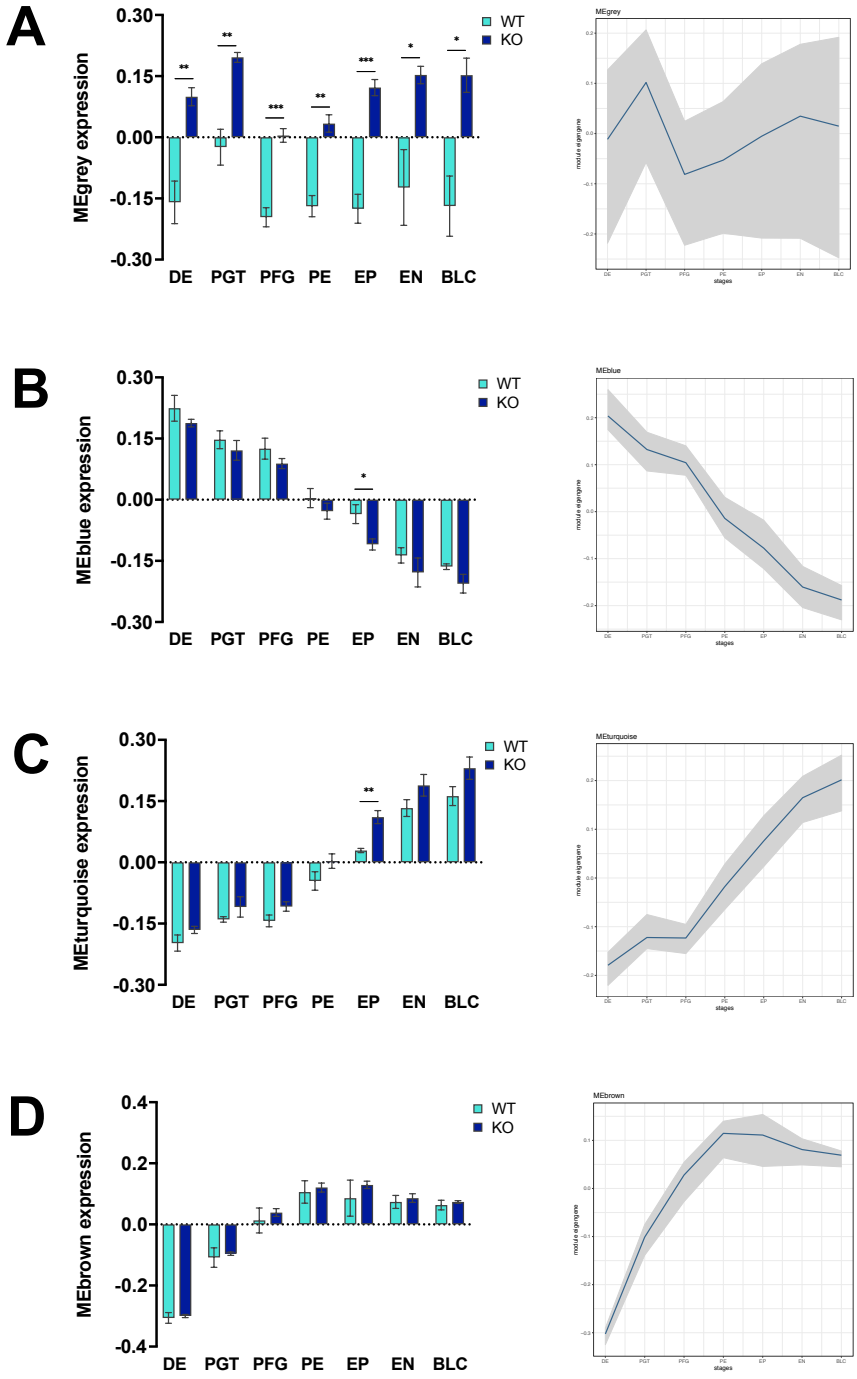


Figure 3.20: Analysis of modules of co-expressed genes using WGCNA Bar plots (left) and ribbon plots (right) showing ME expression of selected modules in *RREB1* KO and WT lines throughout beta cell development. (A) MEgrey. (B) MEblue. (C) MEturquoise. (D) MEbrown. WT n=3, KO n=4; values are displayed as mean±SD ; *p*-values $^{*}<0.05$, $^{**}<0.01$, $^{***}<0.001$ for unpaired t-test followed by Holm-Sidak correction for multiple comparisons; DE, definitive endoderm; PGT, primitive gut tube; PFG, posterior foregut; PE, pancreatic endoderm; EP, endocrine precursors; EN, endocrine-like cells; BLC, beta-like cells; WT, turquoise; KO, darkblue.

(Saad et al. 2011). Higher expression in *RREB1* WT lines (PFG q=0.0013, PE q=2.74x10^{-32}, EP q=9.82x10^{-19}) could indicate that a higher proportion of WT than KO cells followed the intestinal differentiation path. Additionally, *IGF2* which encodes the protein hormone insulin-like growth factor 2 that is structurally similar to insulin has been shown to decrease expression levels of TFs critical in beta cell development and function, such as *FOXA2*, *PDX1*, *NEUROD1* and *HNF4A*, when overexpressed in murine beta cells potentially leading to beta cell dedifferentiation (Casellas et al. 2015). Higher expression levels of *IGF2* concurrent with lower expression levels of *FOXA2*, *HNF4A* and *NEUROD1* in *RREB1* WT lines could therefore indicate a lower percentage of differentiating endocrine cells in WT compared to *RREB1* KO lines. Upstream regulators of genes in the brown module predicted by iRegulon contained motifs bound by hepatocyte nuclear factors including FOXA1 (NES=3.20), FOXA2 (NES=3.21), HNF1A and HNF1B as well as PDX1 or PAX4 (NES=3.01-3.59), TFs implicated in early beta cell development. *FOXA1*, *HNF1A*, *PDX1* and *PAX4* were also found in the brown module, supporting them as transcriptional regulators of genes in this module.

Similar to RREB1, target genes of both RFX2 and RFX3 were mostly found in the blue module (q=1.15x10^{-15} and q=1.27x10^{-19}, respectively). Additional target genes, albeit fewer, clustered in the brown module for RFX2 (q=1.26x10^{-5}) or in the brown and black modules for RFX3 (q=4.91x10^{-6} and q=1.25x10^{-6}, respectively).

3.4 Discussion

This chapters describes the generation and characterisation of various *RREB1* KO and WT hiPSC lines and their differentiation along the endocrine lineage into BLCs. Assessment of key developmental marker expression in differentiated cells at seven distinct stages of *in vitro* beta cell differentiation revealed variable differentiation efficiencies amongst *RREB1* WT clones. While marker expression profiles were comparable between individual clones, *RREB1* WT clone #3 consistently showed higher expression levels compared to WT clones #1 and #2 and clustered closer with *RREB1* KO clones in PCA from stage 3 (PFG) onwards. Interestingly, *RREB1* WT clone #3 only formed a sufficient percentage of DE cells (>60%), when plated at a lower density (0.8 x 10^6 vs 1.3 x 10^6 hiPSCs/well) which might have affected its differentiation capacity. To account for this, transcriptome data was normalised using RUVSeq, which improved clustering of distinct differentiation stages and resulted in a clearer separation of *RREB1* WT and KO lines.

Loss of RREB1 modifies early endocrine differentiation

Evaluation of gene expression and TF networks at individual developmental stages suggested that loss of RREB1 had predominantly affected early endocrine differentiation (stages 4 and 5). Differential gene expression analysis detected the most DEGs (2,611) at stage 5, of which the majority was up-regulated in the *RREB1* KO lines, indicating that RREB1 might act as transcriptional repressor in EP cells, as seen in various other cellular systems (S Zhang et al. 2003; Date et al. 2004; RL Chen et al. 2010; Flajollet et al. 2009; Mukhopadhyay et al. 2007; Milon et al. 2010; Zou et al. 2011). Endocrine progenitor markers (*NEUROG3*, *NKX2.2*, *NEUROD1*, *MFNG*, *RFX3*) were enriched among DEGs up-regulated in the *RREB1* KO lines at stages 4, 5 (most pronounced) and 6, while pancreas progenitor markers (*CPA2*, *HNF1B*, *NOTCH1*) were down-regulated suggesting loss of RREB1 might have resulted in an enhanced proportion of endocrine cells. In line with this, WGCNA detected the most striking ME expression differences between *RREB1* KO and WT lines at stage 5.

The turquoise module, representative of endocrine progenitor and endocrine genes, showed significantly different ME expression between *RREB1* WT and KO clones and was enriched for terms relating to 'synaptic transmission', 'hormone transport' and 'insulin secretion' as well as the regulation of these processes. Genes involved in beta cell survival, cell-to-cell communication, insulin exocytosis and endocrine differentiation were markedly higher expressed in *RREB1* KO than WT clones, signifying the potential positive effect of RREB1 loss on endocrine differentiation.

RREB1 may affect beta cell development through manipulation of RFX2 or RFX3

Predicted upstream regulators of genes in the turquoise module as well as those differentially expressed at individual developmental stages included the TF REST as well as the RFX TF family, in particular RFX2 and RFX3.

RFX TFs are conserved in yeast, flies and vertebrates (Emery et al. 1996b) and members of this TF family have been demonstrated to be crucial for various developmental processes (Horvath et al. 2009; Kistler et al. 2009; Choksi et al. 2014; Magnani et al. 2015; Y Wu et al. 2016). For example, RFX2 is highly expressed in testis (Wolfe et al. 2006) and a key regulator of spermatogenesis in mice (Wolfe et al. 2006; Horvath et al. 2009; Y Wu et al. 2016). Although a role for RFX2 in beta cell development or function has not been reported in the literature so far, gene co-expression and promoter motif analysis predicted RFX2-regulated modules involved in ciliogenesis and synaptic transmission (Ma et al. 2017). For example the RFX2 target gene *CAMK2A* which encodes the Ca^{2+}/calmodulin-dependent protein kinase II has been shown to modulate glucose homeostasis and insulin secretion by regulating Ca^{2+} levels in pancreatic beta cells (Tabuchi et al. 2000; Dadi et al. 2014). Moreover, beta cell *SLC17A7* which encodes a vesicular glutamate transporter is involved in glutamate uptake into secretory vesicles and incretin-induced insulin secretion (Gheni et al. 2014; Murao et al. 2017). *RREB1* KO lines were characterised by markedly higher RFX2/3 motif activity as well as *RFX2*

gene expression throughout all seven stages of beta cell development. Putative RREB1 binding site motifs have been predicted in the *RFX2* promoter (Matys et al. 2006), hence it is likely that RREB1 acts as a potential transcriptional repressor of *RFX2*. Therefore, RFX2 may represent a novel transcriptional regulator of endocrine differentiation, regulated by RREB1.

The TF RFX3 has been characterised more extensively. It plays a key role in ciliogenesis, i.e. regulates genes required for the formation, growth, maintenance and function of cilia (Bonnafe et al. 2004; El Zein et al. 2009) and has been shown in mice to be expressed in pancreatic *Neurog3*-positive endocrine progenitors as well as developing and mature endocrine cells (Ait-Lounis et al. 2007). RFX3 deficiency during endocrine development manifested itself in reduced numbers of alpha and beta cells with a concomitant increase in PP cells in perinatal *Rfx3* KO mice supporting a crucial role for RFX3 in pancreatic endocrine cell differentiation (Ait-Lounis et al. 2007). Early pancreatic development and beta cell lineage specification were unaffected by loss of RFX3. RFX3-deficient embryos revealed similar expression levels of *Pdx1*, *Pax4*, *Nkx2.2*, *Nkx6.1* and *Neurod1* to control littermates (Ait-Lounis et al. 2007). However, the number of endocrine cells co-expressing NKX6.1 and insulin or PDX1 and insulin was markedly reduced and assessment of beta cell specific marker expression revealed reduced levels of *Slc2a2*, encoding the glucose transporter GLUT2, and *Gck* encoding the glucokinase enzyme, both key components of the beta cell glucose-sensing and insulin secretion machinery, implying incomplete or blocked beta cell differentiation and maturation in *Rfx3* KO embryos (Ait-Lounis et al. 2010). The developmental defect led to smaller islets, characterised by altered alpha and beta cell composition, reduced total pancreatic insulin content and impaired glucose tolerance in adult *Rfx3* KO mice, indicating that RFX3 plays a unique role in both beta cell differentiation and function in mice (Ait-Lounis et al. 2007). A recent study identified *RFX3* as direct target gene of PDX1 in human iPSC-derived pancreatic progenitors (X Wang et al. 2018). *RFX3* expression was up-regulated during *in vitro* differentiation, supporting a role for RFX3 during human pancreas development (X Wang et al. 2018).

Loss of RREB1 resulted in markedly higher RFX2/3 motif activity and *RFX3* gene expression throughout all seven stages of beta cell development. As RFX3 has been shown to be crucial for endocrine cell development, derepression of *RFX3* in *RREB1* KO lines might have led to enhanced endocrine differentiation. In line with this, expression of *GCK* which has been shown to be a direct target gene of RFX3 in humans and mice (Ait-Lounis et al. 2010) was significantly higher in *RREB1* KO lines during stages 2 to 5.

In addition to RFX3, the only other RFX protein associated with pancreas development so far is RFX6, which is required for the formation of endocrine cells (alpha, beta, delta, and PP cells) and the maintenance of a functional beta cell phenotype in both human and mouse (Smith et al. 2010; Soyer et al. 2010; Piccand et al. 2014; Chandra et al. 2014). MARA did not identify an effect of loss of RREB1 on RFX6 function, which is interesting as one of the predicted RFX6 binding motifs showed great similarity to the RFX2 and RFX3 motifs and RFX6 has been shown to be the islet-specific TF among RFX proteins (Varshney et al. 2017). Intriguingly, in pancreatic endoderm cells this particular RFX motif is enriched at poised enhancers, which become active in islets, thus RFX TFs have been suggested to be important for the activation of islet enhancers during endocrine cell differentiation (A Wang et al. 2015). Further investigation is needed to distinguish which of the RFX TFs binds to this particular binding motif during endocrine development. In contrast to RFX2 and RFX3, there was less public data available on RFX6 TF binding motifs and target genes, thus it cannot be ruled out that a potential impact of RREB1 deficiency on RFX6 was not picked up during *in vitro* beta cell differentiation due to insufficient data availability.

RREB1 and REST - a tale of double repression during beta cell development?

The transcriptional repressor REST, predicted to be a potential upstream regulator of DEGs between *RREB1* KO and WT lines, plays a unique role during pancreas

development and in the regulation of genes involved in the beta cell exocytotic machinery. It has been shown to prevent the transcription of neuronal genes in most cell types, except in neurons and pancreatic beta cells, where it is hardly expressed (Chong et al. 1995; Atouf et al. 1997). Neuronal genes expressed in beta cells include genes encoding synaptic vesicle proteins (e.g. SNARE proteins), neurotransmitters (glutamate, acetylcholine) (Satin et al. 1998) and neurotransmitter-related enzymes (Yokoi et al. 2016). Ectopic expression of human REST in mature beta cells prevented full expression of these genes and impaired GSIS in mice (D Martin et al. 2008). During pancreas development, REST expression is confined to pancreatic progenitors. While overexpression of human *REST* in PDX-expressing progenitor cells diminished the generation of NEUROG3-positive endocrine progenitors and impaired subsequent differentiation of endocrine cells, loss of REST led to an increase of endocrine progenitor and endocrine signature genes, including *Neurod1*, *Mfng*, *Gjd2* and *Syt4* in mice (D Martin et al. 2015).

Transcriptional activity of REST began to separate progressively between *RREB1* KO and WT lines from stage 4 (PE) onwards with *RREB1* KO clones showing higher REST motif activity. For a transcriptional repressor higher motif activity implies that it was less repressive. *REST* represents a forbidden gene in mature beta cells, hence lower expression and less repression of endocrine signature genes was desired in the end stages of *in vitro* beta cell differentiation. Higher REST motif activity in *RREB1* KO compared to WT lines therefore supports the hypothesis of improved endocrine differentiation when RREB1 gets lost. Motif activities of RREB1 and REST correlated negatively and JASPAR and TRANSFAC TF binding site databases predict REST binding motifs in the *RREB1* promoter. Unlike *REST*, *RREB1* is expressed in adult beta cells (Nica et al. 2013) and therefore unlikely a disallowed gene in endocrine cells. Whether and how these two TFs interact with each other, remains to be elucidated.

Conclusions

Taken together, several analytical approaches suggest a novel role for RREB1 in endocrine differentiation. Loss of RREB1 during beta cell development led to a higher proportion of endocrine progenitors, potentially through enhancement of the NEUROG3 endocrine differentiation programme at stages 4 (PE) and 5 (EP). This might have been mediated by REST, whose transcriptional activity was negatively affected by loss of RREB1, likely leading to a derepression of *NEUROG3* gene expression. As *REST* expression was unchanged in *RREB1* KO lines, it might be more likely that REST and RREB1 interact with each other at the protein than the transcriptional level. Clarification could be provided by either ChIP-Seq for the identification of physical RREB1-DNA binding sites in the genome, or RREB1 co-immunoprecipitation for the identification of protein-protein interactions. To my knowledge, neither of these approaches have been performed genome-wide in differentiated pancreatic cells or any other cell type so far.

The RFX TF family was also affected by loss of RREB1. Increased expression and transcriptional activity of both RFX2 and RFX3 in *RREB1* KO lines suggests RREB1 as a transcriptional regulator of RFX genes, thus, derepression of *RFX2* or *RFX3* transcription in *RREB1* KO lines might be another mechanism underlying enhanced endocrine differentiation.

Having observed an effect of RREB1 deficiency on beta cell development, I was interested whether this was mirrored in mature beta cell identity and/or function. As hiPSC-derived BLCs were expected to be still functionally immature, a more suitable cellular system, the mature human beta cell line EndoC-βH1 in combination with RNA interference or CRISPR/Cas9 was used to assess the effect of RREB1 depletion on beta cell function, which will be discussed in Chapter 4.

4

Determination of the role of RREB1 in mature human beta cells

4.1 Introduction

The majority of T2D- and FG-associated GWAS loci are thought to affect beta cell function (Florez 2008; Ingelsson et al. 2010; Voight et al. 2010; Dupuis et al. 2010; Dimas et al. 2014). Common variants often share association between T2D risk and quantitative measures of islet function, supporting a central role for beta cells in T2D disease susceptibility (Dupuis et al. 2010; Morris et al. 2012; Wood et al. 2017). In addition, profiling of the epigenomic landscape in human islets has revealed that GWAS loci are enriched in regulatory elements, in particular in stretch and super enhancers (Gaulton et al. 2010; Parker et al. 2013; Pasquali et al. 2014; Gaulton et al. 2015; Thurner et al. 2018).

Chapter 3 describes how a hiPSC-based model has successfully been used to study a T2D-associated gene during human beta cell development. As *in vitro* hiPSC-derived BLCs have been shown to still be functionally immature (Basford et al. 2012; Bruin et al. 2014; Russ et al. 2015), identification and characterisation of genes and T2D variants acting via regulation of insulin secretion remain challenging

using this model. Phenotypes associated with impaired metabolic regulation of insulin secretion might be masked by immaturity and not detectable. Human beta cell lines like EndoC-βH1 have been extensively characterised and represent an authentic complementary model when studying mature beta cell function (Hastoy et al. 2018; Lawlor et al. 2019).

The unique beta cell phenotype is associated with expression of genes important for cellular function (*INS*, *GCK*, *SLC30A8*, *PDX1*, *MAFA*, *NKX6.1*) as well as a subset of genes specifically repressed in beta cells (De Vos et al. 1995; Chimienti et al. 2006; T Gao et al. 2014; Taylor et al. 2013; Swisa et al. 2017; Quintens et al. 2008; Pullen et al. 2013). Two independent studies have identified more than 60 disallowed genes in rodent beta cells (Pullen et al. 2010; Thorrez et al. 2011) including genes implicated in metabolic-secretion coupling and insulin exocytosis (*Slc161a* (Pullen et al. 2012), *Ldha* (Sekine et al. 1994)), pro-proliferative genes (*Igfbp4* (Durai et al. 2006), *Pdgfra* (H Chen et al. 2011), *Cxcl12* (Singh et al. 2011)) and *Rest*, which encodes the transcriptional repressor RE-1 silencing TF (D Martin et al. 2008). As REST *cis*-regulated genes are essential to insulin secretion (*Snap25*, *Syt4*, *Gjd2*, *Neurod1*), it has to be actively excluded from mature beta cells to ensure proper cell function (D Martin et al. 2015).

4.1.1 Experimental aims

In Chapter 3, I showed that loss of RREB1 had a positive impact on endocrine differentiation. To evaluate, whether there were any effects associated with loss of RREB1 in mature beta cells, the human beta cell line EndoC-βH1 in combination with either RNA interference-mediated transient KD of *RREB1* or CRISPR/Cas9-mediated stable KO of *RREB1* was characterised for GSIS as well as expression of beta cell signature genes (Kutlu et al. 2009; Nica et al. 2013; Muraro et al. 2016; Lawlor et al. 2017). Findings will be discussed in this chapter and compared to observations that I have made in hiPSC-derived *RREB1* KO BLCs (Chapter 3).

Whether the effect of RREB1 on endocrine cell differentiation is caused by direct transcriptional control of genes important for development or indirectly through manipulation of other TFs is unclear, as little is known about RREB1 *cis*-regulated genes due to the lack of RREB1 ChIP-Seq data. Therefore, I performed ChIP-Seq experiments in EndoC-βH1 cells overexpressing RREB1. Integration with transcriptomic data allowed me to identify genes directly regulated by RREB1 in beta cells, while integration with epigenomic annotations allowed me to assess, whether RREB1 binding sites were enriched in islet regulatory regions.

4.2 Methods

4.2.1 Transcriptional silencing of *RREB1* using RNA interference

Small interfering RNAs targeting four exonic regions in *RREB1* (exon 6/7 boundary, exon 8 (2x) and exon 9) were used to transiently decrease *RREB1* expression in EndoC-βH1 cells. KD experiments were performed as described in Section 2.1.6.

4.2.2 Generation of EndoC-βH1 *RREB1* KO cells

To generate a stable *RREB1* KO EndoC-βH1 line three sgRNAs targeting *RREB1* were taken from the TKO Library v3 (T Hart et al. 2017) and one was designed using the CRISPOR online design tool (http://crispor.tefor.net) (Haeussler et al. 2016). Individual sgRNAs were subcloned into plentiCRISPRv2, lentivirus were produced in HEK293T cells and the functional viral titer was determined as described in Section 2.2. EndoC-βH1 cells were transduced at a MOI of 10 and selected for successful sgRNA genome integration using puromycin. Drug-resistant cells were expanded and assessed for indels in the *RREB1* gene. A control cell line, referred to as EV, was engineered in parallel by transducing EndoC-βH1 cells with plentiCRISPRv2 vector, encoding Cas9, but without containing a sgRNA sequence.

4.2.3 Insulin secretion assays in EndoC-βH1

GSIS was assessed in EndoC-βH1 cells following *RREB1* KD (96h post-transfection) as well as in *RREB1* KO and EV EndoC-βH1 cells using static incubations as described in Section 2.1.9. For *RREB1* KD experiments, additional cells were collected for the analysis of *RREB1* KD efficiency using RT-qPCR. Insulin was measured using the Insulin (human) AlphaLISA Detection Kit (PerkinElmer) and

analysed as described in Section 2.1.9.

4.2.4 Transcriptome profiling

For transcriptome profiling RNA was extracted and processed from EndoC-βH1 cells following *RREB1* KD (96h post-transfection) as well as from *RREB1* KO and EV EndoC-βH1 cells as described in Section 2.5.1. Software packages for RNA-Seq data analysis included the Bioconductor packages DEXSeq (S Anders et al. 2012) and DESeq2 (Love et al. 2014) in R v3.3.3. Gene ontology and TF binding motif enrichment analysis were done using g:Profiler (Raudvere et al. 2019) and the iRegulon (v1.3) Cytoscape (v3.7.0) plugin (Janky et al. 2014), respectively. In addition, motif activity response analysis (MARA) was applied to identify key TFs mediating gene expression variation across *RREB1* KO (KD) and EV (NT) EndoC-βH1 cells using the online tool ISMARA (Integrated System for MARA) (Balwierz et al. 2014). Individual analysis methods are described in more depth in Section 2.5.

4.2.5 ChIP-Seq

For ChIP-Seq experiments, EndoC-βH1 *RREB1* KO cells were transfected with FLAG-tagged pSPORT-*RREB1* (FLAG) or pCMV6-EV (CTRL) expression plasmids as described in Section 2.6. ChIP was performed using the ChIP-IT High Sensitivity® kit (Active Motif), aFLAG polyclonal antibody (1:50, Cell Signalling) and MagReSyn® Protein A beads (ReSyn Biosciences). Preparation of ChIP-Seq libraries and sequencing was done by the Oxford Genomics Centre at the WCHG and data was analysed by Dr Toryn Poolman (OCDEM).

4.3 Results

4.3.1 Reduced *RREB1* levels affect cellular insulin content in EndoC-βH1 cells

RNA interference was used to determine the effect of reduced *RREB1* levels on beta cell activity. siRREB1-targeted cells (*RREB1* KD) revealed significantly lower levels of *RREB1* compared to EndoC-βH1 control cells (transfected with siNT (non-targeting siRNA mix), *RREB1* NT) (-33.8±11.3%, p=0.0169) (Figure 4.1 A).

Partial RREB1 depletion did not affect basal (1mM glucose) or glucose-stimulated (20mM) secretion of insulin (Figure 4.1 B). However, cellular insulin content was significantly reduced in *RREB1* KD EndoC-βH1 cells (-31.3±18.0%, p=0.0420) (Figure 4.1 C), which could not be attributed to a decrease in cell numbers (p=0.4836) (Figure 4.1 D). As a consequence, assessment of insulin secretion measures, expressed as percentage of content, revealed higher insulin release in response to a glucose stimulus in *RREB1* KD beta cells (x1.50±0.06, p=0.0636) (Figure 4.1 E), which was not observed when secretion measures were adjusted for cell count (Figure 4.1 F). In line with reduced cellular insulin content, *INS* gene expression was significantly lower in *RREB1* KD compared to NT EndoC-βH1 cells (-15.6±5.7%, p=0.0053) (Figure 4.1 G).

4.3.2 Gene expression profiling in EndoC-βH1 cells characterised by reduced *RREB1* levels

4.3.2.1 Principal component and correlation analyses identify an outlier among EndoC-βH1 NT samples

To investigate, whether expression of genes implicated in insulin processing and/or the exocytotic machinery, potentially underlying the insulin content phenotype, were affected by partial depletion of RREB1, EndoC-βH1 cells were transfected with either siRREB1 or siNT (control) at six consecutive cell passages to obtain independent

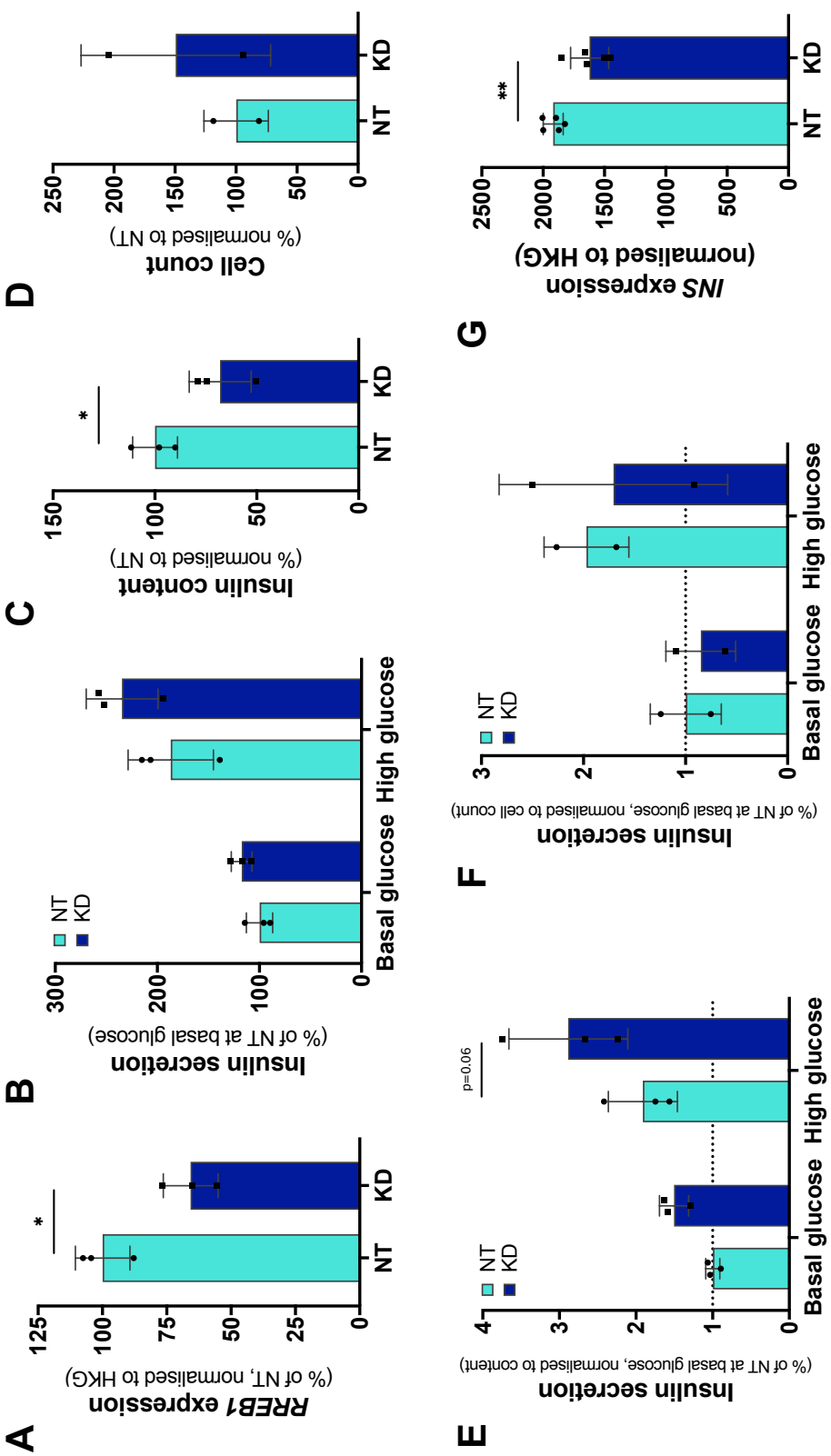


Figure 4.1: Effect of reduced *RREB1* levels on beta cell function EndoC-βH1 cells characterised by reduced levels of *RREB1* (A) were assessed for glucose-stimulated insulin secretion (B), cellular insulin content (C), and cell numbers (D). Insulin secretion measures expressed as percentage of content (E) or adjusted for cell count (F). *INS* expression was down-regulated in *RREB1* KD EndoC-βH1 cells (G). n=2-5; values are displayed as mean±SD; p-values $^{*}<0.05$, $^{**}<0.01$ for unpaired t-test (A, C, D and G) or two-way ANOVA followed by Sidak's multiple comparisons test (B, E and F); HKG, housekeeping genes; NT, (non-targeting control), turquoise; KD (knockdown), darkblue.

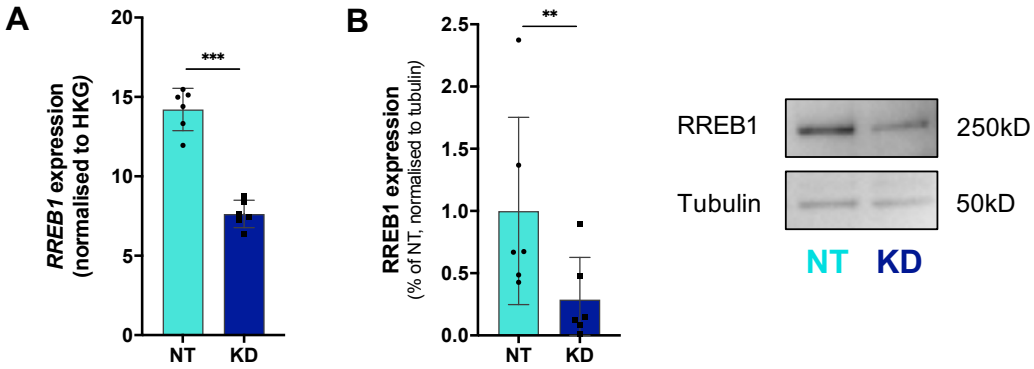


Figure 4.2: RNA interference caused lower *RREB1* transcript and RREB1 protein expression in EndoC-βH1 cells RNA interference-mediated inhibition of *RREB1* mRNA (A) and RREB1 protein (B) expression in EndoC-βH1 cells assessed 96h after transfection. n=6; values are displayed as mean±SD; p-values **<0.01, ***<0.001 for unpaired t-test; HKG, housekeeping genes; NT, turquoise; KD, darkblue.

biological replicates. For each transfected KD/NT pair, RNA samples were collected 96h after transfection and processed for RNA-Seq. siRNA-mediated KD of *RREB1* caused on average a 46.3±4.1% reduction in *RREB1* transcript levels compared to NT EndoC-βH1 cells (p=1.39x10^{-6}) (Figure 4.2 A). At the protein level this led to a reduction of RREB1 expression of 78.3±13.1% (p=0.0085) (Figure 4.2 B).

A total of 12 RNA-Seq libraries were sequenced to a mean read depth of 19.6±0.6 million reads per sample. Evaluation of *RREB1* expression revealed much lower transcript levels in EndoC-βH1 NT sample #2 compared to all other NT as well as *RREB1* KD specimen (Figure 4.3 A). The first two principal components (PC1, PC2) were calculated using normalised gene counts of all 12 *RREB1* KD and NT EndoC-βH1 samples to visualise sample-to-sample distances. While PC1 clearly separated *RREB1* KD from NT samples, PC2 revealed pair-wise variability across samples. NT #2 clustered furthest away from the other five NT samples and did not follow the pair-specific clustering pattern observed for the other five *RREB1* KD/NT pairs (Figure 4.3 B). Normalised count data were also visualised in a heatmap to get an overview over the similarities and diversities between samples. Again, NT #2 was strikingly dissimilar to all other 11 samples, illustrated by the red column (Figure 4.3 C). Taken together, a first assessment of transcriptome data identified NT #2 as likely outlier. As PCA revealed a pair-wise separation between

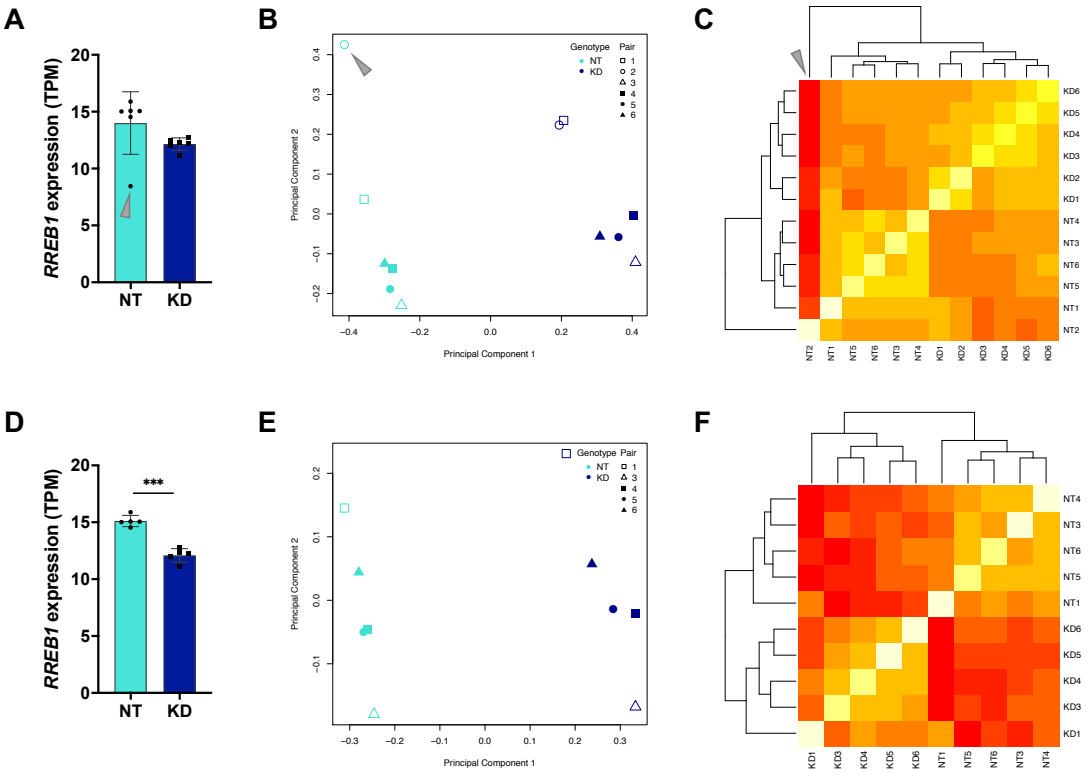


Figure 4.3: Principal component and correlation analyses of the transcriptomes of six *RREB1* KD and six NT samples (A) Evaluation of *RREB1* expression in *RREB1* KD and NT EndoC-βH1 samples showed a reduction for *RREB1* in NT sample #2 (grey arrowhead). (B) The first two principal components were calculated using normalised gene counts of all 12 *RREB1* KD and NT EndoC-βH1 samples. NT sample #2 (grey arrowhead) did not follow the same cluster pattern as the other KD/NT pairs. (C) Heatmap of normalised gene counts showing strong dissimilarity between NT #2 and all other 11 samples (grey arrowhead) (increasing dissimilarity from lightyellow to darkred). (D) *RREB1* expression in *RREB1* KD and NT EndoC-βH1 samples after removal of KD/NT pair #2. Removal of KD/NT pair #2 improved clustering of samples in PCA (E) and revealed a stronger diversity between remaining *RREB1* KD and NT samples (F). n=5-6; values are displayed as mean±SD; p-values ***<0.001 for Wald test followed by Benjamini-Hochberg adjustment (D); TPM, transcripts per million; NT, turquoise; KD, darkblue.

samples originating from different cell passages, transcriptome data of both NT #2 and *RREB1* KD #2 were removed for further analysis. Removal of KD/NT pair #2 led to a significant reduction in *RREB1* transcripts levels in *RREB1* KD samples (Figure 4.3 D) (q=9.87x10^{-10}, $\log_2$FC=-3.024), improved sample clustering in PCA (Figure 4.3 E) and revealed stronger dissimilarities between remaining *RREB1* KD and NT samples (Figure 4.3 F).

Assessment of exon usage in NT and *RREB1* KD samples confirmed partial deficiency of *RREB1* in KD specimen. *RREB1* gene and transcript expression

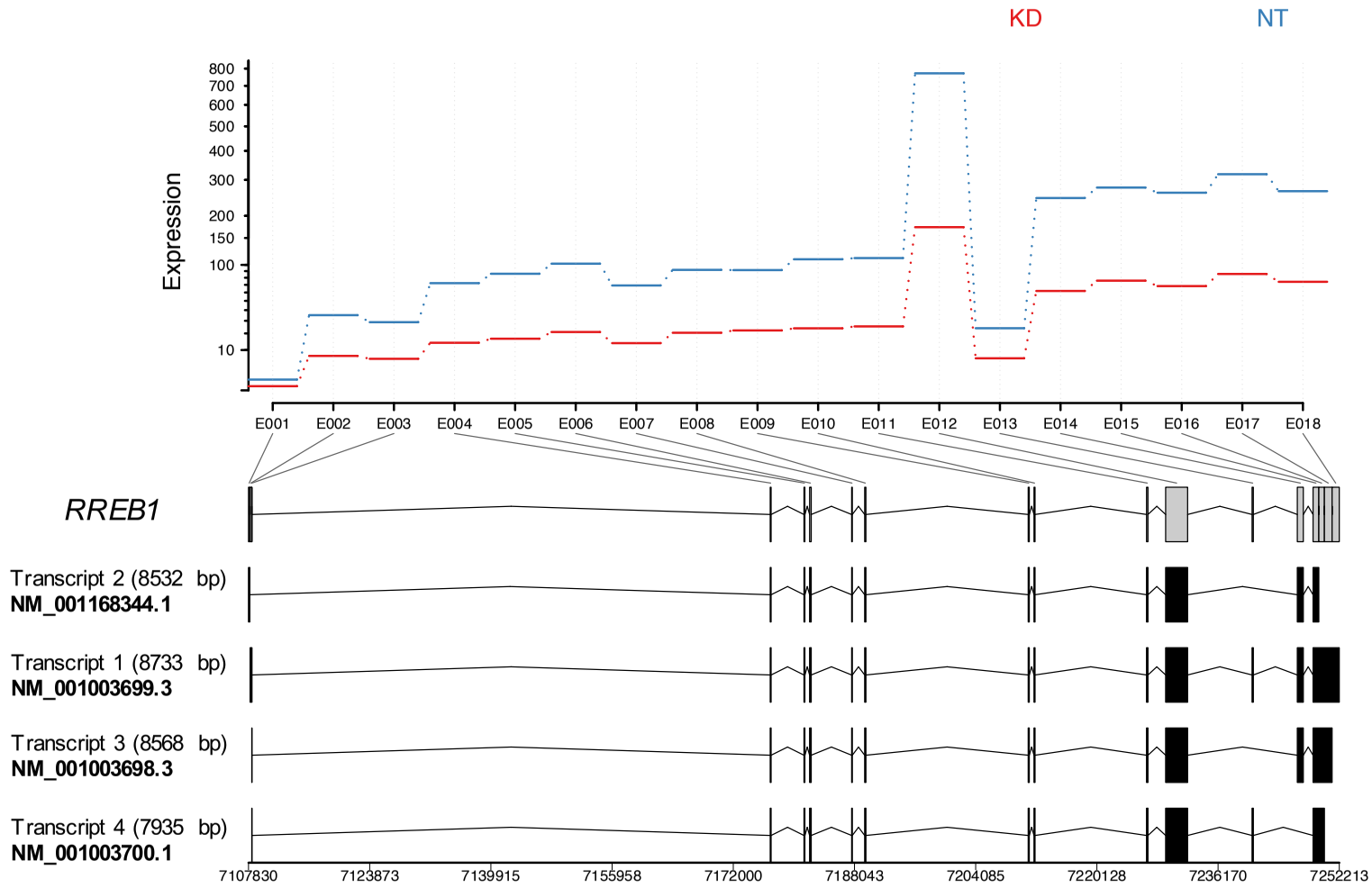


Figure 4.4: *RREB1* gene expression after analysis of differential exon usage using the DEXSeq package The fitted expression estimates for each exon of the *RREB1* gene are plotted for *RREB1* KD (red) and NT (blue) EndoC-βH1 samples. The annotated transcript variants for *RREB1* are specified. n=5; E001-E018 represent exon counting bins of the *RREB1* gene.

showed lower levels of *RREB1* across all exons (Figure 4.4).

4.3.2.2 Differential gene expression analysis

Next, transcriptome data of the remaining five KD/NT pairs were subjected to differential gene expression analysis using the Bioconductor DESeq2 package to identify genes that were differentially expressed as a result of reduced *RREB1* levels in mature human beta cells (see Section 2.5.2.2).

In total, 2,144 DEGs were detected between *RREB1* KD and NT samples, of which the majority (56%) were up-regulated in the KD cells. 55% and 56% of the up-regulated and down-regulated genes, respectively, corresponded to predicted RREB1 target genes identified in the JASPAR and TRANSFAC databases (Mathelier et al. 2014; Matys et al. 2006) (Table 4.1).

Table 4.1: Differentially expressed genes between *RREB1* KD and NT EndoC-βH1 cells

	DEGs	**Up-regulated**	**Down-regulated**
All	2,144	1,204 (56%)	940 (44%)
Predicted RREB1 target genes	1,192 (56%)	661 (55%)	531 (56%)

57,820 genes tested using DESeq2; q<0.01; DEGs, differentially expressed genes.

Enriched biological terms and pathways among all DEGs and those predicted to be regulated by RREB1 are listed in Table 4.2. Up-regulated DEGs were again enriched for biological processes associated with neurons such as 'nervous system development', 'neuronal system', 'synaptic signalling' and 'axon guidance', as seen for DEGs up-regulated in the *RREB1* KO developmental model (see Table 3.4) and likely attributable to the phenotypic similarities between neurons and beta cells (Atouf et al. 1997).

Interestingly, while terms relating to exocytotic processes including 'regulation of exocytosis', 'synaptic vesicle exocytosis', 'transmission across chemical synapses' and

Table 4.2: Subset of enriched biological terms and pathways among DEGs between *RREB1* KD and NT EndoC-βH1 cells

DEGs		# of enriched terms	Examples of enriched terms	padj (q)
Up-regulated	GO	245	nervous system development	7.68×10^{-40}
			synaptic signalling	2.41×10^{-15}
			regulation of exocytosis	4.12×10^{-4}
			synaptic vesicle exocytosis	0.0018
	REACTOME	22	neuronal system	3.37×10^{-10}
			transmission across chemical synapses	6.68×10^{-8}
			axon guidance	7.97×10^{-8}
	KEGG	3	cell adhesion molecules (CAMs)	7.34×10^{-5}
			Hippo signalling pathway	0.0091
Up-regulated RREB1 targets	GO	169	nervous system development	2.77×10^{-21}
			synaptic signalling	6.41×10^{-7}
			synaptic vesicle cycle	2.70×10^{-5}
			synaptic vesicle exocytosis	3.84×10^{-4}
			cell-cell signalling	4.73×10^{-4}
			neurotransmitter secretion	9.42×10^{-4}
	REACTOME	12	axon guidance	2.89×10^{-4}
			oncogenic MAPK signalling	0.0029
	KEGG	3	cell adhesion molecules (CAMs)	0.0028
Down-regulated	GO	76	response to endoplasmic reticulum stress	3.51×10^{-10}
			ER-nucleus signalling pathway	3.06×10^{-6}
			ERAD pathway	2.44×10^{-5}
			endoplasmic reticulum unfolded protein response	0.0013
			ATF6-mediated unfolded protein response	0.0041
	REACTOME	7	regulation of insulin-like growth factor (IFG) transport and uptake by insulin-like growth factor binding proteins (IGFBPs)	5.42×10^{-5}
			post-translational protein phosphorylation	7.09×10^{-5}
	KEGG	4	protein processing in endoplasmic reticulum	9.33×10^{-9}
			lysosome	1.79×10^{-4}
Down-regulated RREB1 targets	GO	24	response to endoplasmic reticulum stress	7.39×10^{-7}
			ER-nucleus signalling pathway	0.0076
			endoplasmic reticulum unfolded protein response	0.0084
	KEGG	3	protein processing in endoplasmic reticulum	8.18×10^{-5}

1,204 up-regulated and 940 down-regulated DEGs were tested for enriched biological terms and pathways using g:profiler (see Section 2.5.3). All human genes annotated in the Ensembl database were used as background. Significance threshold was set to q<0.01 using the tailor-made g:SCS algorithm for multiple testing. DEGs, differentially expressed genes.

'neurotransmitter secretion' were enriched among DEGs up-regulated in RREB1-deficient beta cells, processes and pathways associated with insulin or hormone secretion were not among enriched terms. This was not consistent with observations in hiPSC-derived BLCs and surprising, having observed a reduction in insulin content in *RREB1* KD EndoC-βH1 cells.

Down-regulated DEGs, including only those predicted to be regulated by RREB1, were enriched for terms regarding the ER, including 'response to endoplasmic reticulum stress', 'endoplasmic reticulum unfolded protein response', 'ATF6-mediated unfolded protein response', 'protein processing in endoplasmic reticulum' and 'lysosome' (Table 4.2).

4.3.2.3 Prediction of upstream transcriptional regulators of DEGs

TF enrichment analysis using the iRegulon Cytoscape plugin (Janky et al. 2014), identified RFX proteins, RREB1 and REST as the top three transcriptional regulators of DEGs up-regulated in *RREB1* KD samples (Table 4.3), consistent with findings for hiPSC-derived *RREB1* KO BLCs (see Table 3.5). Equally similar to BLCs, down-regulated DEGs were enriched for various TFs, including the hepatocyte nuclear factors HNF1A, HNF1B and FOX1A (NES=3.2-4.9).

Table 4.3: Predicted upstream regulators for DEGs up-regulated in the *RREB1* KD EndoC-βH1 model

TF	NES	# targets
RFX	3.1-7.6	524
RREB1	3.8-5.0	563
REST	3.0-4.5	629

Top three TF motifs of 9,713 position weight matrices (PWMs) and 1,120 ENCODE ChIP-Seq tracks (centred 10kb around TSS) tested. NES, normalised enrichment score with cutoff set to $\geq$3 (corresponding to a false discovery rate (FDR) of 3-9%; # targets, number of targets for TF motif with highest NES; RFX comprises RFX1-6, 8, RFXANK and RFXAP; DEGs, differentially expressed genes; TF, transcription factor.

4.3.2.4 Transcriptional activity of RFX proteins was modulated as a consequence of reduced *RREB1* levels in mature beta cells

Motif activity response analysis (MARA) strengthened the RFX family of TFs as transcriptional regulators affected by partial loss of RREB1 in beta cells. RFX2 and RFX3 were identified as the top key TFs mediating gene expression changes between *RREB1* KD and NT EndoC-βH1 samples. RFX2/3 motif activity (Z=9.55) was significantly higher in *RREB1* KD samples (q=4.39x10^{-5}) (Figure 4.5 A). While *RFX3* transcript levels were not affected by RREB1 deficiency (Figure 4.5 C), *RFX2* expression was markedly higher in *RREB1* KD compared to NT samples (q=1.50x10^{-9}, $\log_2$FC=0.4087) (Figure 4.5 B), suggesting a transcriptional activator role for RFX2 in line with findings in the hiPSC *RREB1* KO beta cell differentiation model (see Figure 3.15). Expression of *GPR56*, a RFX3 target gene (Bae et al. 2014), was significantly higher in *RREB1* KD than NT EndoC-βH1 cells (q=1.83x10^{-189}, $\log_2$FC=1.6438), implicating that RFX3 also acted as a transcriptional activator in beta cells.

Interrogation of the transcriptome dataset for transcriptional activity of the islet-specific RFX TF RFX6, using three in-house compiled PWMs (see Section 3.3.4.9), revealed no relevant role for the binding motifs RFX6_1 and RFX6_2 (Z=0.01), but activity conveyed by the RFX6_3 motif activity (Z=0.52), although only being ranked 218 out of 503 motifs tested, was significantly lower in *RREB1* KD than NT samples (q=7.94x10^{-5}) (Figure 4.5 D). *RFX6* expression was not affected by reduced levels of *RREB1* (Figure 4.5 E). Intriguingly, the PWMs for RFX2, RFX3 and RFX6 (RFX6_3) were very similar to each other (Figure 4.5 F), thus opposing motif activities for RFX2/3 and RFX6 were not anticipated, although a similar observation had already been made in the hiPSC beta cell developmental model (see Figures 3.15 A and 3.16 B).

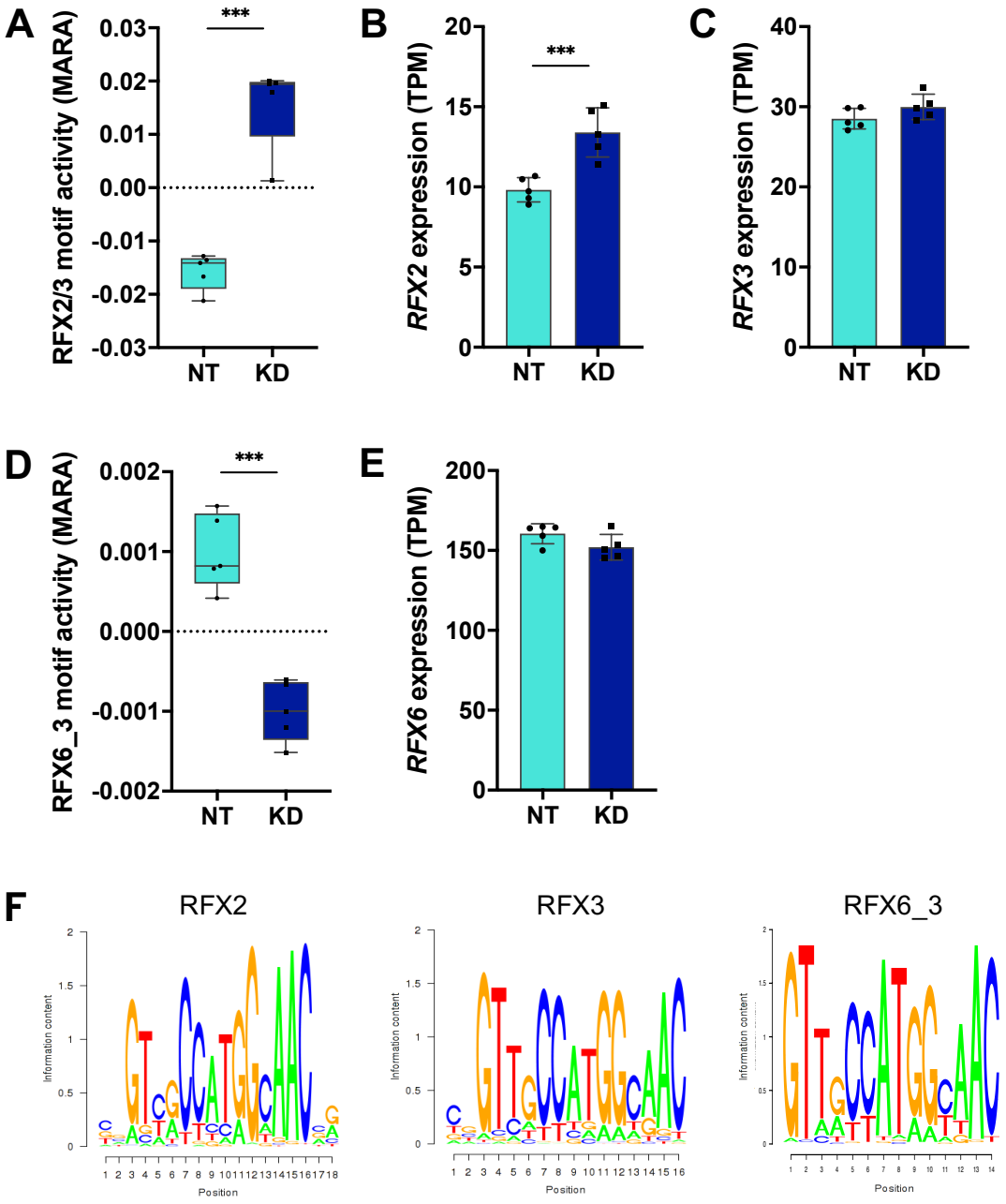


Figure 4.5: RFX2/3 and RFX6 motif activities for *RREB1* KD and NT EndoC-βH1 cells identified by MARA (A) RFX2/3 motif activity was significantly increased in *RREB1* KD EndoC-βH1 cells. (B) Expression of *RFX2* was markedly higher in KD than NT EndoC-βH1 cells. (C) *RFX3* expression was not affected by partial loss of RREB1. (D) RFX6_3 motif activity was significantly lower in *RREB1* KD EndoC-βH1 cells. (E) *RFX6* transcript levels were comparable between *RREB1* KD and NT EndoC-βH1 samples. (F) PWMs of RFX2, RFX3 and RFX6_3. n=5; values are displayed as mean±SD; *p*-values ***<0.001 for unpaired t-test (A and D) or Wald test followed by Benjamini-Hochberg adjustment (B, C and E); PWM, position weight matrix; TPM, transcripts per million; NT, turquoise; KD, darkblue.

4.3.2.5 REST transcriptional activity was up-regulated in *RREB1* KD EndoC-βH1 cells

In Chapter 3, I showed that the TF REST was implicated as a differentially active regulator during *in vitro* beta cell development (see Section 3.3.4.10). In mature beta cells iRegulon also suggested REST as an upstream regulator of DEGs up-regulated as a consequence of RREB1 deficiency (Table 4.3). MARA, which models genome-wide gene expression, ranked the REST motif 99th out of 503 motifs tested (Z=0.97). Transcriptional activity was significantly higher in *RREB1* KD than NT EndoC-βH1 cells (p=0.0394), replicating findings in hiPSC-derived *RREB1* KO BLCs (Figure 4.6 A). *REST* expression was modest, as expected for a beta cell disallowed gene, and tended to be slightly, albeit not significantly, lower in *RREB1* KD EndoC-βH1 cells (q=0.0828, $\log_2$FC=-0.2498) (Figure 4.6 B). Reduced motif activity implied less repression of neuronal (or beta cell) specific genes. Indeed, REST target genes implicated in insulin exocytosis *SNAP25* (q=1.73x10^{-6}, $\log_2$FC=0.1956) (Figure 4.6 C), *SCG2* (q=9.15x10^{-16}, $\log_2$FC=0.3539) (Figure 4.6 D), *NRXN1* (q=2.19x10^{-5}, $\log_2$FC=0.2220) (Figure 4.6 E), *NRXN2* (q=2.70x10^{-36}, $\log_2$FC=0.6584) (Figure 4.6 F), beta cell connectivity *GJD2* (q=2.18x10^{-11}, $\log_2$FC=0.5802) (Figure 4.6 G) and beta cell survival *CDK5R2* (q=0.0011, $\log_2$FC=0.1983) (Figure 4.6 H) were significantly higher expressed in RREB1-depleted EndoC-βH1 cells.

Taken together, my data show that partial loss of RREB1 led to a complex phenotype in EndoC-βH1 cells. While *INS* transcript levels and insulin content were reduced, beta cell signature genes were up-regulated. GSIS was unchanged in RREB1-deficient beta cells, raising the question that up-regulation of genes, implicated in insulin exocytosis, may have compensated for less available insulin. To gain further insight into the role of RREB1 in mature beta cells, I next generated a stable *RREB1* KO EndoC-βH1 line characterised by complete loss of RREB1.

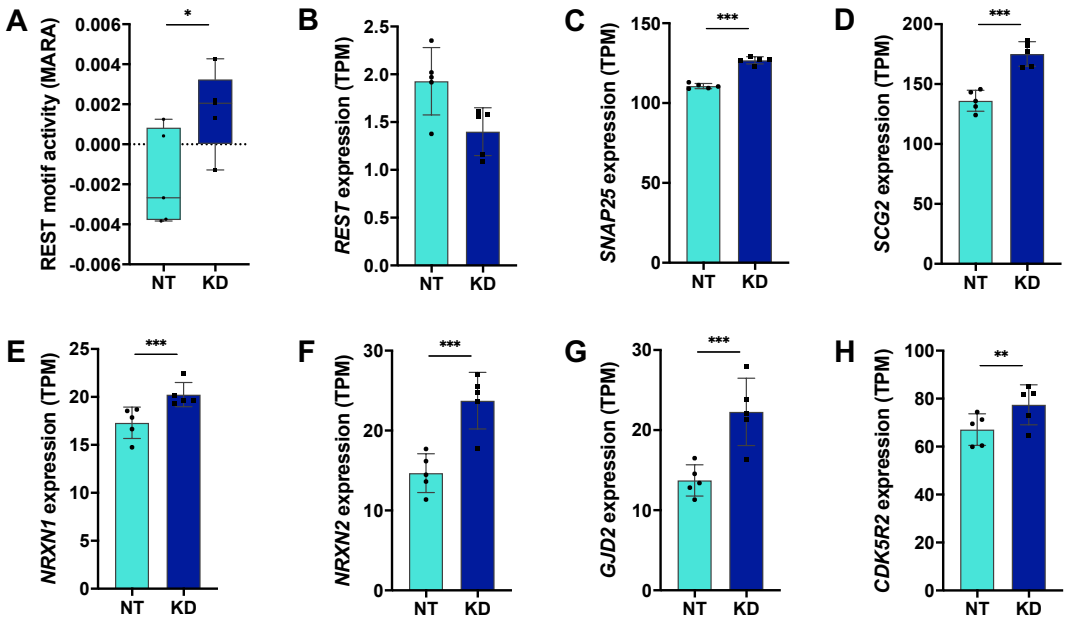


Figure 4.6: REST motif activity and target gene expression in *RREB1* KD and NT EndoC-βH1 cells (A) Transcriptional activity of REST was markedly higher in *RREB1* KD EndoC-βH1 cells. (B) Expression of *REST* in *RREB1* KD and NT EndoC-βH1 cells. Expression of REST target genes involved in insulin secretion *SNAP25* (C), *SCG2* (D), *NRXN1* (E) and *NRXN2* (F) as well as cell-to-cell communication *GJD2* (G) and pro-survival in beta cells *CDK5R2* (H) were significantly increased in RREB1-deficient EndoC-βH1 cells. n=5; values are displayed as mean±SD; *p*-values $^{*}<0.05$, $^{**}<0.01$, $^{***}<0.001$ for unpaired t-test (A) or Wald test followed by Benjamini-Hochberg adjustment (B-H); TPM, transcripts per million; NT, turquoise; KD, darkblue.

4.3.3 Generation and characterisation of a EndoC-βH1 *RREB1* KO model

To generate *RREB1* KO EndoC-βH1 cells a pooled lentiviral CRISPR/Cas9 approach was utilised including four sgRNAs targeting four different exons in *RREB1* (Figure 4.7 A). As EndoC-βH1 cells could not be grown as single clones, edited cells were expanded as a cell pool, thus the resulting *RREB1* KO line was a genetic mosaic, i.e. individual cells were characterised by differing *RREB1* genotypes, but with the high MOI leading to a predominance of LOF in the pool.

PCR amplification of gDNA from pooled *RREB1* KO cells revealed that a proportion of them carried a large deletion between *RREB1* exon 4 and 12, which was not detected in the EV control cells (Figure 4.7 B). WB analysis confirmed complete loss of RREB1 protein in the non-clonal *RREB1* KO EndoC-βH1 line

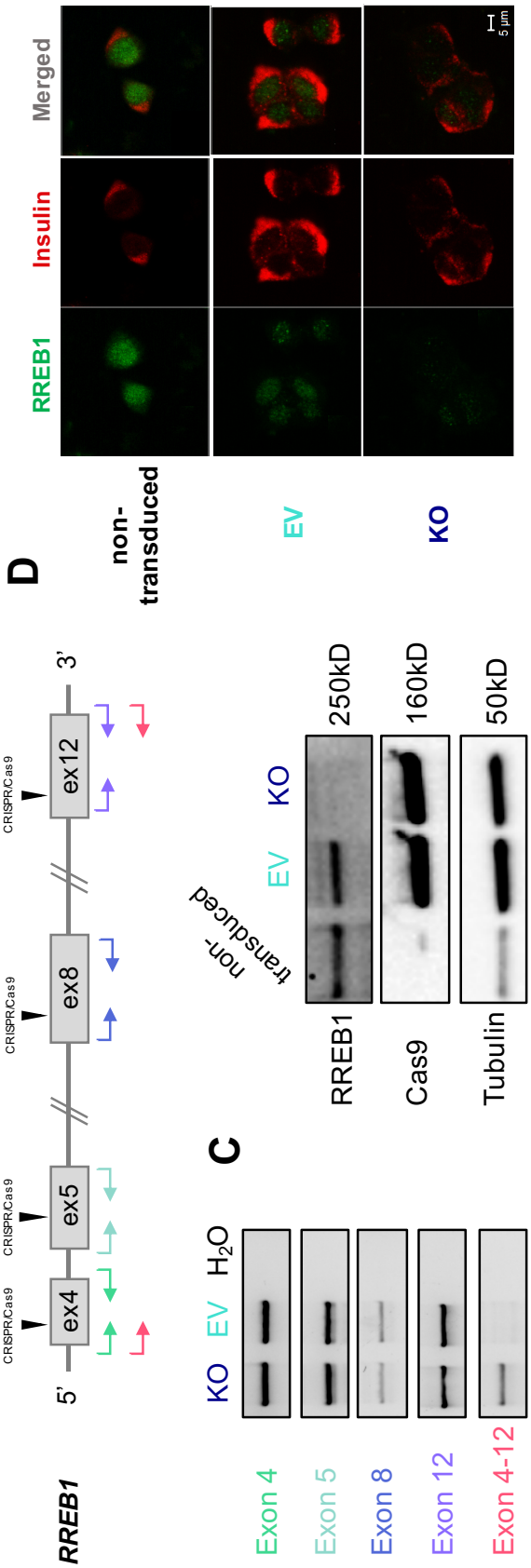


Figure 4.7: Quality control of *RREB1* KO EndoC-βH1 cells (A) Schema of *RREB1* highlighting the locations of the four sgRNAs (black arrowheads) used to generate a *RREB1* KO EndoC-βH1 line. Arrows below represent primers used for *RREB1* genotyping and indicate their location in the genome. Primer pairs are colour-coded. (B) Gel electrophoresis of PCR-amplified fragments. (C) Western blot analysis for validation of loss of RREB1 protein (~250kD) in *RREB1* KO EndoC-βH1 cells. *RREB1* KO and EV cells showed strong expression of the endonuclease Cas9. Tubulin (50kD) was used as protein loading control. (D) Immunofluorescence staining of RREB1 (green) and insulin (red). Only images of EV and KO cells were taken with the same microscopy settings. EV, control cells; non-transduced, unedited parental EndoC-βH1 line.

(Figure 4.7 C). Both *RREB1* KO and EV cells displayed strong expression of the endonuclease Cas9, suggesting successful transduction and selection of transduced cells. As expected, Cas9 expression was absent in the non-transduced parental EndoC-βH1 cell line (Figure 4.7 C). Immunofluorescence staining revealed that RREB1 was confined to the nucleus in non-transduced EndoC-βH1 as well as EV cells (Figure 4.7 D) similar to its subcellular localisation in hiPSCs (see Figure 3.2 F). Full nuclear staining was absent in *RREB1* KO EndoC-βH1 cells, confirming complete loss of RREB1. Insulin was restricted to the cytoplasm of beta cells, tending to aggregate on one site of the cell. Intriguingly, insulin staining seemed to be fainter for *RREB1* KO compared to EV EndoC-βH1 cells, supporting an effect of loss of RREB1 on insulin protein (Figure 4.7 D).

4.3.4 Cellular insulin content is reduced in *RREB1* KO EndoC-βH1 cells

Given that reduced levels of *RREB1* had caused an insulin content phenotype in EndoC-βH1 cells, I next determined the effect of complete loss of RREB1 on GSIS in *RREB1* KO and EV EndoC-βH1 cells. Insulin secretion was neither affected in basal (1mM) nor in high (20mM) glucose by *RREB1* KO in EndoC-βH1 cells (Figure 4.8 A). However, loss of RREB1 led to a significant reduction in cellular insulin content (-44.2%, p=0.0002) (Figure 4.8 B), replicating the insulin content phenotype of partial RREB1-deficient beta cells (see Figure 4.1 C) and confirming previously observed weaker insulin staining in *RREB1* KO EndoC-βH1 cells. To test whether reduction in insulin content was due to reduced cell number, cell viability was measured directly after glucose stimulation. Cell counts did not differ between *RREB1* KO and EV cell lines (p=0.8603) (Figure 4.8 C). Assessment of insulin mRNA levels revealed decreased *INS* gene expression (-32.9±13.6%, p=0.0357) (Figure 4.8 D), suggesting that reduced insulin content could be due to a downstream effect caused by a decrease in *INS* transcript levels. As a consequence of reduced insulin content, insulin released to a glucose stimulus was significantly

higher in *RREB1* KO beta cells when secretion measures were adjusted for insulin content (insulin secretion expressed as percentage of content) (x1.84±0.53, p=0.0012) (Figure 4.8 E), consistent with observations in partial RREB1-depleted beta cells (see Figure 4.1 E). Basal insulin secretion, normalised to cellular insulin content, also tended to be higher in RREB1-depleted beta cells, but did not reach statistical significance (x1.47±0.27, p=0.3000), thus stimulation index (ratio of insulin released to a glucose stimulus vs basal unstimulated secretion) was similar in *RREB1* KO and EV EndoC-βH1 cells (p=0.3355) (Figure 4.8 F). Normalisation by adjusting for cell count (insulin secretion expressed as percentage of cell numbers) did not reveal any differences in insulin secretion between *RREB1* KO and EV beta cells (Figure 4.8 G).

Next, I was interested in how decreased insulin availability affected beta cells under conditions of prolonged high insulin demand. To deplete *RREB1* KO and EV cells of docked insulin granules, they were stimulated with 20mM glucose in combination with 10μM forskolin (FRSK), a cAMP-elevating agent, for 30min prior to assessment of GSIS. The amount of insulin secreted per cell in response to FRSK stimulation was significantly lower in *RREB1* KO compared to EV beta cells (-40.3±18.0%, p=0.0470) (Figure 4.8 H). Assessment of insulin release per cell after FRSK-mediated docked granule depletion, revealed significantly decreased insulin secretion in basal glucose (-37.7±16.2%, p=0.0453) and in response to a glucose stimulus (-48.3±9.6%, p=0.0080) in *RREB1* KO EndoC-βH1 cells (Figure 4.8 I). While both EV and *RREB1* KO beta cells showed a blunted response to glucose after the FRSK challenge (ratio of high to basal secretion, 1.31±0.13 and 1.10±0.05, respectively), *RREB1* KO EndoC-βH1 cells tended to recover more poorly than EV beta cells (p=0.0576) (Figure 4.8 J). Insulin content per cell was also reduced in EndoC-βH1 cells lacking RREB1 (-52.3±23.6%, p=0.0124) (Figure 4.8 K), thus insulin secretion, when normalised by adjusting for insulin content per cell, did not show any difference in GSIS between *RREB1* KO and EV beta cells (Figure 4.8 L).

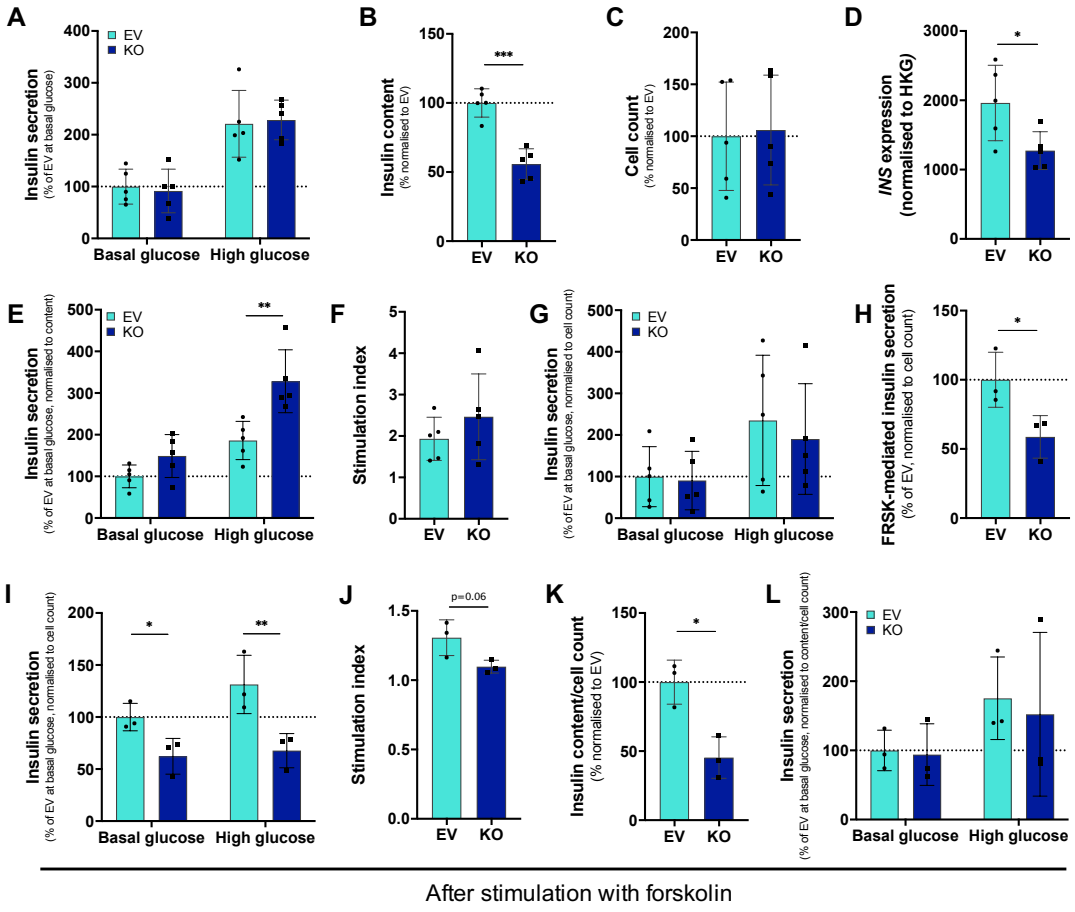


Figure 4.8: Effect of RREB1 loss on beta cell function *RREB1* KO EndoC-βH1 cells were assessed for glucose-stimulated insulin secretion (A), cellular insulin content (B), cell numbers (C) and insulin gene (*INS*) expression (D). Normalised insulin secretion by adjusting for insulin content revealed significantly higher insulin release in response to 20mM glucose (E) without an effect on stimulation index (ratio of insulin release in response to high vs basal glucose) (F). Insulin release adjusted for cell numbers showed no difference in GSIS between *RREB1* KO and EV EndoC-βH1 cells (G). To determine how *RREB1* KO EndoC-βH1 cells behave under increased insulin demand, beta cell function was characterised after stimulation with 20mM glucose and 10μM forskolin by forskolin-stimulated insulin secretion per cell (H), glucose-stimulated insulin secretion after stimulation with forskolin per cell (I), stimulation index (J), insulin content per cell (K) and glucose-stimulated insulin secretion normalised to cellular insulin content per cell (L). n=3-5; values are displayed as mean±SD; *p*-values $^{*}<0.05$, $^{**}<0.01$, $^{***}<0.001$ for two-way ANOVA followed by Sidak's multiple comparisons test (A, E, G, I and L) or unpaired t-test (B, C, D, F, H, J and K); HKG, housekeeping genes; FRSK, forskolin; EV, turquoise; KO, darkblue.

4.3.5 Loss of RREB1 did not lead to distinct morphological changes in mature beta cells

To determine whether loss of RREB1 had an effect on the ultrastructural features of beta cells, *RREB1* KO and EV EndoC-βH1 cells were assessed via electron microscopy (EM). EM analysis did not reveal any obvious structural differences in regard to the occurrence of cell organelles including mitochondria and lysosomes (Figure 4.9 A to C). Insulin granules were present in all cells and tended to accumulate on one site in some cells (Figure 4.9 A), consistent with the observation of insulin immunostaining in EndoC-βH1 cells (see Figure 4.7 D).

4.3.6 Transcriptome profiling of mature beta cells characterised by loss of RREB1

To further characterise how loss of RREB1 transcriptional function affected mature beta cells, EndoC-βH1 *RREB1* KO and EV cells were collected over a period of six subsequent cell passages and processed for RNA-Seq. Twelve RNA-Seq libraries were sequenced to a mean read depth of 20.5±2.6 million reads per sample.

4.3.6.1 *RREB1* transcript levels were up-regulated in *RREB1* KO EndoC-βH1 cells

As already observed in the developmental *RREB1* KO model, *RREB1* expression was markedly higher in *RREB1* KO compared to EV beta cells (q=2.84x10^{-21}, $\log_2$FC=0.7444). Analysis of exon usage revealed differential expression for exons 1 to 3 as well as 8 to 12 (Figure 4.10). While exons 8 to 12 were significantly less expressed in *RREB1* KO lines (q<0.02), exons 1 to 3 displayed markedly increased usage (q<0.001). The four sgRNAs used to generate the *RREB1* KO EndoC-βH1 line, targeted exons 4, 5, 8 and 12, thus reduced expression of exons 8 to 12 was indicative of a deletion event comprising this particular genomic region of *RREB1* in KO cells. Increased expression of exons 1 to 3, which are part of the

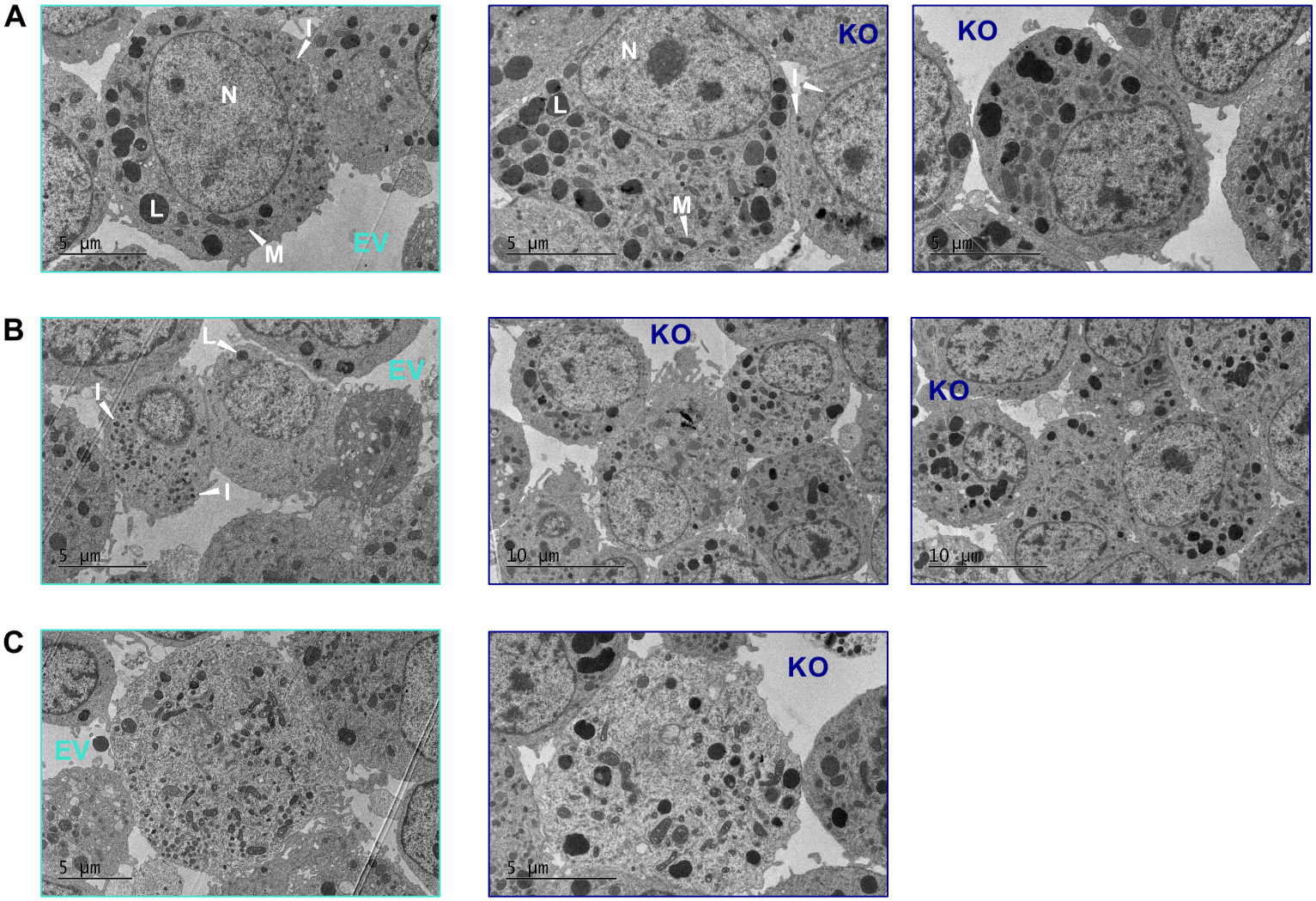


Figure 4.9: Analysis of beta cell ultrastructural features in *RREB1* KO EndoC-βH1 cells Ultrastructural features including nucleus (N), mitochondria (M), lysosomes (L) and insulin granules (I) in *RREB1* KO and EV EndoC-βH1 cells assessed via electron microscopy. (A) Single cell in focus. (B) Cell cluster. (C) Cells without visible nucleus. EV, turquoise; KO, darkblue.

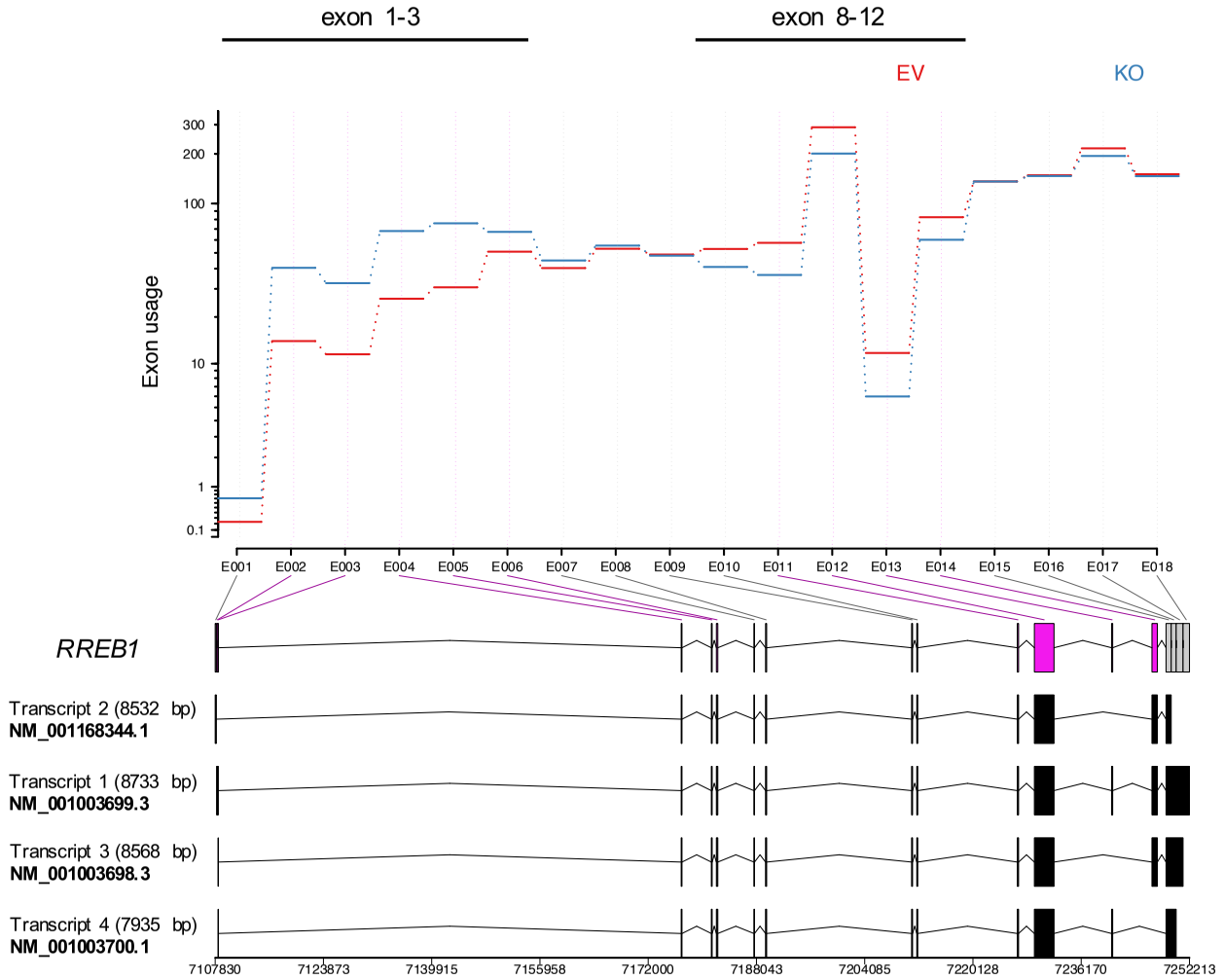


Figure 4.10: *RREB1* gene expression after analysis of differential exon usage using the DEXSeq package The fitted expression after subtraction of the overall changes in *RREB1* gene expression estimates for each exon of the *RREB1* gene are plotted for *RREB1* KO (blue) and EV (red) EndoC-βH1 samples. Exons that show significant differential exon usage are shown in pink. The annotated transcript variants for *RREB1* are specified. n=6; E001-E018 represent exon counting bins of the *RREB1* gene.

5'-UTR (untranslated region), was likely reflective of a compensatory up-regulation of *RREB1* gene expression caused by loss of RREB1 protein.

4.3.6.2 Principal component analysis identified a pair-wise pattern among *RREB1* KO and EV EndoC-βH1 samples

PCA of the 12 transcriptomes showed a clear separation between *RREB1* KO and EV samples, mediated by PC1. PC2 revealed additional pair-wise separation between samples originating from different cell passages, similar to earlier observations of *RREB1* KD and NT beta cell samples (Figure 4.11 A). Visualisation of transcriptome data in a heatmap uncovered greater differences between *RREB1* KO and EV samples than among individual KO or individual EV samples and mirrored pair-wise variability observed in PCA (Figure 4.11 B). This variability was taken into account, when transcriptome data of KO/EV pairs were subjected to differential gene expression analysis using the DESeq2 package (see Section 2.5.2.2).

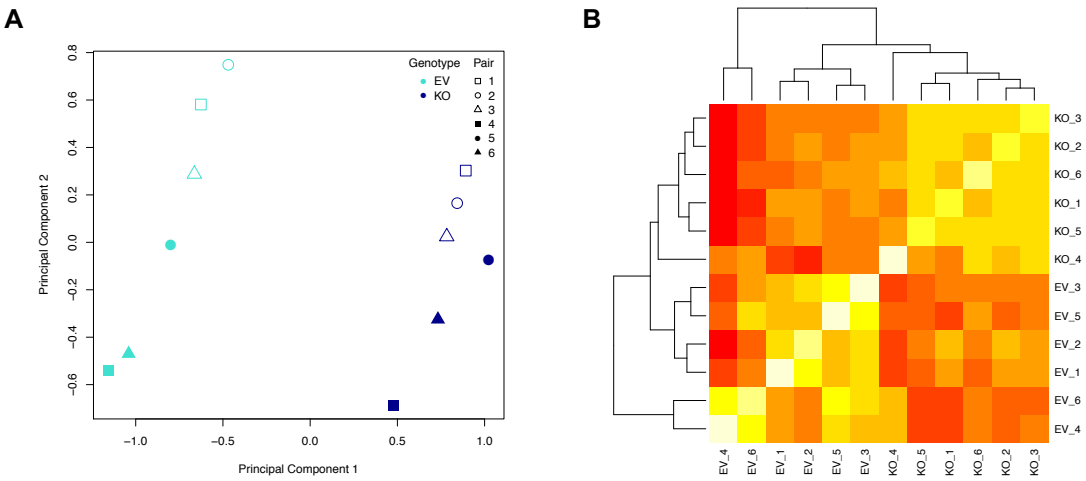


Figure 4.11: Principal component and correlation analyses of the transcriptomes of six *RREB1* KO and six EV beta cell samples (A) The first two principal components were calculated using normalised gene counts of all 12 *RREB1* KO and EV EndoC-βH1 samples. (B) Heatmap of normalised gene counts showing dissimilarity between *RREB1* KO and EV samples (increasing dissimilarity from lightyellow to darkred). n=6; EV, turquoise; KO, darkblue.

4.3.6.3 Differential gene expression analysis

A total of 2,604 DEGs were detected between *RREB1* KO and EV samples with the majority (66%) being up-regulated in the KO cells. Among those, 49% corresponded to predicted RREB1 target genes, while 53% of the down-regulated DEGs matched genes likely to be regulated by RREB1 (Table 4.4). Comparison with DEGs identified in either the *RREB1* KD beta cell model or the hiPSC-derived *RREB1* KO BLCs (Chapter 3) revealed a striking overlap in up-regulated DEGs. Partial and complete RREB1-deficient EndoC-βH1 cells shared 736 genes, while RREB1-depleted BLCs and *RREB1* KO beta cells had 512 up-regulated genes in common.

Table 4.4: Differentially expressed genes between *RREB1* KO and EV EndoC-βH1 cell lines

	DEGs	**Up-regulated**	**Down-regulated**
All	2,604	1,721 (66%)	883 (34%)
Predicted RREB1 target genes	1,047 (40%)	853 (49%)	466 (53%)
Overlap with *RREB1* KD model	930 (of 2,144)	736 (of 1,204)	194 (of 940)
Overlap with *RREB1* KO BLCs	639 (of 1,948)	512 (of 1,108)	127 (of 840)

57,820 genes tested using DESeq2; q<0.01; DEGs, differentially expressed genes.

Enriched biological terms and pathways among DEGs are listed in Table 4.5. Up-regulated DEGs, including the subset predicted to be regulated by RREB1, were again enriched for biological terms relating to neurons ('nervous system development', 'neuronal system', 'axon guidance'). Interestingly, while all up-regulated DEGs contained a significant number of genes involved in MAPK signalling, the subset of predicted RREB1 target genes did not, despite RREB1 having been shown to be a downstream effector of this pathway. In contrast to *RREB1* KD beta cells, several terms relating to exocytosis, insulin secretion and regulation of these processes were

Table 4.5: Subset of enriched biological terms and pathways among DEGs between *RREB1* KO and EV EndoC-βH1 cells

DEGs		# of enriched terms	Examples of enriched terms	padj (q)
Up-regulated	GO	324	nervous system development	$8.48x10^{-50}$
			synaptic vesicle exocytosis	$8.78x10^{-7}$
			insulin secretion	$2.72x10^{-5}$
			regulation of insulin secretion	$7.40x10^{-5}$
			hormone transport	$1.87x10^{-4}$
			regulation of calcium ion-dependent exocytosis	$3.30x10^{-4}$
			calcium-ion regulated exocytosis	$3.97x10^{-4}$
			regulation of synaptic vesicle exocytosis	0.0017
	REACTOME	33	neuronal system	$1.93x10^{-26}$
			transmission across chemical synapses	$2.60x10^{-13}$
			protein-protein interactions at synapses	$8.11x10^{-7}$
			regulation of insulin secretion	0.0017
			presynaptic depolarization and calcium channel opening	0.0058
	KEGG	17	axon guidance	$3.62x10^{-5}$
			glutamatergic synapse	$1.8x^{-4}$
			cAMP signalling	2.67^{-4}
			insulin secretion	0.0030
			MAPK signalling pathway	0.0043
Up-regulated RREB1 targets	GO	206	nervous system development	$6.13x10^{-24}$
			synaptic signalling	$5.78x10^{-15}$
			regulation of exocytosis	$6.12x10^{-4}$
			regulation of secretion	$6.48x10^{-4}$
	REACTOME	10	axon guidance	$1.42x10^{-6}$
			neuronal system	$2.15x10^{-6}$
			extracellular matrix organisation	$1.06x10^{-4}$
			transmission across chemical synapses	0.0080
	KEGG	4	axon guidance	$5.34x10^{-4}$
			cell adhesion molecules (CAMs)	0.0071
Down-regulated	GO	9	nuclear DNA replication	$9.46x10^{-5}$
			cellular response to stress	0.0014
	REACTOME	13	extension of telomeres	$1.96x10^{-4}$
			cell cycle	$7.83x10^{-4}$
	KEGG	2	DNA replication	$2.02x10^{-4}$
Down-regulated RREB1 targets	GO	1	cellular response to stress	$1.00x10^{-4}$
	REACTOME	2	DNA replication	0.0018
	KEGG	1	DNA replication	0.0027

1,721 up-regulated and 883 down-regulated DEGs were tested for enriched biological terms and pathways using g:profiler (see Section 2.5.3). All human genes annotated in the Ensembl database were used as background. Significance threshold was set to q<0.01 using the tailor-made g:SCS algorithm for multiple testing. DEGs, differentially expressed genes.

found to be enriched among DEGs up-regulated in *RREB1* KO beta cells, pointing to a potential role for RREB1 in regulation of beta cell function.

Down-regulated DEGs did not show an enrichment of any beta cell specific biological processes.

4.3.6.4 Prediction of upstream transcriptional regulators of DEGs

TF enrichment analysis using iRegulon predicted roles for the RFX TF family, REST and RREB1 as transcriptional regulators of DEGs up-regulated in the *RREB1* KO beta cell line (Table 4.6), as previously observed for *RREB1* KD beta cells and hiPSC-derived BLCs. Down-regulated DEGs were mainly enriched for E2F TFs, which are involved in cell cycle regulation and DNA synthesis (NES=3.0-4.7), consistent with the enrichment in biological terms 'DNA replication' and 'cell cycle' predicted by g:profiler (see Table 4.5). MARA reinforced the TFs RFX2, RFX3 and REST as differential transcriptional regulators in beta cells as a consequence of loss of RREB1, similar to the observations in partial RREB1-deficient beta cells and hiPSC-derived BLCs.

Table 4.6: Predicted upstream regulators for DEGs up-regulated in the *RREB1* KO EndoC-βH1 model

TF	NES	# targets
RFX	3.1-6.2	766
RREB1	3.6-4.7	927
REST	3.2-4.3	375

Top three TF motifs of 9,713 PWMs and 1,120 ENCODE ChIP-Seq tracks (centred 10kb around TSS) tested. NES, normalised enrichment score with cutoff set to $\geq$3 (corresponding to a false discovery rate (FDR) of 3-9%; # targets, number of targets for TF motif with highest NES; RFX comprises RFX1-6, 8, RFXANK and RFXAP; DEGs, differentially expressed genes.

4.3.6.5 Loss of RREB1 led to increased REST transcriptional activity in mature beta cells

As expected for a beta cell disallowed gene, *REST* expression was low in EV cells and tended to be even lower, albeit not to a significant extent, in *RREB1* KO

EndoC-βH1 cells (q=0.2831, $\log_2$FC=-0.3579) (Figure 4.12 A). Two independent TF enrichment prediction tools (ISMARA and iRegulon) highlighted the repressor REST as a differential regulator in *RREB1* KO and EV transcriptomic datasets. REST motif activity was significantly increased in RREB1-deficient beta cells (Z=3.72, p=1.58x10^{-6}), similar to findings in *RREB1* KD and *RREB1* KO hiPSC-derived beta cells (Figure 4.12 B). In line with this, REST target genes involved in insulin exocytosis including *SNAP25* (q=3.72x10^{-25}, $\log_2$FC=0.7427), *CHGB* (q=0.0013, $\log_2$FC=0.3275), *SCG2*, (q=1.91x10^{-21}, $\log_2$FC=0.8103), *GRIA2*, (q=3.22x10^{-4}, $\log_2$FC=0.8057), *NRXN1* (q=2.04x10^{-11}, $\log_2$FC=0.4951) and *NRXN2* (q=8.22x10^{-18}, $\log_2$FC=1.7516), genes important in cell-to-cell communication such as *GJD2* (q=1.18x10^{-15}, $\log_2$FC=1.2188) and genes associated with pro-survival activity like *CDK5R2* (q=1.56x10^{-4}, $\log_2$FC=0.3179) and *PTPRN* (q=2.03x10^{-11}, $\log_2$FC=0.4465) were significantly higher expressed in *RREB1* KO compared to EV EndoC-βH1 cells, as has already been observed in the developmental as well as the partial RREB1-depleted beta cell models.

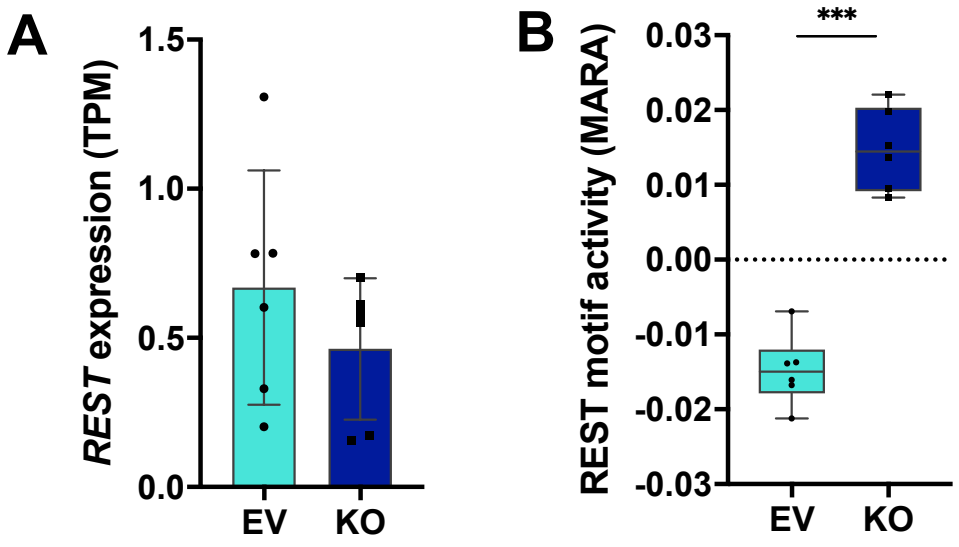


Figure 4.12: REST motif activity and expression profiles in *RREB1* KO and EV EndoC-βH1 cells (A) *REST* transcript levels in *RREB1* KO and EV EndoC-βH1 cells. (B) REST transcriptional activity was significantly increased as a consequence of loss of RREB1. n=6; values are displayed as mean±SD; p-values ***<0.001 for Wald test followed by Benjamini-Hochberg adjustment (A) or unpaired t-test (B); TPM, transcripts per million; EV, turquoise; KO, darkblue.

4.3.6.6 Transcriptional activity of RFX proteins was modulated as a consequence of loss of RREB1 in mature beta cells

RFX2 and RFX3 were again among the top TFs mediating gene expression variation between *RREB1* KO and EV EndoC-βH1 samples, displaying significantly higher motif activity (Z=8.45) in *RREB1* KO samples (q=1.73x10^{-6}) (Figure 4.13 A). *RFX2* transcript levels (q=6.47x10^{-5}, $\log_2$FC=0.3629) as well as RFX2 protein expression (x2.26±0.10, p=0.0123) were markedly increased in *RREB1* KO EndoC-βH1 cells (Figure 4.13 B and C). While there was no effect of RREB1 loss on gene expression, RFX3 protein expression was slightly, albeit not significantly increased in *RREB1* KO EndoC-βH1 cells (x1.52±0.38, p=0.1004) (Figure 4.13 D and E), consistent with the findings in partial RREB1-depleted beta cells (see Figure 4.5). Increased RFX2 motif activity, protein expression and target gene expression (*CAMK2A*, q=8.71x10^{-109}, $\log_2$FC=3.9429) highlighted RFX2 as transcriptional activator. Slightly increased RFX3 protein levels as well as significantly higher target gene expression (*GPR56*, q=5.24x10^{-283}, $\log_2$FC=3.1262 and *GCK* (Ait-Lounis et al. 2010), q=0.0235, $\log_2$FC=0.2160) strongly supported a transcriptional activator role for RFX3 in mature beta cells, similar to RFX2. RFX6_3 motif activity (Z=0.57, q=0.0382) and *RFX6* gene expression (q=3.06x10^{-8}, $\log_2$FC=-0.4188) were significantly lower in *RREB1* KO than EV samples (Figure 4.13 F and G), while RFX6 protein levels were not affected by loss of RREB1 (Figure 4.13 H).

4.3.6.7 Reduced RFX6 motif activity and gene expression did not underlay the transcriptional phenotype of *RREB1* KO beta cells

To investigate, whether reduced RFX6 transcriptional activity might partially underlay the transcriptional phenotype of *RREB1* KO beta cells, I next compared the transcriptomic signature of *RREB1* KO EndoC-βH1 cells with gene expression patterns in human and mouse beta cells characterised by reduced expression or complete loss of RFX6 (Chandra et al. 2014; Piccand et al. 2014). Table 4.7 gives

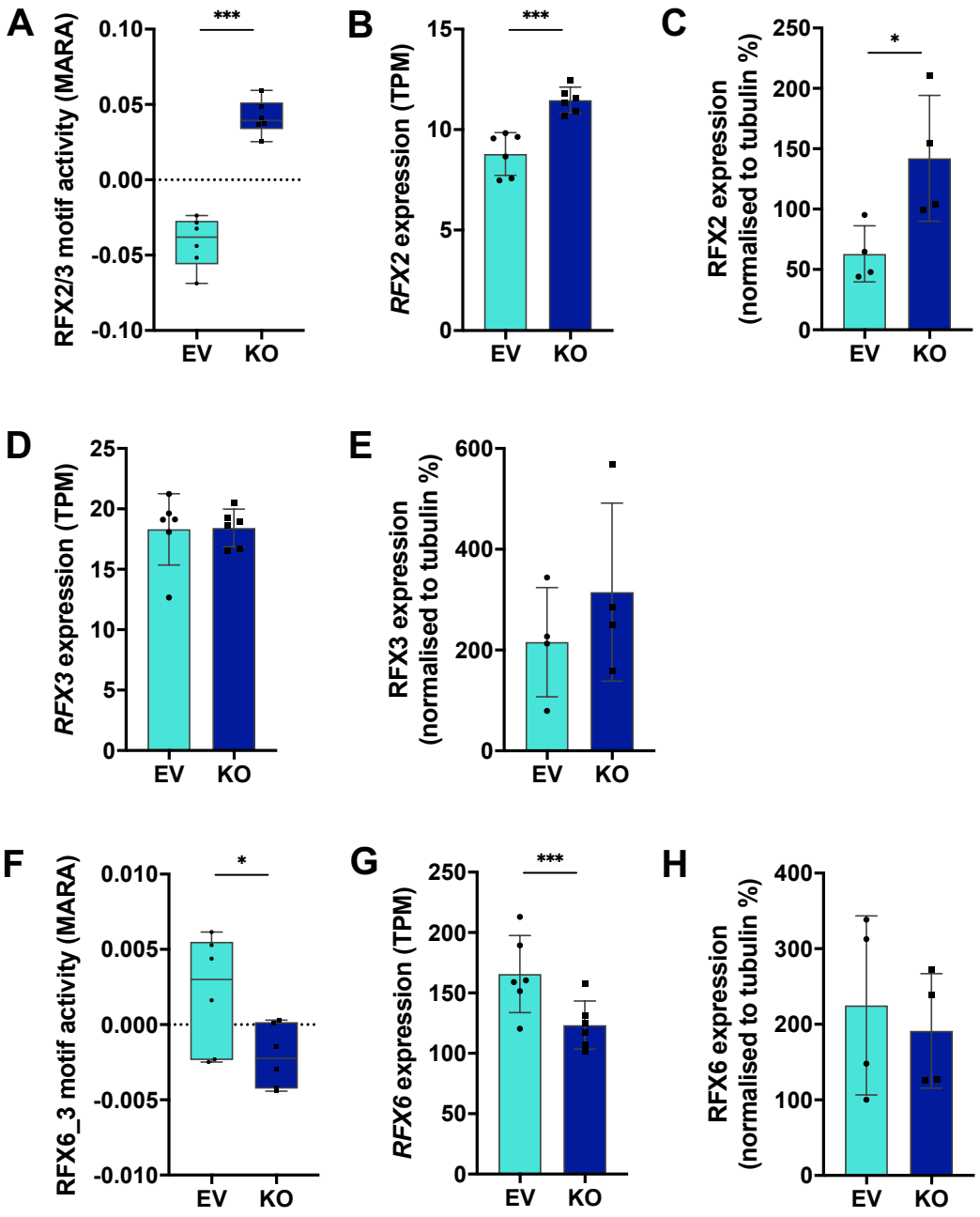


Figure 4.13: Loss of RREB1 in mature beta cells affected transcriptional activity and expression of RFX2 ad RFX3 (A) RFX2/3 motif activity was significantly higher as a consequence of loss of RREB1 in EndoC-βH1 cells. Expression of *RFX2* gene (B) and RFX2 protein (C) were markedly increased in *RREB1* KO EndoC-βH1 cells. (D) *RFX3* expression was not affected by loss of RREB1. (E) RFX3 protein expression tended to be slightly higher in *RREB1* KO cells. RFX6_3 motif activity (F) and *RFX6* transcript levels (G) were significantly lower in *RREB1* KO compared to EV EndoC-βH1 cells. (H) RFX6 protein expression was not affected by loss of RREB1. n=4-6; values are displayed as mean±SD; *p*-values $^{*}<0.05$, $^{***}<0.001$ for unpaired t-test (A, C, E, F and H) or Wald test followed by Benjamini-Hochberg adjustment (B, D and G); TPM, transcripts per million; EV, turquoise; KO, darkblue.

an overview of interrogated genes and the directional effect on mRNA expression as a consequence of either RREB1 or RFX6 depletion.

While *RFX6* KD in human beta cells had no effect on genes encoding beta cell TFs (Chandra et al. 2014), *Neurod1*, *Pax6* and *Isl1* expression was negatively affected by loss of RFX6 in rodent beta cells (Piccand et al. 2014). RREB1 deficiency in EndoC-βH1 revealed a similar effect for *NEUROD1* expression (p=6.96x10^{-8}, $\log_2$FC=-0.3433), and additional down-regulation of *MAFB* (p=0.0226, $\log_2$FC= -0.2212) and *NKX2.2* (p=0.0290, $\log_2$FC=-0.1667), but a reciprocal increase in *ISL1* expression (p=0.0047, $\log_2$FC=0.2407).

ABCC8 and *GCK*, direct RFX6 target genes in mouse and significantly down-regulated by loss of RFX6 in both human and murine beta cells (Chandra et al. 2014; Piccand et al. 2014), were either unaffected by loss of RREB1 (*ABCC8*) or revealed slightly increased expression levels (*GCK*, p=0.0235, $\log_2$FC=0.2160) in *RREB1* KO EndoC-βH1 cells. On the other hand, while expression of *SLC2A2*, encoding the glucose-facilitated transporter GLUT2 was unaffected by RFX6 depletion (Piccand et al. 2014), expression of both *SLC2A1* (p=1.07x10^{-6}, $\log_2$FC=-0.3693) and *SLC2A2* (p=0.0033, $\log_2$FC=-0.6623) was significantly lower in *RREB1* KO beta cells. Genes encoding voltage-gated Ca^{2+} channel subunits, significantly down-regulated as a consequence of RFX6 loss (Chandra et al. 2014; Piccand et al. 2014), were consistently up-regulated in *RREB1* KO EndoC-βH1 cells.

Furthermore, while RFX6 depletion had caused increased expression of beta cell disallowed genes (Piccand et al. 2014), a characteristic of immature beta cells, *RREB1* KO EndoC-βH1 cells revealed an enrichment of beta cell forbidden genes in down-regulated DEGs (p=2.15x10^{-6}). Those included pro-proliferative genes such as *PDGFRA*, encoding the platelet-derived growth factor receptor α, required for the age-dependent decreasing proliferative capacity in islet cells (H Chen et al. 2011) and *IGBP4* which has been shown to bind to insulin-like growth factor 1 (IGF1) thereby inhibiting its pro-proliferative signalling (Durai et al. 2006).

Table 4.7: Comparison of gene expression profiles between *RREB1* KO EndoC-βH1 cells and RFX6-deficient beta cell models

Gene	***RREB1* KO EndoC-βH1 model**	**RFX6 deficiency in mature human beta cells[1]**	**RFX6 deficiency in adult mouse beta cells[2]**
RFX6	↓ (p=3.06x10^{-8}, $\log_2$FC=-0.4188)	↓	↓
PDX1	↔	↔	↔
MAFA	↔	NA	↔
MAFB	↓ (p=0.0226, $\log_2$FC=-0.2212)	↔	↔
NEUROD1	↓ (p=6.96x10^{-8}, $\log_2$FC=-0.3433)	↔	↓
PAX6	↔	↔	↓
NKX2.2	↓ (p=0.0290, $\log_2$FC=-0.1667)	↔	↔
NKX6.1	↔	NA	↔
ISL1	↑ (p=0.0047, $\log_2$FC=0.2407)	NA	↓
INS*	↓ (p=0.0107, $\log_2$FC=-0.4159)	↓	*Ins1* ↓, *Ins2* ↔
PCSK1	↑ (p=0.0359, $\log_2$FC=0.2058)	NA	↔
*GCK***	↑ (p=0.0235, $\log_2$FC=0.2160)	↓	↓
SLC2A1	↓ (p=1.07x10^{-6}, $\log_2$FC=-0.3693)	NA	NA
SLC2A2	↓ (p=0.0033, $\log_2$FC=-0.6623)	NA	↔
*ABCC8***	↔	↓	↓
KCNJ11	↔	↓ (n.s.)	↔
UCN3	↑ (p=5.19x10^{-6}, $\log_2$FC=0.5440)	NA	↓
CACNA1A	↔ ↑ in RT-qPCR (p=0.0345, x1.24±0.21)	↓	↓
*CACNA1C***	↑ (p=0.0013, $\log_2$FC=0.2349)	↓	↓
CACNA1B	↑ (p=2.58x10^{-11}, $\log_2$FC=0.5776)	NA	↓
CACNA1D	↑ (p=3.81x10^{-12}, $\log_2$FC=0.5913)	↓	↓
CACNA1E	↑ (p=5.90x10^{-7}, $\log_2$FC=1.1858)	NA	↓
CACNB1	↑ (p=5.46x10^{-10}, $\log_2$FC=0.7944)	NA	↓
*CACNB2***	↑ (p=0.0153, $\log_2$FC=0.3123)	↓	↓
PDGFRA	↓ (p=0.0038, $\log_2$FC=-0.7951)	NA	↑
IGFBP4	↓ (p=0.0031, $\log_2$FC=-0.7011)	NA	↑
CD302	↓ (p=6.95x10^{-5}, $\log_2$FC=-0.2866)	NA	↑
SELENBP1	↓ (p=1.77x10^{-10}, $\log_2$FC=-1.4107)	NA	↑
ZFP36L1	↓ (p=5.28x10^{-4}, $\log_2$FC=-0.5450)	NA	↑
FCGRT	↓ (p=3.90x10^{-13}, $\log_2$FC=-0.6666)	NA	↑
GALM	↓ (p=0.0016, $\log_2$FC=-0.3860)	NA	↑

[1]Chandra et al. 2014; [2]Piccand et al. 2014; ↓, down-regulated; ↔, unchanged; ↑, up-regulated; NA, data not available; n.s., not significant; **overlapping phenotype**; *, direct RFX6 target gene in human; **, direct RFX6 target gene in mouse; from top: TFs involved in endocrine differentiation, genes involved in insulin secretion, genes encoding Ca^{2+} channel subunits, beta cell disallowed genes.

Consistent with findings so far, expression of the beta cell maturation marker *UCN3* was significantly up-regulated in *RREB1* KO beta cells (q=5.19x10^{-6}, $\log_2$FC=0.5440), hence, in contrast to the reduced expression observed in *Rfx6* KO rodent beta cells (Piccand et al. 2014).

Taken together, apart from decreased *RFX6*, *INS* and *NEUROD1* transcript levels, beta cell signature genes revealed a directionally opposite gene expression pattern in *RREB1* KO EndoC-βH1 cells compared to *RFX6* deficiency in adult human and rodent beta cells (Chandra et al. 2014; Piccand et al. 2014). Therefore, reduced RFX6 transcriptional activity was unlikely to be responsible for the transcriptional phenotype associated with loss of RREB1 in mature beta cells.

4.3.6.8 Transcriptional activity of *INS* gene regulators PDX1 and NEUROD1 were reduced in *RREB1* KO beta cells

NEUROD1 and PDX1, two well known regulators of the *INS* gene promoter (Glick et al. 2000), showed differential transcriptional activity between *RREB1* KO and EV EndoC-βH1 cells. NEUROD1 motif activity (Z=1.50, p=0.0320) as well as gene expression (q=6.96x10^{-8}, $\log_2$FC=-0.3433) were significantly reduced in *RREB1* KO beta cells (Figure 4.14 A and B). For PDX1, transcriptional activity was markedly lower in RREB1-deficient beta cells (Z=1.51, p=8.96x10^{-4}), however, this was not mirrored at the transcript level (Figure 4.14 C and D). Therefore, both NEUROD1 and PDX1 represent TFs potentially accountable for down-regulation of *INS* gene expression.

4.3.7 The RFX TF family and RREB1

In light of the impact of RREB1 deficiency on transcriptional activity of RFX2 and RFX3 in mature beta cells as well as hiPSC-derived BLCs, I sought to determine, whether effects on gene expression profiles associated with loss of RREB1, could be mirrored by modulating RFX proteins in beta cells. Therefore, I characterised gene expression profiles of *RREB1* KO and EV EndoC-βH1 cells after RNA interference

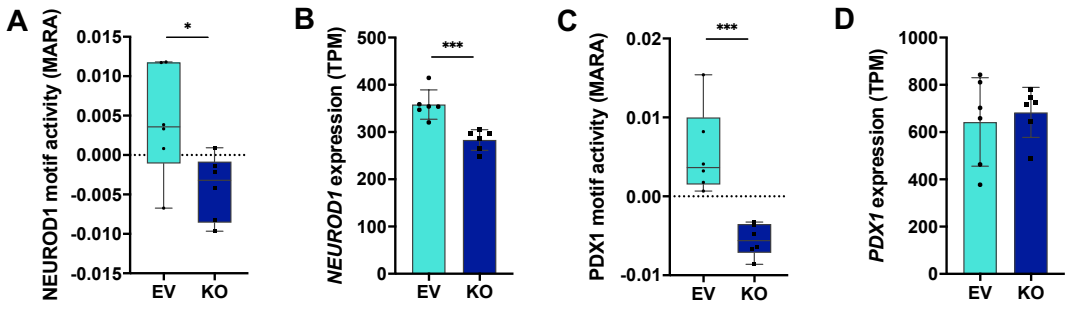


Figure 4.14: NEUROD1 and PDX1 motif activities were reduced in *RREB1* KO EndoC-βH1 cells NEUROD1 motif activity (A) and *NEUROD1* transcript levels (B) were significantly lower in *RREB1* KO compared to EV EndoC-βH1 cells. (C) PDX1 motif activity in *RREB1* KO and EV EndoC-βH1 cells calculated by MARA. (D) Loss of RREB1 did not affect *PDX1* gene expression. n=6; values are displayed as mean±SD; p-values $^{*}<0.05$, $^{***}<0.001$ for unpaired t-test (A and C) or Wald test followed by Benjamini-Hochberg adjustment (B and D); TPM, transcripts per million; EV, turquoise; KO, darkblue.

mediated inhibition of either *RFX2* or *RFX3* expression. Interrogated genes included known and computationally-predicted RFX2 and RFX3 target genes, significantly up-regulated in *RREB1* KO EndoC-βH1 cells such as *CAMK2A*, *CACNA1A*, *GCK*, *GPR56*, *DLG4* and *MAP6*.

4.3.7.1 Loss of RREB1-associated gene expression changes could not be attributed to RFX2

siRNA-mediated KD of *RFX2* caused a reduction of 44.7±8.8% in *RFX2* transcript levels (p=1.48x10^{-4}), corresponding to a 67.6±31.9% reduction in RFX2 protein expression (p=0.0571) in EV EndoC-βH1 cells. *RREB1* KO beta cells revealed a slightly stronger decrease in *RFX2* transcript and RFX2 protein levels of 53.0±13.3% (p=1.12x10^{-5}) and 85.8±8.9% (p=0.0286), respectively, resulting in similar residual RFX2 expression in *RREB1* KO and EV beta cells (Figure 4.15 A-C). Expression of selected target genes was unaffected by reduced RFX2 levels, except for *CACNA1A*, whose expression was significantly up-regulated in RFX2-depleted EV beta cells (x1.56±0.29, p=0.0159), while remaining unchanged in RFX2-deficient *RREB1* KO EndoC-βH1 cells (Figure 4.15 D).

RREB1 KO beta cells were characterised by increased *RFX2* and increased *CACNA1A* transcript levels. If RFX2 was solely responsible for up-regulated

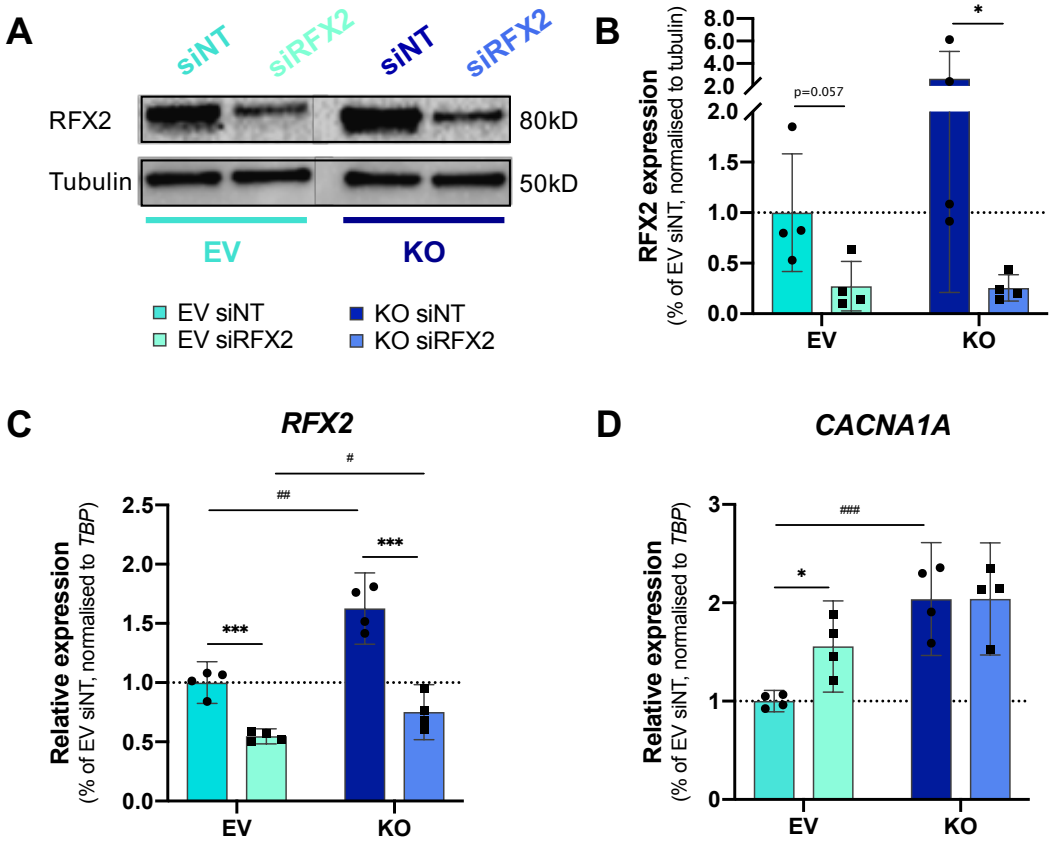


Figure 4.15: Analysis of gene expression profiles in *RREB1* KO and EV EndoC-βH1 cells following *RFX2* silencing (A) Example WB showing effect of *RFX2* silencing on RFX2 protein expression. (B) siRFX2 treatment resulted in similar RFX2 protein expression levels in *RREB1* KO and EV EndoC-βH1 cells. (C) *RFX2* transcript levels after RNA interference mediated inhibition of gene expression. (D) Reduced levels of RFX2 caused an increase in *CACNA1A* expression in EV, but not *RREB1* KO beta cells. n=4; values are displayed as mean±SD; p-values * or $^{\#}$<0.05, ** or $^{\#\#}$<0.01, *** or $^{\#\#\#}$<0.001 for two-way ANOVA followed by Tukey's multiple comparisons test (B, C and D); EV, turquoise; KO, darkblue.

CACNA1A gene expression, reduced levels of *RFX2* in EV beta cells were expected to result in decreased *CACNA1A* expression levels, however the opposite effect was observed.

4.3.7.2 *RFX3* silencing partially rescued up-regulated *CAMK2A* and *GPR56* gene expression in *RREB1* KO EndoC-βH1 cells

RNA interference significantly reduced *RFX3* transcript levels in both *RREB1* KO (79.9-±2.2%, p=6.66x10^{-9}) and EV (75.9-±4.9%, p=2.32x10^{-8}) EndoC-βH1

cells, resulting in an almost complete loss of RFX3 at the protein level (Figure 4.16 A-C). EV cells revealed a reduction of 93.9±10.4% (p=0.0286) and *RREB1* KO beta cells showed on average a 94.4±9.3% (p=0.0286) decrease in RFX3 expression (Figure 4.16 B). Reduced *RFX3* levels partially rescued increased *CAMK2A* and *GPR56* expression in *RREB1* KO beta cells. *CAMK2A* transcript levels dropped by 35.9±6.0% (p=0.0074), implicating RFX3 as potential regulator of *CAMK2A* transcription. A small decrease in *CAMK2A* gene expression on top of the already low expression levels was also observed in EV cells, but did not reach statistical significance (-18.0±11.5%, p=0.2910). As reduced levels of RFX2 did not impact *CAMK2A* gene expression, RFX3 was highly likely responsible for the up-regulation of *CAMK2A* in *RREB1* KO beta cells. Indeed, a search in human sequence-specific motif and regulatory site databases revealed TF binding sites for RFX3 in *CAMK2A* gene promoters (Jolma et al. 2010; Jolma et al. 2013; Pachkov et al. 2013). *GPR56*, an established RFX3 target gene, was down-regulated in both *RREB1* KO (43.7-±10.0%, p=0.0077) and EV (-51.2±4.8%, p=0.0021) EndoC-βH1 cells as a consequence of RFX3 depletion, strengthening RFX3 as transcriptional regulator affected by loss of RREB1 in mature beta cells.

Taken together, siRNA-mediated individual silencing of *RFX2* or *RFX3* in *RREB1* KO and EV EndoC-βH1 cells, highlighted RFX3 as the most likely transcriptional regulator affected by RREB1 deficiency. Solely reduced RFX3 expression resulted in a partial rescue of the transcriptional phenotype associated with loss of RREB1 in beta cells.

4.3.7.3 *RFX* depletion affected RREB1 protein expression in mature beta cells

Interrogation of sequence-specific binding of human TFs identified binding sites for RFX2 and RFX3 in *RREB1* gene promoters (Jolma et al. 2010; Jolma et al. 2013; Matys et al. 2006). Hence, I was interested to assess, whether RNA-mediated inhibition of *RFX2* or *RFX3* expression had an impact on RREB1 protein levels.

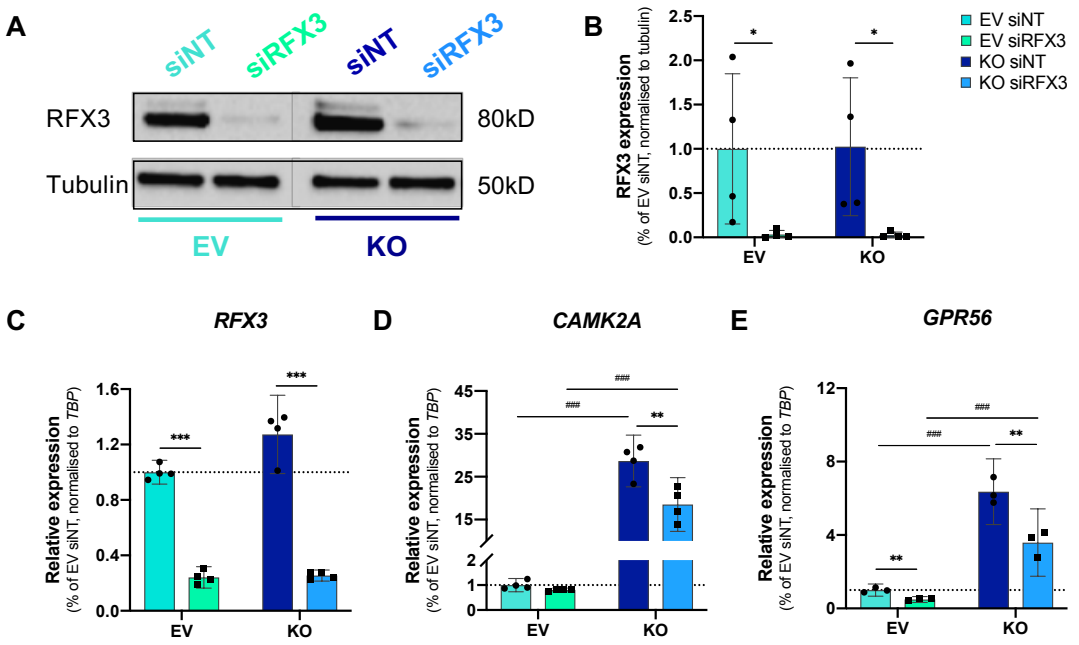


Figure 4.16: *RFX3* silencing partially rescued up-regulated *CAMK2A* and *GPR56* gene expression in *RREB1* KO EndoC-βH1 cells (A) Exemplary WB for RFX3 after *RFX3* silencing. (B) siRFX3 treatment resulted in almost complete loss of RFX3 protein expression in *RREB1* KO and EV EndoC-βH1 cells. (C) *RFX3* transcript levels were significantly reduced by RNA interference. (D) RFX3 depletion led to a down-regulation of *CAMK2A* expression in *RREB1* KO EndoC-βH1 cells. (E) *GPR56* transcript levels were significantly reduced in RFX3-deficient *RREB1* KO and EV beta cells. n=3-4; values are displayed as mean±SD; p-values $^{*}<0.05$, $^{**}<0.01$, *** or $^{\#\#\#}<0.001$ for two-way ANOVA followed by Tukey's multiple comparisons test (B, C, D and E); EV, turquoise; KO, darkblue.

WB analysis revealed a significant reduction of RREB1 expression as a consequence of RFX2 or RFX3 deficiency in EV EndoC-βH1 cells (-49.2±8.5%, p=0.0074 and -52.8±15.8%, p=0.0047, respectively), raising the question that RFX proteins may exert their effect through alterations in RREB1 activity.

4.3.8 Identification of *cis*-regulated RREB1 target genes

Transcriptome profiling identified a considerable set of genes affected by loss of RREB1, but did not discriminate whether DEGs were direct targets of RREB1 (*cis*-regulated genes) or rather controlled by other TFs which in turn were regulated by RREB1 (indirect target or *trans*-regulated genes). ChIP-Seq represents an approach that can aid identification of physical TF-DNA binding sites in the

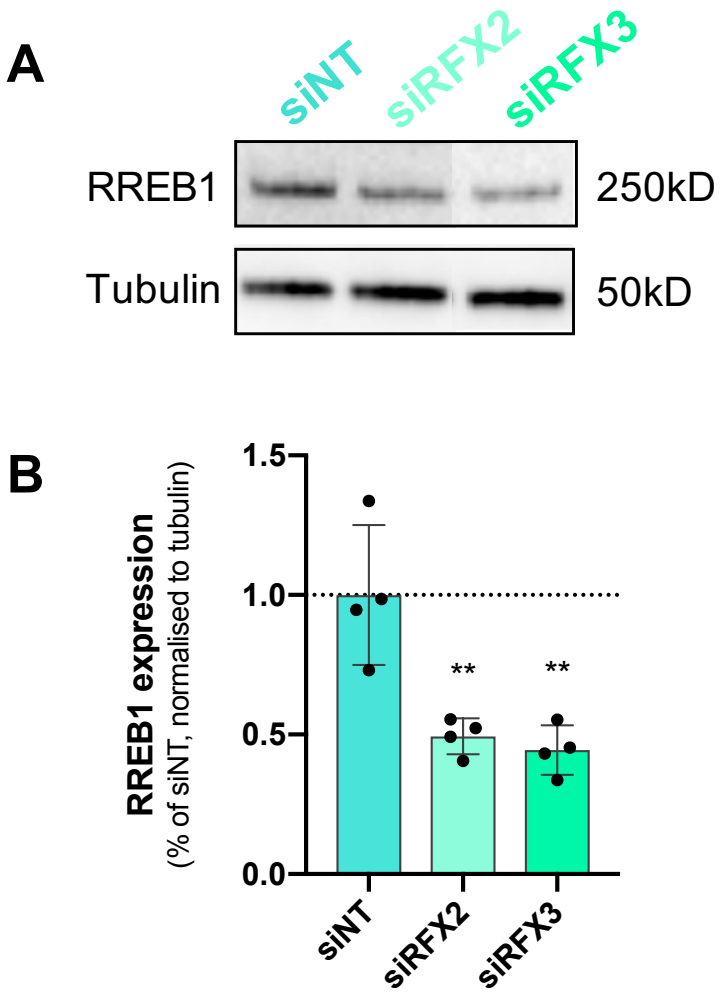


Figure 4.17: RFX depletion affected RREB1 expression in EndoC-βH1 cells (A) Exemplary WB showing reduced RREB1 expression as a consequence of *RFX2* or *RFX3* silencing. (B) Individual siRNA-mediated KD of *RFX2* and *RFX3* caused a significant decrease in RREB1 expression. n=4; values are displayed as mean±SD; *p*-values $^{**}<0.01$ for two-way ANOVA followed by Dunnett's multiple comparisons test (B); EV, turquoise; KO, darkblue.

genome. Functional binding can be determined by combining ChIP-Seq with gene expression profiling (RNA-Seq).

Thus, to determine which DEGs might be *cis*-regulated targets of RREB1, ChIP-Seq was performed in *RREB1* KO EndoC-βH1 cells overexpressing FLAG-tagged RREB1 and integrated with gene expression profiling data which allowed prediction of RREB1 gene expression regulation in beta cells. ChIP was performed with a FLAG antibody, as multiple commercially available RREB1 antibodies that

I tested were of insufficient specificity for IP.

A total of 1,090 peaks were identified in control samples (no RREB1 expression), while 12,912 peaks were detected in RREB1-FLAG samples, of which ∼20% were found in gene promoters, within 3kb of the transcription start site (TSS) (Figure 4.18 A). Evaluation of peak count frequencies within 3kb up- and downstream of the TSS revealed a bell shape-like distribution curve around the TSS only for RREB1-FLAG ChIP-Seq signals (Figure 4.18 B). Functional enrichment analysis of annotated nearest genes identified various cell signalling pathways including 'signalling by receptor tyrosine kinases', 'signalling by VEGF' or 'netrin-1 signalling' as well as processes at synapses such as 'transmission across chemical synapses' and 'protein-protein interactions at synapses' (Figure 4.18 C), matching enriched biological terms ('synaptic vesicle exocytosis', 'transmission across chemical synapses', 'protein-protein interactions at synapses', 'cAMP signalling' and 'MAPK signalling pathway'), identified among DEGs up-regulated in *RREB1* KO beta cells (see Table 4.5).

Identified RREB1-FLAG peaks were annotated to 7,358 unique genes, of which 658 (∼38%) overlapped with DEGs up-regulated and 244 (∼28%) with DEGs down-regulated in *RREB1* KO EndoC-βH1 cells (p=1.51x10^{-164} and p=3.02x10^{-33}, respectively). RREB1-FLAG ChIP-Seq signals mapped to some of the strongest up-regulated DEGs including *CEND1* (q=1.25x10^{-143}, $\log_2$FC=5.0878), *CELF5* (q=1.30x10^{-78}, $\log_2$FC=4.1932), *RNF112* (q=4.84x10^{-96}, $\log_2$FC=4.2242), *TTBK1* (q<2.23x10^{-308}, $\log_2$FC=3.419) and *ASTN1* (q=6.17x10^{-62}, $\log_2$FC= 3.6581), as well as to some of the strongest down-regulated DEGs such as *ADCYAP1* (q=4.91x10^{-23}, $\log_2$FC=-1.900), *CD9* (q=7.46x10^{-19}, $\log_2$FC=-1.7968), *COL21A1* (q=1.63x10^{-11}, $\log_2$FC=-1.6497), *KITLG* (q=1.36x10^{-15}, $\log_2$FC=-1.5781) and *MSS51* (q=3.12x10^{-11}, $\log_2$FC=-1.4943). Comparison of computationally predicted and experimentally identified RREB1 *cis*-regulated genes revealed a significant overlap (p<1.00x10^{-414}). In total 2,528 of the ChIP-Seq identified genes (34.4%) were found among predicted RREB1 target genes.

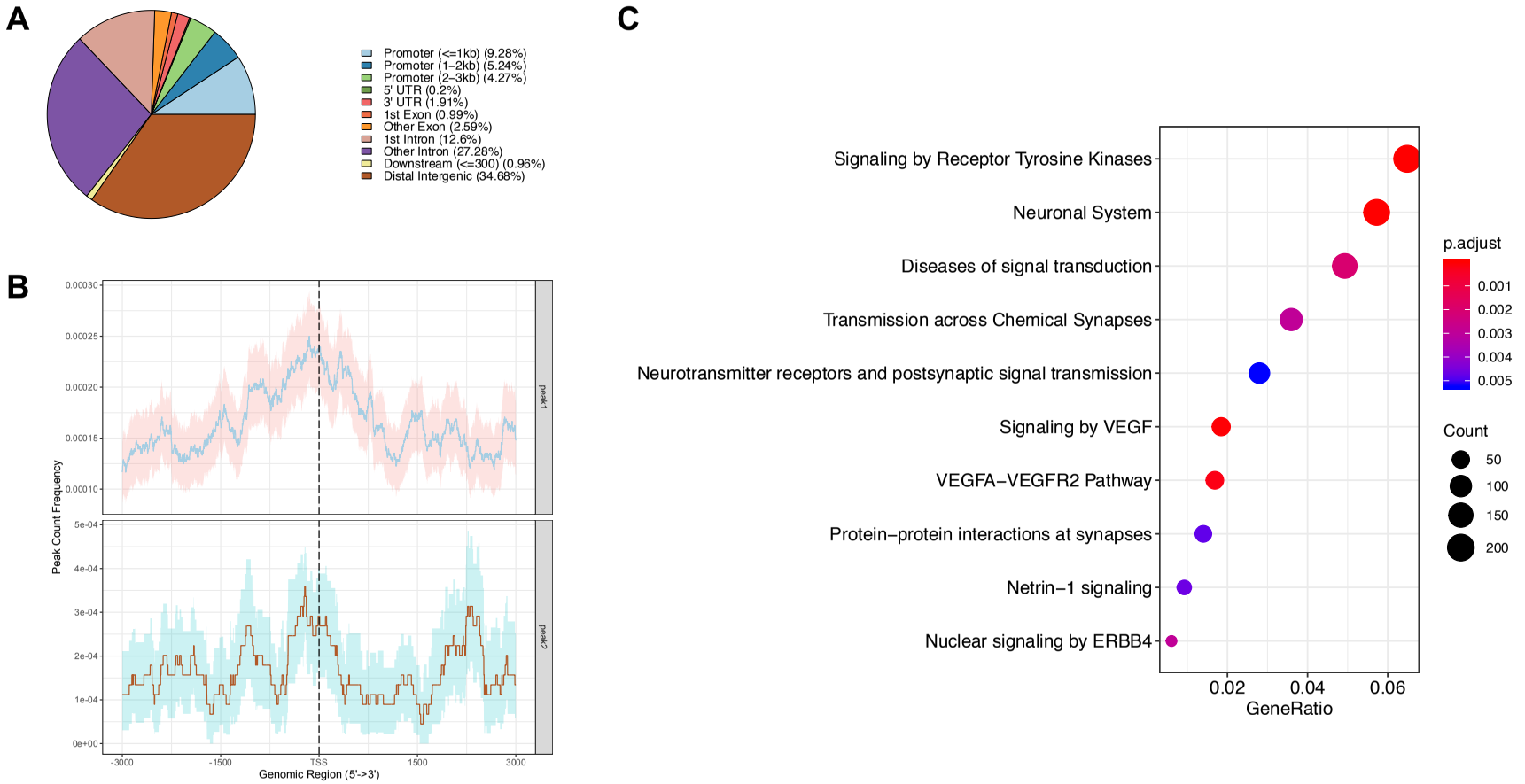


Figure 4.18: Genomic annotations and functional enrichment analysis of ChIP-Seq peaks (A) Genomic annotation of ChIP-Seq peaks. (B) RREB1-FLAG (top) and control (bottom) peak count frequencies within 3kb of the TSS displayed as mean and 95% CI. (C) Reactome pathways identified in functional enrichment analysis. TSS, transcription start site. Figures were generated by Dr Toryn Poolman.

Table 4.8 lists a selected subset of ChIP-Seq signals including their annotated genes and genomic locations as well as the effect on gene expression as a consequence of loss of RREB1 in EndoC-βH1 cells. Two unique peaks were annotated to *RREB1*, one mapping to the *RREB1* promoter ($\leq$1kb) and the other one in an intergenic region distal of the *RREB1* gene, strongly suggesting RREB1 as a regulator of its own gene expression and explaining the significant up-regulation of *RREB1* transcript levels in *RREB1* KO beta cells.

Both *RFX2* and *RFX3* have been reported to contain TF binding sites in their promoter regions for RREB1 (Matys et al. 2006; Marbach et al. 2016). Five peaks were detected in the genomic regions flanking *RFX2*, two mapped to the promoter region ($\leq$1kb and 1-2kb from the TSS, respectively), two to intronic and one to a distal intergenic region (Table 4.8). Expression of *RFX2* was significantly up-regulated in *RREB1* KO beta cells, suggesting RREB1 as transcriptional repressor. No peak was annotated to *RFX3*, fitting with unchanged transcript levels, however, a ChIP-Seq signal was observed just upstream of *RFX3*, annotated to *LINC01231*.

RREB1 binding was also observed in distal intergenic regions of the *REST* and *RCOR2* promoters as well as in an intronic region of *RCOR1*, suggesting RREB1 as direct regulator of these three genes. *RCOR1* and *RCOR2* encode REST co-repressors, which bind to the C-terminal domain of REST to form a repressor/co-repressor complex involved in neural gene expression regulation (Andrés et al. 1999). While expression of *RCOR1* and *RCOR2* was unaffected by loss of RREB1, *REST* transcript levels were hardly detectable in beta cells, but REST repressive activity was significantly down-regulated in *RREB1* KO EndoC-βH1 cells. Interestingly, there was a substantial overlap of RREB1 *cis*-regulated genes and REST target genes (p=7.86x10^{-290}). A total of 1,713 genes (23.3%) were identified to be regulated by both TFs, including *RREB1* itself. Surprisingly, iRegulon only predicted REST (not RREB1) as upstream regulator of newly identified RREB1 *cis*-regulated genes (NES=3.0), as well as of the subset regulated by both RREB1 and REST (NES=3.5-7.3).

Table 4.8: Subset of annotated ChIP-Seq peaks in RREB1-FLAG samples and integration with gene expression profiling in *RREB1* KO EndoC-βH1 cells

Gene	# of peaks	Genomic annotation	Expression in ***RREB1*** **KO beta cells**
RREB1	2	promoter (≤1kb) distal intergenic	up-regulated
RFX2	5	promoter (≤1kb) promoter (1-2kb) intron (2 peaks) distal intergenic	up-regulated
REST	1	distal intergenic	unchanged
RCOR1	1	intron	unchanged
RCOR2	1	distal intergenic	unchanged
KLF4	4	distal intergenic (4 peaks)	unchanged
NANOG	1	distal intergenic	not expressed
GATA4	1	distal intergenic	unchanged
GATA6	1	distal intergenic	not expressed
HHEX	1	distal intergenic	not expressed
HNF1B	2	exon distal intergenic	hardly expressed
HNF4G	2	distal intergenic (2 peaks)	unchanged
NOTCH2	2	intron distal intergenic	unchanged
NKX6.1	1	distal intergenic	unchanged
NKX2.2	1	promoter (2-3kb)	down-regulated
MAFB	13	promoter (1-2kb) distal intergenic (12 peaks)	down-regulated
GCK	1	exon	up-regulated
PCSK1	2	intron (2 peaks)	up-regulated
CHGB	2	distal intergenic (2 peaks)	up-regulated
ABCC8	1	exon	unchanged
KCNJ11	1	3' UTR	unchanged
SNAP25	1	intron	up-regulated
SCG2	3	distal intergenic (3 peaks)	up-regulated
CACNA1A	1	intron	up-regulated
CACNA1B	1	intron	up-regulated
CACNA1C	4	intron (3 peaks) exon	up-regulated

Continued on next page

Gene	# of peaks	Genomic annotation	Expression in ***RREB1*** KO beta cells
CACNA1D	2	intron exon	up-regulated
CACNA1E	2	intron (2 peaks)	up-regulated
GJD2	1	distal intergenic	up-regulated
NCAM1	3	distal intergenic (3 peaks)	up-regulated
PTPRN	1	promoter (≤1kb)	up-regulated
PDGFRA	1	promoter (≤1kb)	down-regulated
ZFP36L1	1	distal intergenic	down-regulated

From top: genes encoding TFs and associated proteins involved in endocrine differentiation, genes involved in insulin exocytosis, genes involved in cell-to-cell communication, forbidden genes in beta cells.

In light of this unexpected finding, I next sought to determine whether RFX TFs also shared common target genes with RREB1. ChIP-Seq-identified RREB1 *cis*-regulated genes were enriched for both RFX2 and RFX3 predicted target genes with 1,202 (16.3%) and 1,526 (20.7%) suggested to be equally regulated by RREB1, respectively (p=1.86x10^{-175} and p=1.66x10^{-212}, respectively). Again, TF enrichment analysis using iRegulon only predicted members of the RFX TF family, but not RREB1, as putative upstream regulators of overlapping genes (NES=3.1-9.1 for RFX2/RREB1 shared genes and NES=3.1-7.3 for RFX3/RREB1 common genes).

DEGs associated with insulin secretion, including *GCK*, *PCSK1*, *CHGB*, *SNAP25*, *SCG2* and genes encoding voltage-sensitive Ca^{2+} channel subunits (*CACAN1A*, *CACNA1B*, *CACAN1C*, *CACAN1D* and *CACAN1E*), all up-regulated in *RREB1* KO EndoC-βH1 cells, contained RREB1-FLAG ChIP-Seq signals, highlighting RREB1 as a direct transcriptional repressor of these genes. ChIP-Seq peaks were also annotated to up-regulated DEGs involved in cell-to-cell communication such as *GJD2*, *NCAM1* and *PTPRN*. Additionally, binding of RREB1 was found in two of the markedly down-regulated beta cell forbidden genes (*PDGFRA* and *ZFP36L1*), suggesting RREB1 as a direct regulator which activates transcription. Furthermore, ChIP-Seq peaks were annotated to *ABCC8* and *KCNJ11*, genes encoding K_{ATP} channel subunits, however, expression of these genes was not affected by loss of RREB1.

A subset of ChIP-Seq peaks were annotated to genes encoding TFs that perform crucial biological functions in early beta cell development including hPSC-specific genes *KLF4* and *NANOG* or genes involved in endoderm formation such as *GATA4* (Xuan et al. 2016), *GATA6* (Xuan et al. 2016; Tiyaboonchai et al. 2017; Fisher et al. 2017) and *HHEX* (Rankin et al. 2011). While *HHEX* was not expressed in EndoC-βH1 cells, transcript levels were significantly increased in *RREB1* KO hiPSC-derived DE cells (q=0.0078, $\log_2$FC=0.6748), suggesting RREB1 as transcriptional repressor of *HHEX*. In addition, hepatocyte nuclear factor genes *HNF1B* and *HNF4G*, as well as *NOTCH2* and *NKX6.1*, encoding TFs active in pancreatic progenitors contained binding sites for RREB1. While *HNF1B* was hardly expressed in mature beta cells, it was significantly down-regulated in *RREB1* KO cells at stage 2 (PGT, q=0.0045, $\log_2$FC=-0.5457) and stage 5 (EP, q=0.0032, $\log_2$FC=-0.6167). Interestingly, while *NOTCH2* transcript levels were unaffected by loss of RREB1 in mature beta cells, expression was markedly up-regulated in *RREB1* KO cells at stage 2 (PGT, q=1.99x10^{-4}, $\log_2$FC=0.6167) and significantly down-regulated at stage 5 (EP, q=0.0044, $\log_2$FC=-0.7484). Endocrine progenitor marker genes *NKX2.2* and *MAFB* were also identified as *cis*-regulated RREB1 target genes. Consistent with this, *RREB1* KO endocrine progenitor cells at stage 4 (PE) revealed a significant up-regulation of both *NKX2.2* (q=1.70x10^{-6}, $\log_2$FC=0.8303) and *MAFB* (q=0.0177, $\log_2$FC=0.830) expression and a marked increase in just *NKX2.2* transcript levels at stage 5 (EP, q=1.79x10^{-5}, $\log_2$FC=1.0708), supporting a role for RREB1 in endocrine cell differentiation.

4.3.8.1 RREB1 binding sites were enriched in islet promoter regions

As T2D GWAS signals have been shown to be enriched in islet regulatory regions where they modulate TF binding sites, we next mapped RREB1 binding sites to epigenomic annotations in islets, beta cells and pancreatic progenitors including active promoters, open chromatin sites identified in ATAC-Seq analysis, different enhancer classes (active, inactive, stretch and super enhancers), genomic regions bound

Table 4.9: Enrichment of RREB1 binding sites in epigenomic features identified in human islets, beta cells and pancreatic progenitors

Epigenomic feature	Source	Cell type	Enrichment factor	padj (q)
Active promoters	Miguel-Escalada et al. 2019	human islets	24.2588	9.99x10 $^{-4}$
ATAC-Seq islets Stitzel	Lawlor et al. 2019	human islets	11.1940	0.0020
Consistent set of mediator peaks	Miguel-Escalada et al. 2019	human islets	9.6540	9.99x10 $^{-4}$
ATAC-Seq islets Oxford	Thurner et al. 2018	human islets	9.4470	9.99x10 $^{-4}$
Accessible chromatin Gaulton	Greenwald et al. 2019	human islets	9.4139	9.99x10 $^{-4}$
Consistent CTCF peaks	Miguel-Escalada et al. 2019	human islets	7.7234	9.99x10 $^{-4}$
ATAC-Seq Endos Stitzel	Lawlor et al. 2019	EndoC-βH1 cells	5.0083	0.0240
ATAC-Seq consistent peaks	Miguel-Escalada et al. 2019	human islets	4.8172	9.99x10 $^{-4}$
Strong CTCF	Miguel-Escalada et al. 2019	human islets	4.1812	9.99x10 $^{-4}$
Consistent H3K27ac peaks	Miguel-Escalada et al. 2019	human islets	2.5331	9.99x10 $^{-4}$
Inactive open chromatin	Miguel-Escalada et al. 2019	human islets	2.0017	9.99x10 $^{-4}$
Inactive enhancers	Miguel-Escalada et al. 2019	human islets	2.0004	9.99x10 $^{-4}$
Active enhancers I	Miguel-Escalada et al. 2019	human islets	1.6183	9.99x10 $^{-4}$
Open strong enhancers	Thurner et al. 2018	human islets	1.5935	9.99x10 $^{-4}$
PATs	Miguel-Escalada et al. 2019	human islets	1.5073	9.99x10 $^{-4}$
TAD-like domains	Miguel-Escalada et al. 2019	human islets	1.4958	9.99x10 $^{-4}$
Active enhancers III	Miguel-Escalada et al. 2019	human islets	1.4791	0.0040
Super enhancers	Miguel-Escalada et al. 2019	human islets	1.3936	0.0030
Active enhancers II	Miguel-Escalada et al. 2019	human islets	1.2290	0.0120
MPC selective enhancers	Cebola et al. 2015	human pancreatic multipotent progenitors	1.0897	0.3407
MPC enhancers	Cebola et al. 2015	human pancreatic multipotent progenitors	1.0547	0.3227
Active enhancers	Pasquali et al. 2014	human islets	0.9941	0.5464
Stretch enhancers	Parker et al. 2013	human islets	0.9574	0.85514

ATAC-Seq, reveals open chromatin sites (promoters and enhancers); CTCF, CCCTC-binding factor (forms chromatin loops); H3K27ac, active enhancers; active enhancers I-III, subclassification according to Mediator, cohesin and H3K27ac occupancy patterns; active enhancers I, high H3K27ac and Mediator occupancy; open strong enhancers, H3K4me1 and H3K27ac marks; PATs, promoter associated 3D spaces; TAD-like domains, topologically associating domains (self-interacting genomic regions); super enhancer, high levels of H3K27ac marks; MPC, pancreatic multipotent progenitors; MPC selective enhancers, active enhancers that do not overlap with active enhancers in at least 6 of 7 non-pancreatic tissues; stretch enhancer, enhancer $\geq$3kb.

by CTCF, promoter associated three dimensional spaces (PAT) and topologically associating domains (TADs) (Table 4.9). Permutation-based enrichment analysis revealed that RREB1 binding was significantly enriched in islet open chromatin, in particularly in active promoters (x24.26, q=9.99x10^{-4}). While there also seemed to be a slight, but still significant enrichment for islet enhancers, RREB1 binding sites were not enriched at islet super or stretch enhancers (Figure 4.19).

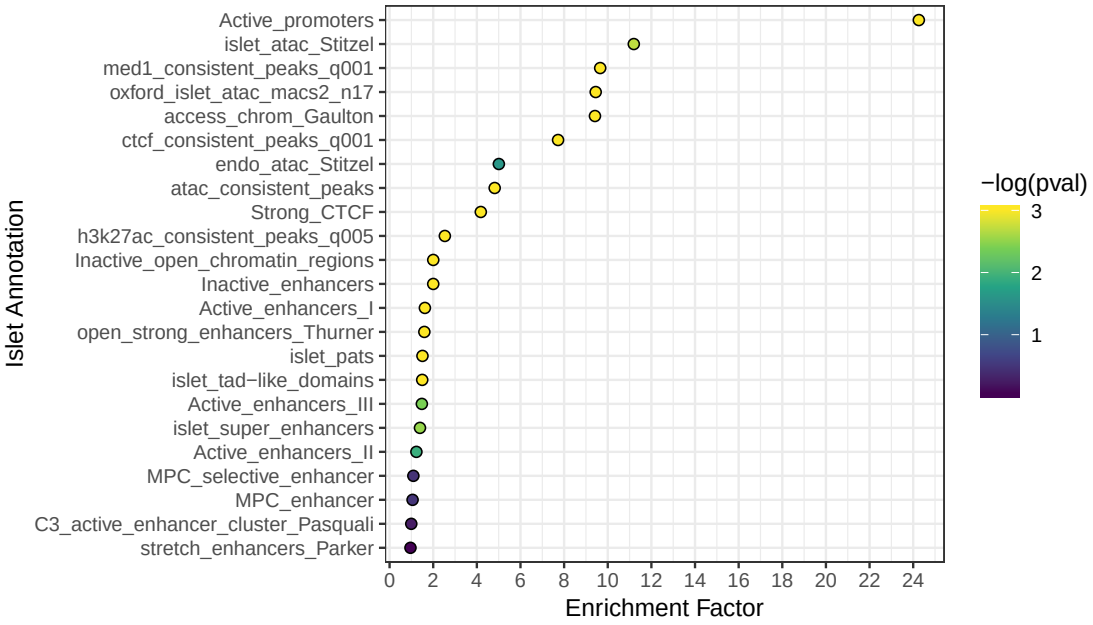


Figure 4.19: RREB1 binding sites were enriched in active islet promoters Enrichment of newly identified RREB1 binding sites in islet, beta cell and pancreatic progenitor regulatory regions. MPC, multipotent pancreatic progenitors; TSS, transcription start site. Graph was generated by Dr Jason Torres.

4.4 Discussion

Loss of RREB1 has been associated with a positive effect on endocrine cell differentiation as discussed in Chapter 3. Less is known about the TF's function in mature beta cells. Here, I have addressed the functional consequences of inactivating RREB1, either partially and transiently via siRNA-mediated gene silencing (KD) or completely and permanently, using a pooled lentiviral CRISPR/Cas9 targeting approach (KO), in the glucose-responsive insulin-secreting human beta cell line EndoC-βH1.

Loss of RREB1 modifies insulin content in mature beta cells through a yet unknown mechanism

Partial ($\sim$78% reduction) or complete loss of RREB1 in EndoC-βH1 cells resulted in reduced *INS* transcript levels and significantly lower cellular insulin content, but did not affect the amount of insulin secreted in basal and high glucose. Under higher insulin demand conditions (high glucose in combination with cAMP-elevating agent forskolin) RREB1-depleted beta cells secreted significantly less insulin compared to control beta cells and after depletion of docked insulin granules, insulin release in response to a glucose stimulus remained markedly lower. By adjusting the amount of secreted insulin for insulin content these apparent deficits could be cancelled out, suggesting that loss of RREB1 had presumably not negatively affected the beta cell secretory machinery.

Reduced cellular insulin content could possibly be a consequence of defective *INS* gene transcription, impaired proinsulin processing or enhanced degradation of insulin granules. *INS* gene expression was compromised in *RREB1* KD and KO EndoC-βH1 cells. Transcriptional activity of NEUROD1 and PDX1, well established regulators of the *INS* promoter (Glick et al. 2000), were markedly reduced as a consequence of RREB1 loss in mature beta cells, potentially underlying significantly lower *INS* transcript levels.

Down-regulated DEGs in *RREB1* KD EndoC-βH1 cells were enriched for biological processes associated with the ER, which in beta cells is an essential cellular component in insulin synthesis, i.e. in the ER proinsulin is folded and its disulphide bonds are formed before it gets transported to the Golgi complex for further maturation. Accumulation of unfolded proinsulin leads to ER stress. To adapt to stress conditions, beta cells activate the UPR pathway which regulates various processes to reduce ER workload (Goodge et al. 2000; Scheuner et al. 2005; Ron et al. 2007; Oslowski et al. 2010). Reduction of insulin production has been shown to have a positive impact on ER stress in beta cells, thus down-regulation of genes involved in ER stress might be reflective of the decrease in *INS* gene expression observed in *RREB1* KD EndoC-βH1 cells (Szabat et al. 2016). Less ER stress in line with up-regulation of *PCSK1* and *PCSK2*, genes encoding the proprotein convertases 1 and 2, which are responsible for proinsulin processing in secretory granules (Kaufmann et al. 1995), indicates that reduced insulin content was likely not attributable to impaired proinsulin processing in RREB1-deficient beta cells.

Lysosomes have been implicated in the degradation of proinsulin and insulin (Pasquier et al. 2019). Neither their occurrence or appearance seemed to be different between *RREB1* KO and EV EndoC-βH1 cells. This might indicate that it is rather unlikely that increased insulin degradation is underlying the observed reduction in cellular insulin content in RREB1-depleted beta cells. It seems more likely that down-regulation of *INS* gene transcription is responsible for reduced cellular insulin content.

RREB1, a potential novel regulator of genes important for beta cell function and identity

Here, I report for the first time, to my knowledge, *cis*-regulated target genes of RREB1 in mature beta cells. Integration of ChIP-Seq experiments with gene expression profiling revealed a considerable set of genes up-regulated in RREB1-deficient beta cells involved in glucose sensing, insulin processing and granule

exocytosis (*GCK*, *PCSK1*, *SNAP25*), which were directly regulated by RREB1. Furthermore, genes implicated in beta cell connectivity (*GJD2*), beta cell maturity (*UCN3*) and cell-cell adhesion (*NCAM1*, *CDH1*) as well as genes displaying pro-survival activity in beta cells (*CDK5R2*, *PTPRN*) showed markedly higher expression in RREB1-deficient beta cells. Among those, binding of RREB1 was observed for *GJD2*, *NCAM1* and *PTPRN*.

Parallel to increased beta cell maturity gene expression, expression of beta cell disallowed genes (*PDGFRA* (RREB1 *cis*-regulated gene), *IGFBP4*, *CD302*) was markedly reduced in *RREB1* KO EndoC-βH1 cells. Repression of disallowed genes has been suggested as essential for beta cells to maintain their identity and functional maturity (Pullen et al. 2013). Loss of RREB1, therefore, seemed to have a positive effect on the differentiation state of mature beta cells, which was rather unexpected considering the insulin content phenotype of *RREB1* KO EndoC-βH1 cells.

Mapping of RREB1 TFBS to regulatory elements in human islets and beta cells revealed that RREB1 binding sites were enriched in active promoters, further supporting a crucial role for RREB1 in beta cell function and representing an important finding for future T2D GWAS loci fine-mapping efforts, as it has previously been shown that T2D- and FG-associated variants are enriched in islet regulatory elements (Parker et al. 2013; Pasquali et al. 2014; Thurner et al. 2018).

Loss of RREB1 modulates transcriptional activity of REST and RFX TFs in mature beta cells

In the *RREB1* KO beta cell developmental model, REST was identified as differentially active TF, likely co-responsible for enhanced endocrine differentiation of *RREB1* KO cells (Chapter 3). RREB1-depleted mature beta cells equally revealed markedly increased REST transcriptional activity, and ChIP-Seq analysis suggested binding of RREB1 to an intergenic region distal of the *REST* promoter. *REST* was hardly expressed in EndoC-βH1 cells, as expected for a beta cell disallowed

gene, but still tended to be even lower in RREB1-depleted compared to control beta cells, supporting RREB1 as modulator of REST.

Loss of RREB1 in EndoC-βH1 cells resulted in differential activity and expression of RFX2 and RFX3. While RFX3 has been reported to play an important role in the regulation of beta cell development and function (Ait-Lounis et al. 2007; Ait-Lounis et al. 2010), a relevant role for RFX2 in glucose homeostasis has not been described, yet. For example, mouse adult beta cells depleted of *Rfx3* were characterised by significantly reduced levels of *Slc2a2* and *Gck* and displayed elevated basal insulin secretion, while cellular insulin content and expression of *Nkx6.1*, *Nkx2.2* or *Pcsk1* were not affected by loss of RFX3 (Ait-Lounis et al. 2010). *RFX2* was identified as direct target of RREB1 and expression was significantly up-regulated as a consequence of loss of RREB1, raising questions whether RFX2 might play a role in mature beta cells. In vertebrates RFX2 and RFX3 have been shown to have similar tissue expression patterns and contain functional domains allowing them to dimerise as either homo- or heterodimers (Reith et al. 1994; Emery et al. 1996a; Horvath et al. 2004; Rual et al. 2005; Aftab et al. 2008). Both RFX2 and RFX3 have been reported to be expressed in ciliated tissues and modulate FOXJ1-mediated cilia gene regulation in vertebrates (Bonnafe et al. 2004; Chung et al. 2012; Didon et al. 2013; Quigley et al. 2017). Pancreatic endocrine cells contain cilia, but their function is still unknown. Islet cells of *Rfx3* KO mice were characterised by a reduced number and shorter cilia (Ait-Lounis et al. 2007), however, whether defective cilia contributed to impaired endocrine cell development and beta cell function remains to be elucidated. RREB1-depleted mature beta cells were characterised by a significant increase in RFX2/3 motif activity. To assess, whether RFX2 or RFX3 might have contributed to the transcriptional phenotype of *RREB1* KO EndoC-βH1 cells, I profiled gene expression in beta cells depleted of either RFX genes alone or RFX and RREB1. siRNA-mediated individual silencing of *RFX2* or *RFX3* in *RREB1* KO and EV EndoC-βH1 cells, highlighted RFX3 as the most likely transcriptional regulator affected by RREB1 deficiency. Merely

reduced RFX3 expression resulted in partial rescue of the transcriptional phenotype associated with loss of RREB1 in beta cells.

A substantial number of DEGs up-regulated in *RREB1* KO beta cells, were suggested to be *cis*-regulated or computationally-predicted target genes of the transcriptional repressor REST as well as of the TFs RFX2 and RFX3. Due to hitherto insufficient publicly available data on RREB1-regulated genes in mature beta cells, it might be possible that DEGs were falsely interpreted as REST or RFX target genes in upstream regulator prediction analysis. In support of this, TF enrichment analysis using iRegulon did not predict RREB1 as upstream regulator of newly identified RREB1 *cis*-regulated genes, nor of the subsets of genes suggested to be regulated by RREB1 and REST, RREB1 and RFX2 or RREB1 and RFX3, while REST and the RFX TF family were identified as key transcriptional mediators.

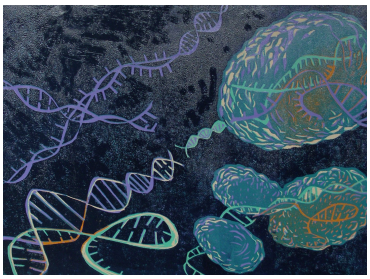
Conclusions

Loss of RREB1 resulted in beta cells with a somewhat 'confused' identity. On the one hand, RREB1-depleted EndoC-βH1 cells displayed a few characteristics resembling immature beta cells including reduced insulin content (JSE Liu et al. 2017) and down-regulation of *NEUROD1* (C Gu et al. 2010). Additionally, expression of glucose transporter-encoding genes *SLC2A1* and *SLC2A2* were reduced. On the other hand, up-regulation of maturity genes, paralleled by down-regulation of disallowed genes and lower REST transcriptional activity as well as increased insulin secretion in response to a glucose stimulus, when adjusted for reduced cellular insulin content, were reminiscent of mature beta cell features. In light of this and considering the observations that reduced insulin production causes less ER stress and less beta cell fragility (Szabat et al. 2016), loss of RREB1 seemed to have had a positive rather than a negative impact on mature beta cells.

While ChIP-Seq revealed RREB1 as a novel regulator of genes implicated in insulin secretion as well as beta cell identity, the transcriptional phenotype of RREB1-depleted adult beta cells could also, at least in part be attributed to increased

transcriptional activity of RFX3, strongly highlighting RREB1 as modulator of this TF. As RREB1 could not be confirmed as transcriptional regulator of the *RFX3* gene, the mechanism underlying up-regulated RFX3 motif activity is currently unknown.

The exact regulatory pathways or networks underlying reduced *INS* expression equally remain to be elucidated. ChIP-Seq experiments did not reveal binding of RREB1 to the *INS* promoter, thus, down-regulated *INS* transcription was likely mediated by an effect of RREB1 on other TFs. Potential mediators include NEUROD1 and PDX1, as their transcriptional activities were negatively affected as a consequence of loss of RREB1. Especially, NEUROD1, which has been shown to co-occupy the *Ins1* and *Ins2* promoters together with RREB1 and the C-terminal-binding protein (CtBP) in the murine pancreatic beta cell line β-TC6 represents a very promising candidate (Ray et al. 2014).



5

Effects of T2D-associated *RREB1* alleles on beta cell development and function

5.1 Introduction

The *RREB1* locus contains multiple GWAS signals, which share association between T2D risk and glycaemic traits (Scott et al. 2012; Mahajan et al. 2015; Mahajan et al. 2014; Fuchsberger et al. 2016; Mahajan et al. 2018a; Mahajan et al. 2018b). Fine-mapping specified three independent signals for T2D association, represented by a non-coding (rs112498319), a non-coding (rs9505097) in LD with a coding (rs35742417, p.S1499Y), and a coding (rs9379084, p.D1171N) variant, which were identified through meta-analysis of GWAS data in individuals of European ancestry performed by the T2D-GENES and GoT2D consortia (Mahajan et al. 2015; Fuchsberger et al. 2016; Mahajan et al. 2018a and Anubha Mahajan, personal communication) (see Figure 1.5). While subsequent coding variant fine-mapping confirmed p.D1171N (MAF=11%, OR=1.09) as causative (posterior probability of driving T2D association PPA=0.92) and highlighted *RREB1* as effector transcript, causality could not be established for p.S1499Y (MAF=21.1%, OR=1.04, PPA<0.001) (Mahajan et al. 2018b). One of the 20 fine-mapped non-coding regulatory variants in LD with p.S1499Y are more likely to be driving

the T2D and FG levels associations (Anubha Mahajan, personal communication). For the third independent signal, fine-mapping has identified a credible set of 200 variants. Of these rs11759956 has been shown to be located in an islet promoter and therefore represents the most likely causal regulatory variant underlying the T2D association (Anubha Mahajan, personal communication).

For the single established causal coding variant p.D1171N the minor allele p.N1171 is associated with reduced T2D risk and lower FG levels, suggesting a protective effect and pointing to a likely role for pancreatic beta cells in diabetes susceptibility. Interrogation of the developmental and functional consequences of loss of RREB1 during beta cell differentiation as well as in mature beta cells, described in Chapters 3 and 4, suggest a role for RREB1 in endocrine cell development as well as in beta cell function, supporting a potential impact of the T2D-associated *RREB1* p.D1171N variant on beta cells.

Human PSC-derived beta cell models have been successfully applied to interrogate monogenic diabetes-associated SNPs and thereby shed light on putative molecular mechanisms underlying disease pathogenesis (Saarimäki-Vire et al. 2017; Balboa et al. 2018; Ng et al. 2019). One of the first studies interrogating the role of a GWAS variant to gain more insights into the potential mechanisms underlying the T2D risk association using a hiPSC-based model, was performed at the *SLC30A8* locus (Dwivedi et al. 2019). Dwivedi et al. introduced the protein truncating p.R138X variant into a hiPSC line using CRISPR/Cas9 and differentiated successfully edited clones along the endocrine lineage into BLCs. *SLC30A8* expression was reduced in BLCs heterozygous for p.R138X and cyclohexamide studies showed that nonsense mediated decay of the protective p.R138X allele led to haploinsufficiency for the *SLC30A8* gene (Dwivedi et al. 2019).

Human research islet programmes, facilitating researchers broad access to gene expression, chromatin state and functional data from large collections of human islets, have opened new opportunities to interrogate the impact of diabetes-associated genetic variants on human islet function (Kaddis et al. 2009; Niland et al. 2010;

Kulkarni et al. 2014; Lyon et al. 2015; Mularoni et al. 2017; NJ Hart et al. 2019). For example, by assessing insulin secretion measurements in islets of individuals heterozygous for a low-frequency variant in *PAM* and age-, sex- and BMI-matched donors, it could be shown that the T2D risk allele at rs35658696 significantly reduced insulin secretion in response to a glucose stimulus, confirming observations in cellular *PAM* KD studies and highlighting a potential mechanism underlying the T2D risk association at the *PAM* locus (Thomsen et al. 2018).

5.1.1 Experimental aims

The impact of T2D-associated *RREB1* alleles on RREB1 expression and function are currently unknown. To investigate whether the T2D risk and protective alleles have differential effects on RREB1 protein expression and/or stability, RREB1 p.D1171 and RREB1 p.N1171 protein was produced in HEK293 cells, which represent a suitable, easy-to-manipulate human expression system.

As complete loss of RREB1 had shown a positive impact on beta cell development (Chapter 3), I was interested, whether the T2D protective allele also affected endocrine cell differentiation. Therefore, the SB Ad3.1 hiPSC line, which had previously been shown to be heterozygous for the *RREB1* T2D-associated variant rs9379084 (c.3511G>A, p.D1171N), was genetically engineered to be homozygous for the T2D protective allele (p.N1171N) using CRISPR/Cas9 genome engineering. Generation of *RREB1* hiPSC lines homozygous for the T2D risk allele (p.D1171D, referred to as *RREB1* WT lines in Chapter 3) has already been described in Section 3.3.1. Parallel differentiation of *RREB1* N1171N, D1171N and D1171D hiPSC lines along the endocrine lineage into BLCs allowed to assess the impact of the full *RREB1* allelic spectrum at rs9379084 on beta cell differentiation, and comparison to differentiated *RREB1* KO clones shed light on the putative functional effect of the T2D protective allele.

To study the impact of the T2D protective allele on beta cell function, insulin secretion and content measures in islets of donors heterozygous for p.D1171N were compared to those from individuals homozygous for the T2D risk allele (p.D1171D).

5.2 Methods

5.2.1 RREB1 expression studies

A mutation corresponding to the T2D protective allele of p.D1171N was introduced into pCMV6-*RREB1* isoform 1, 2 and 3 expression plasmids via site-directed mutagenesis as described in Section 2.1.1. HEK293 cells were transfected with 1μg pCMV6-*RREB1* isoform 1, 2, or 3 encoding either p.D1171 (T2D risk allele) or p.N1171 (T2D protective allele) plasmid DNA, harvested either 24 or 48h post-transfection and analysed for RREB1 expression via WB using an aFLAG monoclonal antibody (Sigma Aldrich, F3165) (see Section 2.1.3). Statistical significance was tested using a two-way ANOVA followed by Sidak's multiple comparisons test.

5.2.2 Genome editing

A sgRNA targeting as close as possible to the region of *RREB1* exon 10 encoding aa 1171 was designed using the MIT CRISPR online design tool (http://crispr.mit.edu). Oligonucleotides encoding the target site were purchased from Eurofins Genomics and subcloned into pX330-Puro to produce pX330-Puro-*RREB1*-exon 10 as described in Section 2.3. To generate *RREB1* N1171N hiPSC lines a ssODN repair template was designed and co-transfected with pX330-Puro-*RREB1*-exon 10. Transiently puromycin-resistant cells were genotyped via PCR amplification and Sanger sequencing (see Tables A.1 and A.2). Expression of RREB1 was evaluated by RT-qPCR and WB analysis as described in Sections 2.1.7 and 2.1.3, respectively. Correctly edited cells were evaluated for expression of pluripotency markers via FACS (see Section 2.1.5), genomic integrity via chromosome counting, and screened for common mutations in *TP53* (see Tables A.1 and A.2).

5.2.3 Differentiation experiments

To model pancreas development, a directed seven step *in vitro* differentiation protocol was followed as described in Section 2.4.

5.2.4 Transcriptome profiling

For transcriptome profiling RNA was extracted and processed from differentiated *RREB1* D1171D, D1171N and N1171N cells at seven distinct stages of *in vitro* beta cell development as described in Section 2.5.1. Software packages for RNA-Seq data analysis included the Bioconductor packages RUVSeq (Risso et al. 2014) and DESeq2 (Love et al. 2014) in R v3.3.3. Gene co-expression network analysis was performed using the R software package WGCNA (v.1.51) (Langfelder et al. 2008; B Zhang et al. 2005). Gene ontology and TF binding motif enrichment analysis were done using g:Profiler (Raudvere et al. 2019) and the iRegulon (v1.3) Cytoscape (v3.7.0) plugin (Janky et al. 2014), respectively. In addition, motif activity response analysis (MARA) was applied to identify key TFs mediating gene expression variation across *RREB1* D1171D, D1171N and N1171N samples using the online tool ISMARA (Integrated System for MARA) (Balwierz et al. 2014). Individual analysis methods are described in more depth in Section 2.5.

5.2.5 Insulin secretion assays in primary human islets

Details on the islet isolation process and insulin secretion assays performed at the Alberta Diabetes Institute IsletCore are described in Section 2.1.10. In their database 14 individuals heterozygous for *RREB1* rs9379084 (p.D1171N) were identified, matched to 14 control donors (p.D1171D) for sex, age (p=0.7689), BMI (p=0.9860) as well as cold ischaemia time (CIT, p=0.8441) and evaluated for insulin content and secretion measures. Mean insulin measurements were calculated and are displayed as raw values. Statistical significance between control (D1171D) and D1171N islets was tested using a two-tailed paired t-test.

5.3 Results

5.3.1 Impact of T2D risk and protective alleles on RREB1 protein expression

To assess, whether the T2D-associated alleles at rs9379084 have an effect on RREB1 protein expression, the three protein-coding *RREB1* transcript variants (NM_001003699.3, NM_001003698.3, NM_001003700.1), carrying either the T2D risk (p.D1171) or the protective (p.N1171) allele were individually overexpressed in HEK293 cells. WB analysis revealed that neither transcript variant nor the T2D-associated alleles had a significant impact on RREB1 protein expression 24h or 48h post-transfection (Figure 5.1 A and B).

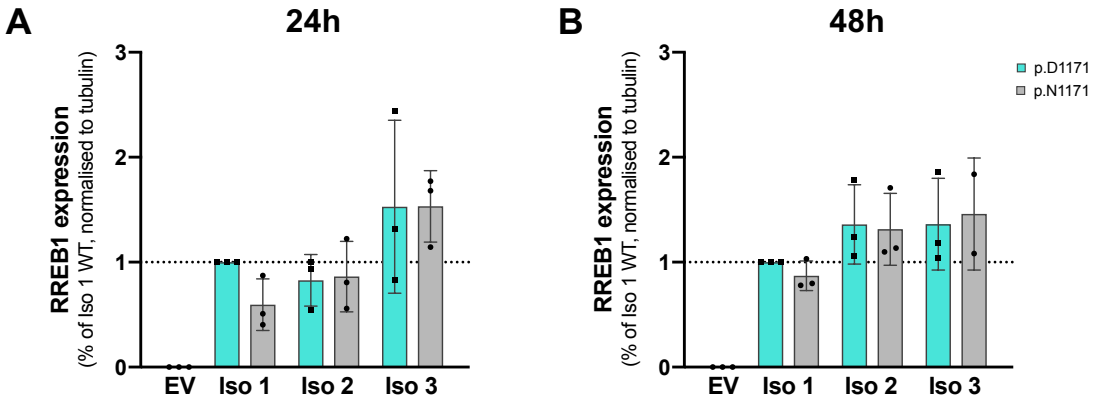


Figure 5.1: Effects of rs9379084 T2D-associated alleles on RREB1 protein expression HEK293 cells were transfected with three different *RREB1* protein-coding transcript variants carrying either the rs9379084 T2D risk (p.D1171) or protective (p.N1171) allele. Neither transcript variant nor T2D-associated allele affected RREB1 protein expression as assessed 24h (A) or 48h (B) after transfection. n=3; values are displayed as mean±SD; two-way ANOVA followed by Sidak's multiple comparisons test (A and B); EV, empty vector control; p.D1171, turquoise; p.N1171, darkgrey.

5.3.2 Generation and characterisation of hiPSC *RREB1* T2D protective allele lines

To characterise the individual impact of the T2D-associated *RREB1* alleles on beta cell development, CRISPR/Cas9 genome engineering was utilised to replace

the T2D risk (c.3511G) with the protective (c.3511A) allele, generating *RREB1* N1171N hiPSC lines homozygous for the protective allele. The genome engineering strategy was similar to the one used to generate *RREB1* D1171D hiPSC clones, except that the ssODN repair template contained the protective allele sequence (c.3511A) (see Section 2.3). To account for potential confounding effects caused by CRISPR/Cas9 genome engineering a *RREB1* D1171N hiPSC clone (*RREB1* D1171N ctrl), which had passed through the CRISPR/Cas9 pipeline, but remained unedited, was characterised in parallel.

5.3.2.1 Genotyping of edited *RREB1* hiPSC lines

Clones showing ssODN-directed repair were identified via PCR amplification of a 447bp fragment and *Hinc*II restriction enzyme digest using the same approach as described in Section 3.3.1.1. Sanger sequencing confirmed ssODN-directed repair on both alleles for three hiPSC clones. Sequencing traces were homozygous for c.3511A at rs9379084 (p.N1171N) and the two silent mutations at c.3510G>C (ssODN-introduced *Hinc*II restriction enzyme site) as well as c.3507G>C (ssODN-introduced PAM modifying mutation) (Figure 5.2 A). As genome edited *RREB1* D1171D lines had equally been characterised by biallelic ssODN-directed repair, resulting *RREB1* D1171D and N1171N lines were isogenic except for the T2D-associated SNP at rs9379084, while the *RREB1* D1171N ctrl line differed in regard to the two ssODN-specific silent mutations (Figure 5.2 A).

5.3.2.2 Characterisation of *RREB1* N1171N hiPSC lines for RREB1 expression and typical hPSC properties

The T2D risk and protective alleles at rs9379084 did not affect *RREB1* transcript levels or RREB1 protein expression in hiPSCs consistent with the observation in RREB1 overexpression studies in HEK293 cells (Figure 5.2 B-D). Table 5.1 gives an overview of the QC results for picked *RREB1* D1171N ctrl and N1171N clones. In the undifferentiated state, *RREB1* D1171D, D1171N ctrl and N1171N lines were characterised by a typical hiPSC colony morphology and >90% of

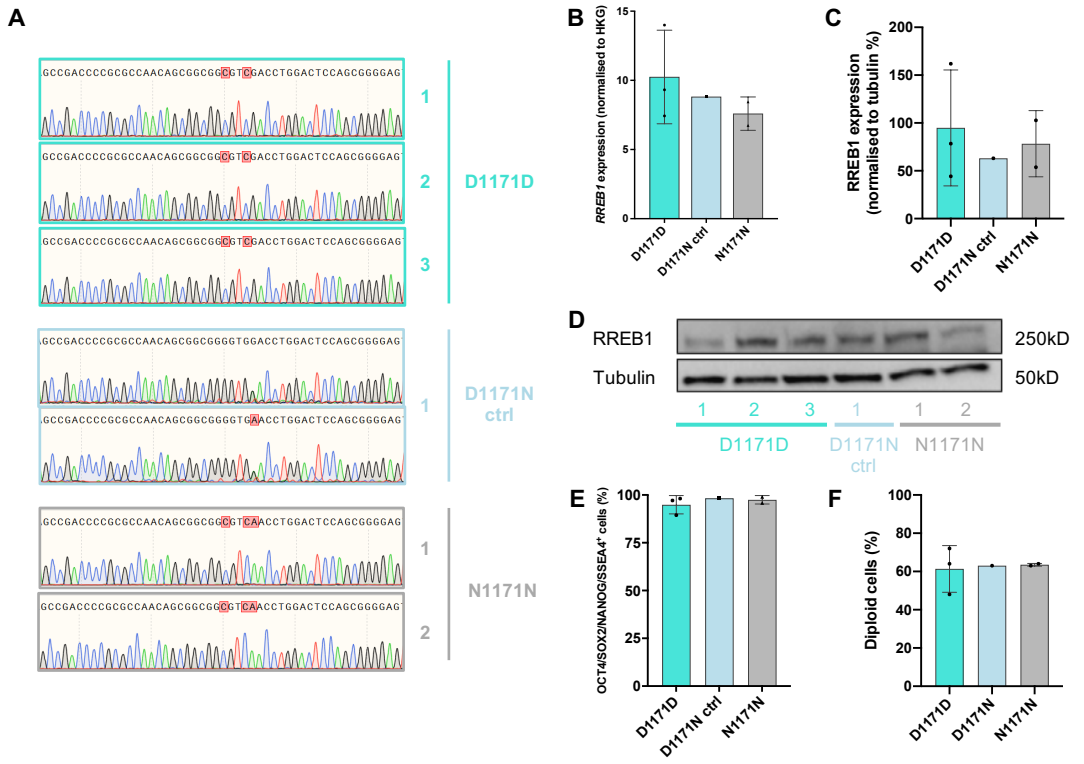


Figure 5.2: Characterisation of *RREB1* T2D variant hiPSC lines for typical stem cell properties (A) Sanger sequencing traces of three independently generated *RREB1* D1171D hiPSC lines (top). All lines were characterised by biallelic 'D1171D' ssODN-directd repair; they revealed homozygosity for c.3511G at rs9379084 and two silent mutations at c.3510G>C (ssODN-introduced *Hinc*II restriction enzyme site) as well as c.3507G>C (ssODN-introduced PAM modifying mutation). Monoallelic sequencing traces of unedited SB Ad3.1 hiPSCs (D1171N ctrl) (middle). Sequencing confirmed biallelic 'N1171N' ssODN-directed repair for two hiPSC lines. (B) *RREB1* mRNA expression in *RREB1* D1171D, D1171N ctrl and N1171N hiPSC clones. (C) Quantification of RREB1 protein expression for different *RREB1* hiPSC lines. (D) RREB1 WB analysis of *RREB1* D1171D, D1171N ctrl and N1171N hiPSC lines. Tubulin (50kD) was used as protein loading control. (E) Edited hiPSC lines were assessed for expression of common pluripotency markers including OCT4, NANOG, SOX2 and SSEA-4. (F) Karyotyping was done via chromosome counting of DAPI-stained metaphases. D1171D n=3, D1171N ctrl n=1, N1171N n=2; values are displayed as mean±SD; one-way ANOVA followed by Tukey's multiple comparisons test (E and F); D1171D, turquoise; D1171N ctrl (unedited control), lightblue; N1171N, darkgrey.

hiPSCs were quadruple-positive for expression of the pluripotency markers OCT4, SOX2, NANOG and SSEA-4 comparable to the percentage previously observed in *RREB1* D1171D hiPSC lines (D1171D, 94.9±4.8%; D1171N ctrl, 98.3%; N1171N, 97.5±2.2%) (Figure 5.2 E).

Chromosome counting of DAPI-stained metaphases was performed by Dr Daniela Moralli at the Chromosome Dynamics Core (WCHG) to monitor genomic stability of *RREB1* hiPSC lines. Only two of three *RREB1* N1171N and the unedited D1171N ctrl hiPSC lines revealed a median count of 46 chromosomes per cell and no sign of aneuploidy. For the unedited *RREB1* D1171N ctrl hiPSC line 63% of the counted metaphases were characterised by 46 chromosomes/cell, while DAPI-stained chromosome counting of the two *RREB1* N1171N lines showed a mean of 63.5±0.7% diploid cells (Figure 5.2 F). Screening for coding mutations in *TP53*, commonly found in genome edited hPSC lines, revealed that neither *RREB1* D1171N ctrl nor N1171N hiPSC lines had acquired such a mutation (Table 5.1).

Table 5.1: QC overview of *RREB1* N1171N and D1171N hiPSC lines

	# Clones				
	Picked	**Correct genotype**	**Without *TP53* mutations**	**Pluripotent (OCT4, NANOG, SOX2 & SSEA-4)**	**Diploid (46 chr/cell)**
N1171N	192	3 (1.6%)	3	3	2 (1%)
D1171N ctrl	N/A	1	1	1	1

For QC overview of *RREB1* D1171D hiPSC clones see Table 3.2 (WT clones).

5.3.3 Characterisation of hiPSC *RREB1* D1171D, D1171N and N1171N lines differentiated along the endocrine lineage into beta-like cells

To investigate whether and how T2D risk and protective alleles at rs9379084 influence pancreatic beta cell development, two *RREB1* N1171N, three *RREB1* D1171N (unedited ctrl and parental SB Ad3.1 line in duplicate) and two *RREB1*

D1171D hiPSC lines were differentiated along the endocrine lineage in parallel, following the directed step-wise *in vitro* beta cell differentiation protocol described in Sections 1.4.4.1 and 2.4.

Unfortunately, *RREB1* D1171D line #2 showed a bacterial infection in hiPSC-derived EP cells and had to be discarded at stage 5. Assessment of cell morphology at distinct stages of beta cell development did not show an obvious difference in appearance of differentiated *RREB1* D1171D, D1171N and N1171N clones (Figure 5.3, data only shown for *RREB1* D1171D #1 and D1171N SB #1 lines).

5.3.3.1 Expression of stage-specific key developmental markers

RNA samples of *RREB1* lines were collected at all seven distinct stages of beta cell development and evaluated for expression of *RREB1* as well as stage-specific key developmental markers. In parallel, differentiated cells were assessed for their ability to form endoderm ($SOX17^+$) and characterised for co-expression of TFs essential in endocrine lineage specification at stages 4 and 5 or co-expression of NKX6.1 and C-peptide (surrogate marker for insulin) at stages 6 and 7 by FACS.

RREB1 expression was significantly higher in DE cells compared to hiPSCs and remained stably expressed throughout all stages of beta cell development for *RREB1* D1171N (unedited ctrl and parental SB) and *RREB1* N1171N clones (p=4.53x10^{-4} and p=0.0496, respectively). While the same was observed for *RREB1* D1171D line #2, *RREB1* transcript levels in D1171D clone #1 only started to rise at stage 5, equalling *RREB1* expression in *RREB1* D1171N and N1171N clones at stage 7 (BLC) (Figure 5.4).

Stages 1 & 2 - definitive endoderm and posterior foregut

Expression levels of *OCT4* were similar between *RREB1* D1171D, D1171N and N1171N lines measured in hiPSCs and PGT cells, respectively (Figure 5.5 A). While hiPSC lines expressed high levels of the pluripotency marker, expression was drastically reduced in the majority of DE cells and stayed low in PGT cells.

Figure 5.3: Cell morphology of differentiated *RREB1* D1171D and D1171N clones Morphology of *RREB1* D1171D #1 and D1171N SB #1 pancreatic endoderm (DE, S1D1 and S1D3), primitive gut tube (PGT, S2D2), posterior foregut (PFG, S3D2), pancreatic endoderm (PE, S4D3) endocrine precursors (EP, S5D3), endocrine-like (EN, S6D6) and beta-like cells (BLC, S7D6). S, stage; D, day.

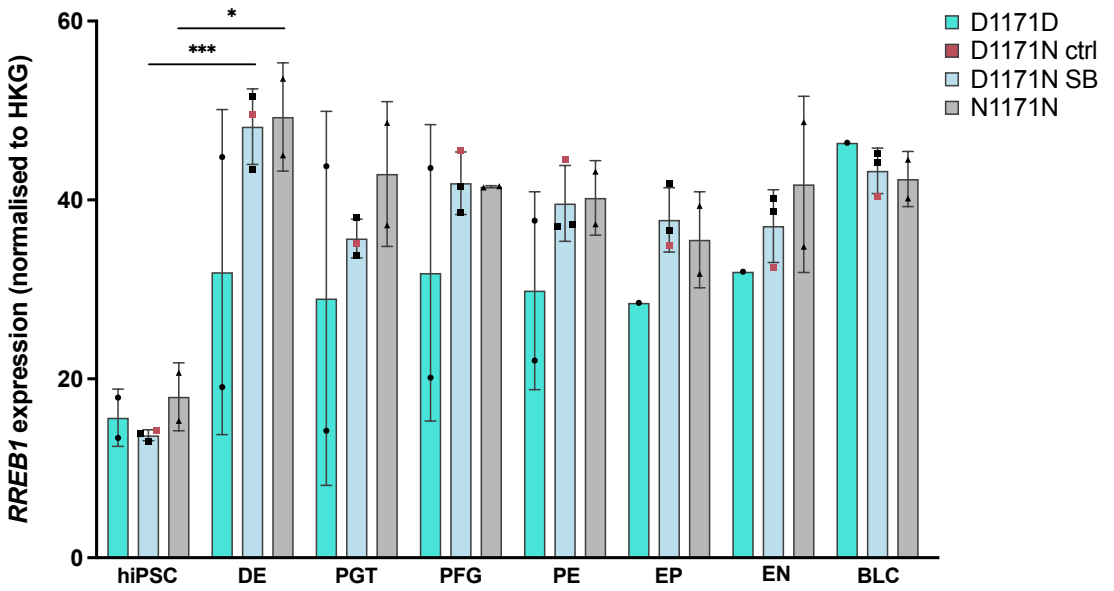


Figure 5.4: *RREB1* mRNA expression profiles during beta cell development in *RREB1* D1171D, D1171N and N1171N lines RREB1 was expressed throughout all seven distinct stages of beta cell development in *RREB1* D1171D, D1171N and N1171N lines. D1171D n=1-2, D1171N n=3, N1171N n=2; values are displayed as mean±SD; *p*-values *<0.05, ***<0.001 for unpaired t-test between hiPSC and DE; HKG, housekeeping genes; hiPSC, human induced pluripotent stem cells; DE, definitive endoderm; PGT, primitive gut tube; PFG, posterior foregut; PE, pancreatic endoderm; EP, endocrine precursors; EN, endocrine-like cells; BLC, beta-like cells; D1171D, turquoise; D1171N ctrl, red; D1171N SB, lightblue; N1171N, darkgrey.

RREB1 D1171D clone #1 and *RREB1* N1171N cell line #1, however, revealed less efficient DE formation, i.e. *OCT4* remained high and levels of *SOX17* at least for *RREB1* D1171D line #1 were considerably lower compared to those observed for the other *RREB1* cell lines (Figure 5.5 A and B). The same was reflected on the protein level. While for the majority of *RREB1* clones >70% of hiPSC-derived DE cells were positive for SOX17 and negative for the pluripotency marker SOX2, only 38.6% of *RREB1* D1171D #1 hiPSCs had formed DE (Figure 5.5 C).

Stages 3, 4 & 5 - posterior foregut, pancreatic endoderm and endocrine precursors

Expression of *PDX1* was first observed at stage 3 (PFG) and persisted until stage 7 (BLC), with the unedited *RREB1* D1171N ctrl clone displaying slightly higher transcript levels compared to the other *RREB1* lines, particularly at stage

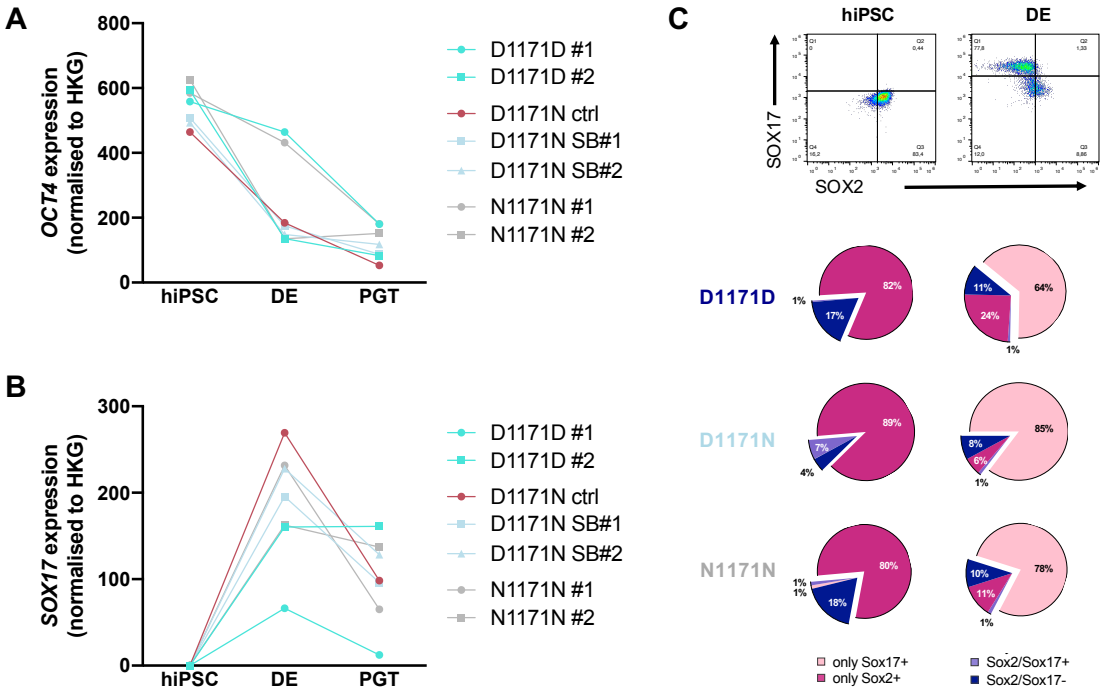


Figure 5.5: Stages 1 & 2 - Assessment of key developmental stage marker expression in differentiated *RREB1* D1171D, D1171N and N1171N cells Differentiated *RREB1* D1171D, D1171N and N1171N cells were evaluated for expression of *OCT4* (A) and *SOX17* (B). (C) Formation of endoderm was assessed by expression of SOX17. D1171D n=2, D1171N n=3, N1171N n=2; values are displayed as mean±SD; one-way ANOVA followed by Tukey's multiple comparisons test; hiPSC, human induced pluripotent stem cells; DE, definitive endoderm; PGT, primitive gut tube; D1171D, turquoise; D1171N ctrl, red; D1171N SB, lightblue; N1171N, darkgrey.

4 (PE) (Figure 5.6 A). *PDX1* expression was proceeded by *NKX6.1*. Again the unedited *RREB1* D1171N ctrl clone was characterised by slightly elevated *NKX6.1* transcript levels compared to the remaining *RREB1* clones. At stage 4 *NKX6.1* expression was only detectable in the *RREB1* D1171N ctrl clone and *RREB1* N1171N line #1 (Figure 5.6 B). In line with this, the percentage of PE cells positive for NKX6.1 protein was below 10% across all *RREB1* lines and the proportion of cells double-positive for PDX1 and NKX6.1 was even lower (0.76±0.20% for D1171D, 2.87±3.44% for D1171N and 4.58±5.47% for N1171N clones) (Figure 5.6 C). While *RREB1* D1171N and N1171N lines revealed on average 32.87±21.74% and 31.90±17.97% PDX1-positive PE cells, respectively, only 5.85±2.19% *RREB1* D1171D PE cells expressed PDX1.

The endocrine progenitor marker *NEUROG3* was transiently expressed in PE

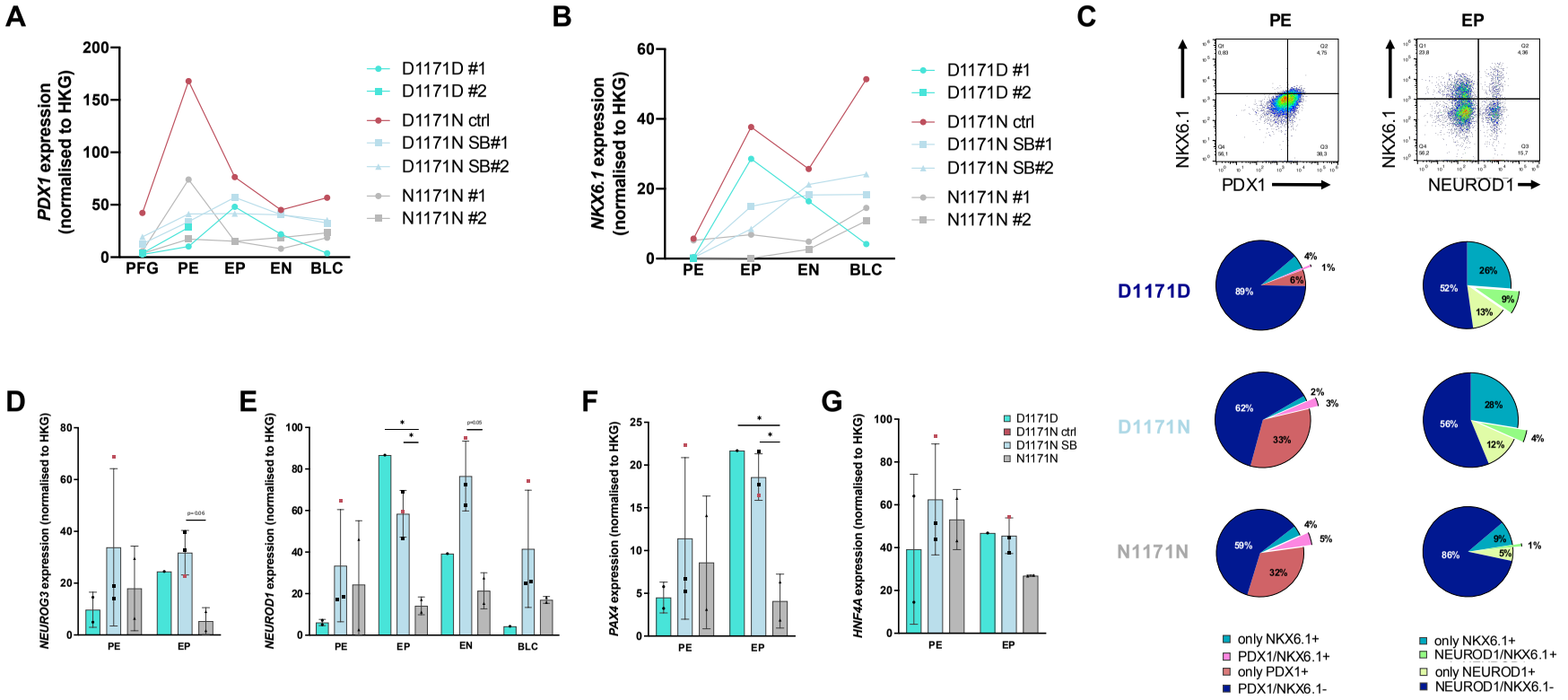


Figure 5.6: Stages 3, 4 & 5 - Assessment of key developmental stage marker expression in differentiated *RREB1* D1171D, D1171N and N1171N cells Differentiated cells were evaluated for the expression of genes encoding the TFs *PDX1* (A), *NKX6.1* (B), *NEUROG3* (D), *NEUROD1* (E), *PAX4* (F) and *HNF4A* (G). (C) Differentiated cells were evaluated for co-expression of PDX1 and NKX6.1 at stage 4 as well as NKX6.1 and NEUROD1 at stage 5 via FACS. D1171D n=1-2, D1171N n=3, N1171N n=2; values are displayed as mean±SD; p-values $^{*}<0.05$ for one-way ANOVA followed by Tukey's multiple comparisons test; PFG, posterior foregut; PE, pancreatic endoderm; EP, endocrine precursors; EN, endocrine-like cells; BLC, beta-like cells; D1171D, turquoise; D1171N ctrl, red; D1171N SB, lightblue; N1171N, darkgrey.

and EP cells. Expression levels in *RREB1* N1171N EP cells were decreased in comparison to those observed in *RREB1* D1171D and D1171N clones (p=0.2115 and p=0.0531, respectively) (Figure 5.6 D). In line with this, NEUROG3 target genes *NEUROD1* and *PAX4* were significantly lower expressed in *RREB1* N1171N clones compared to *RREB1* D1171D and D1171N clones at stage 5 (EP) (p<0.05) (Figure 5.6 E and F). Reduced NEUROD1 was also observed at the protein level in *RREB1* N1171N lines. Less than 8% of EP cells were positive for NEUROD1, and less than 2% were double-positive for NEUROD1 and NKX6.1 (Figure 5.6 C). The *RREB1* D1171D clone was characterised by ∼21% NEUROD1 positive cells and ∼9% of EP cells co-expressed NKX6.1, whilst *RREB1* D1171N lines revealed between 12 and 20% NEUROD1-expressing cells, of which ∼16-33% co-expressed NKX6.1 (Figure 5.6 C).

HNF4A showed variable expression levels for the two *RREB1* D1171D clones at stage 4 with D1171D clone #2 displaying transcript levels similar to the remaining *RREB1* clones, while D1171D clone #1 was characterised by markedly reduced expression levels. *RREB1* N1171N hiPSC-derived EP cells showed decreased, albeit not significant, *HNF4A* expression compared to *RREB1* D1171D and D1171N clones at stage 5 (p=0.1851 and p=0.1128, respectively) (Figure 5.6 G).

Stages 6 & 7 - endocrine-like and beta-like cells

INS and *GCG* expression tended to be lower in *RREB1* N1171N lines compared to *RREB1* D1171D and D1171N clones at stage 6 (EN) (Figure 5.7 A and B). FACS analysis revealed that less than 1% of *RREB1* N1171N EN cells co-expressed NKX6.1 and C-peptide, while ∼3% D1171D and between ∼2 and 4% D1171N cells were positive for NKX6.1 and C-peptide at stage 6. After six additional days of differentiation the percentage of BLCs co-expressing NKX6.1 and C-peptide was very variable between individual *RREB1* clones. While the unedited *RREB1* D1171N ctrl clone showed the highest percentage of double-positive cells (∼12%), *RREB1* D1171D clone #1 and *RREB1* N1171N clone #1 BLCs contained ∼3-4% double-positive cells, and the remaining *RREB1* D1171N and N1171N clones

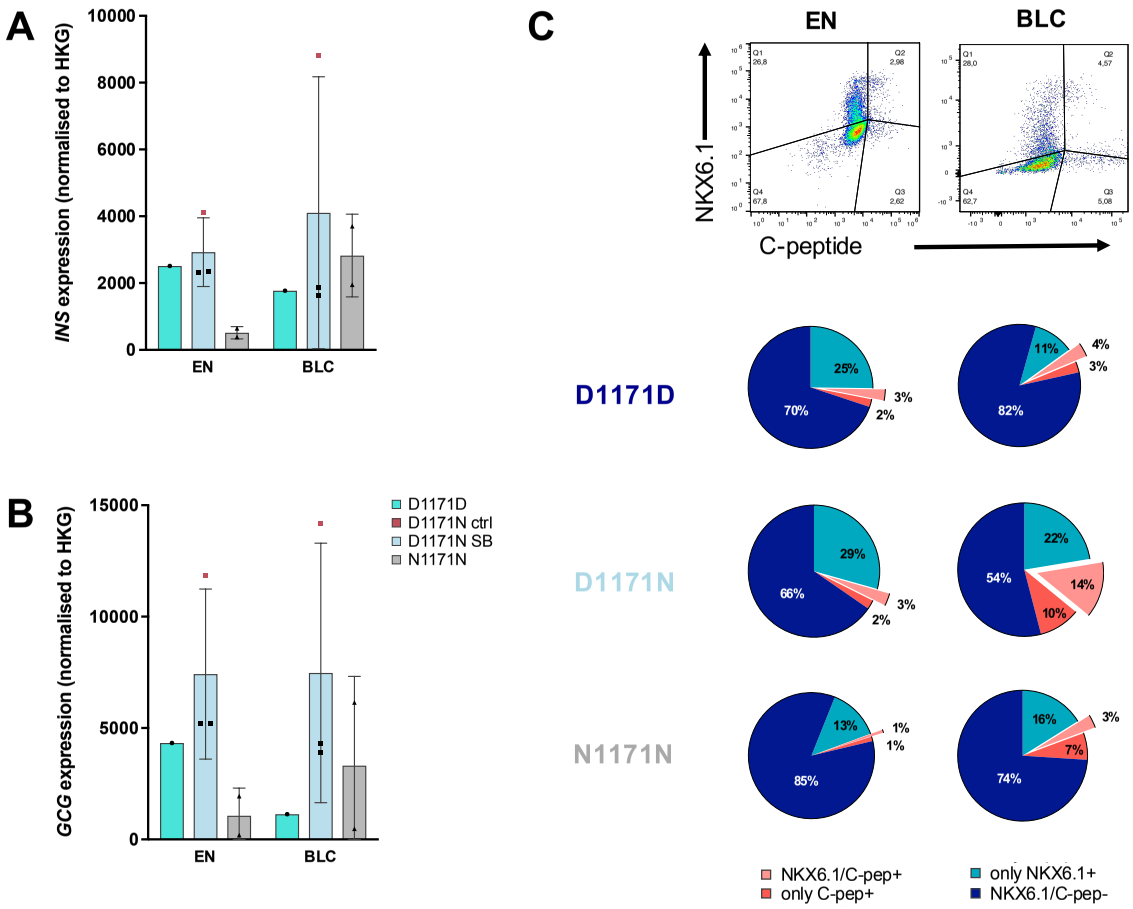


Figure 5.7: Stages 6 & 7 - Assessment of key developmental stage marker expression in differentiated *RREB1* D1171D, D1171N and N1171N cells Endocrine-like and beta-like cells were evaluated for expression of the hormones *INS* (A) and *GCG* (B). (C) Proportions of cells positive for NKX6.1 and C-peptide, just NKX6.1, just C-peptide or negative for both at stages 6 and 7 assessed by FACS. D1171D n=1, D1171N n=3, N1171N n=2; values are displayed as mean±SD; one-way ANOVA followed by Tukey's multiple comparisons test; EN, endocrine-like cells; BLC, beta-like cells; GCG, glucagon; D1171D, turquoise; D1171N ctrl, red; D1171N SB, lightblue; N1171N, darkgrey.

were characterised by only ∼1-2% cells co-expressing NKX6.1 and C-peptide (Figure 5.7 C).

Taken together, despite, sometimes high variability in stage-specific TF expression, characterisation of differentiated *RREB1* D1171D, D1171N and N1171N cells revealed that all lines followed the beta cell lineage programme and generated BLCs positive for both NKX6.1 and C-peptide (insulin) to varying degrees. Although, the unedited *RREB1* D1171N ctrl clone tended to show higher expression levels compared to the parental *RREB1* D1171N SB clones, for a subset of TFs (in particular at stage 4), it did not drive the significant differences observed between *RREB1* N1171N and the remaining *RREB1* lines, thus differences were likely attributable to a genotype effect. *RREB1* N1171N cells were characterised by lower levels of endocrine precursor markers including *NEUROG3* and its target genes *NEUROD1* as well as *PAX4* at stage 5, suggesting that the *RREB1* T2D protective allele might have had a negative impact on endocrine differentiation. As loss of RREB1 had resulted in increased expression of endocrine progenitor markers at stages 4 and 5, hence revealed a positive impact on endocrine differentiation, it seems that c.3511A at rs9379084 acted as a GOF rather than a LOF mutation during beta cell development.

5.3.4 Transcriptome profiling of *RREB1* D1171D, D1171N and N1171N hiPSC lines differentiated along the endocrine lineage

Having observed a potential effect of the *RREB1* T2D protective allele on endocrine differentiation, I sought to investigate, whether the T2D protective allele might have had an impact on transcriptional networks during beta cell development. Therefore, gene expression profiling was performed in *RREB1* D1171D, D1171N and N1171N lines at seven distinct stages of beta cell differentiation. A total of 46 RNA-Seq libraries were sequenced to a mean read depth of 35.9±3.6 million reads per sample.

For *RREB1* D1171D lines, transcriptome data was only available for one cell clone for stages 5 to 7, thus, to increase statistical power, gene expression data of *RREB1* WT clones #1 and #2 discussed in Section 3.3.4 and referred to as D1171D #3 and #4 in the remaining part of this chapter, were included in the RNA-Seq analysis. Correction for batch effects and normalisation of different sequencing depths was done using the RUVSeq package in R (see Section 2.5.2.1).

Boxplots of unnormalised RLE revealed large distributional differences between library preparations. Additionally, the first two principal components showed that libraries did not cluster according to their developmental stage, in particular, D1171D #3 and #4 (Figure 5.8 A). Library preparation effects were attenuated by upper-quartile normalisation, i.e. both PCA clustering and RLE boxplot clustering around zero improved. However, variability observed for the library preparation of some samples, e.g. D1171D #3 DE (first green boxplot), could not be fully removed as exemplified by the median still not being equal to zero (Figure 5.8 B). Application of RUVs shrank the expression measures for the D1171D #3 DE library more towards the median across all libraries, thereby fully adjusting for library preparation effects, and improved sample separation between developmental stages (Figure 5.8 C). Thus, further analysis was performed on RUVs normalised counts.

5.3.4.1 Differential gene expression analysis

Genetic model analysis detected significant association between *RREB1* rs9379084 and T2D risk in both an additive and a dominant inheritance model, but not in a recessive model (Anubha Mahajan, personal communication). While in an additive model homozygous carriers of a risk allele have double the disease risk compared to heterozygous risk allele carriers, in a dominant inheritance model disease risk is the same for carriers of one risk allele (heterozygous) as for individuals carrying two risk allele copies (homozygous) (Salanti et al. 2009). Concerning *RREB1* rs9379084, an additive model would assume that individuals homozygous for the p.D1171 allele (*RREB1* D1171D) are at highest risk, heterozygous carriers (*RREB1*

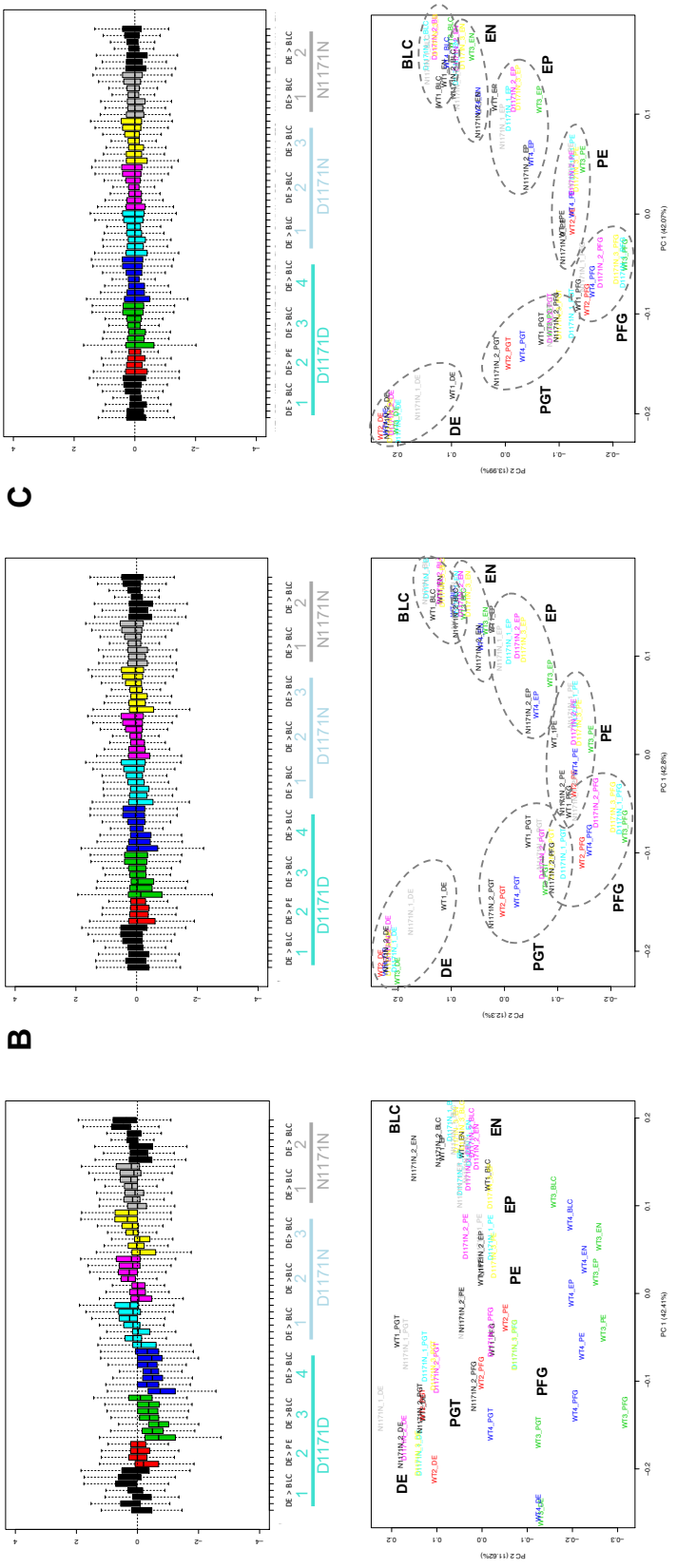


Figure 5.8: Removal of unwanted variation RLE (boxplots, top lane) and PCA (scatter plots of first two principal components, bottom lane) of unnormalised counts (A), counts after upper-quartile normalisation (B) and counts after RUVs (C). D1171D n=3-4, D1171N n=3, N1171N n=2; RLE, relative log expression; PCA, principal component analysis (log scale, centred); boxplots: bottom and top of box represent the 1st and 3rd quartiles, respectively, while the black line inside indicates the median; whiskers show at 1.5 of the interquartile range (IQR) above and below the box; DE, definitive endoderm; PGT, primitive gut tube; PFG, posterior foregut; PE, pancreatic endoderm; EP, endocrine precursors; EN, endocrine-like cells; BLC, beta-like cells.

D1171N) at medium risk and individuals homozygous for the T2D protective allele p.N1171 (*RREB1* N1171N) at lowest risk for developing T2D. In contrast, in a dominant model individuals heterozygous or homozygous for the risk allele p.D1171 are exposed to the same T2D risk, while individuals homozygous for the T2D protective allele p.N1171 are at lower risk of developing T2D.

As genetic association supported both an additive and a dominant model, assessment of DEGs was performed assuming either of the two models separately. Differential gene expression analysis based on rs9379084 p.N1171 allele number (additive model) only detected a notable amount of DEGs at stage 3 (PFG) between *RREB1* D1171D, D1171N and N1171N clones (Table 5.2). A total of 643 DEGs were identified, of which 174 (∼27%) showed increased and 469 (∼73%) displayed decreased expression with increasing p.N1171 allele number. Down-regulated DEGs were associated with biological processes relating to 'cell division', 'cell cycle' and 'microtubule-based process' (Table 5.3).

Analysis assuming a dominant inheritance model revealed ∼4x more DEGs at stage 3 (2,696 genes) compared to the additive model as well as a few at stages 5 and 6 (Table 5.2). The majority of these were down-regulated in *RREB1* N1171N lines. While DEGs up-regulated in *RREB1* N1171N clones at stage 3 were enriched for biological terms related to the ER ('protein localisation to endoplasmic reticulum' or 'protein processing in endoplasmic reticulum'), down-regulated DEGs contained genes associated with gene expression regulation and cilia ('cilium assembly' and 'cilium organisation'). Up-regulated DEGs at stage 5 (EP) were enriched for genes implicated in 'tissue development', while down-regulated DEGs were associated with synaptic processes, reflected in an enrichment of biological terms including 'chemical synaptic transmission', 'synaptic signalling' and 'potassium channels'.

5.3.4.2 Prediction of upstream transcriptional regulators of DEGs

TF enrichment analysis using iRegulon predicted that DEGs up-regulated in *RREB1* N1171N compared to D1171D and D11171N clones (dominant model) at stage 3

Table 5.2: Differentially expressed genes between *RREB1* D1171D, D1171N and N1171N lines at seven stages of beta cell development

	DE	PGT	PFG	PE	EP	EN	BLC
Additive model							
DEGs	1	6	643	4	4	2	2
Up-regulated			174 (27%)				
Predicted RREB1 target genes			114 (66%)				
Down-regulated			469 (73%)				
Predicted RREB1 target genes			206 (44%)				
Dominant model							
DEGs	0	1	2,696	0	284	14	0
Up-regulated			1,206 (45%)		79 (28%)	9 (64%)	
Predicted RREB1 target genes			783 (65%)		34 (43%)	2 (22%)	
Down-regulated			1,490 (55%)		205 (72%)	5 (36%)	
Predicted RREB1 target genes			659 (44%)		76 (37%)	1 (20%)	

57,820 genes tested using DESeq2; genes differentially expressed at q<0.01 for p.N1171 allele additive and dominant models; DEGs, differentially expressed genes; DE, definitive endoderm; PGT, primitive gut tube; PFG, posterior foregut; PE, pancreatic endoderm; EP, endocrine precursors; EN, endocrine-like cells; BLC, beta-like cells.

were enriched for TFs of the ETS family (NES=3.0-5.7), which has been implicated in the regulation of insulin secretion (Gutierrez-Aguilar et al. 2014; Luo et al. 2014; F Chen et al. 2016). Additionally, at stage 5, a subset of DEGs down-regulated in *RREB1* N1171N were associated with the TF REST (NES=3.2-7.9). DEGs identified assuming an additive model were not enriched for a particular TF family.

Motif activity response analysis (MARA) identified differential transcriptional activity for the RFX2/3 (Z=6.24) and REST (Z=5.03) motifs. RFX2/3 motif activity in *RREB1* N1171N lines was significantly decreased during stages 3 to 6 compared to *RREB1* D1171N clones (p<0.01) as well as at stages 5 and 6 compared to *RREB1* D1171D clones (p<0.05) (Figure 5.9 A). In line with this, expression of *RFX2* and *RFX3* seemed to be slightly, albeit not significantly lower in *RREB1* N1171N clones at stages 5 to 7 for *RFX2* and stages 3 to 7 for *RFX3* compared to *RREB1* D1171D and D1171N clones (Figure 5.9 B and C). Transcriptional activity

Table 5.3: Enriched biological terms and pathways among DEGs between *RREB1* D1171D, D1171N and N1171N lines at distinct stages of beta cell development

Stage		# enriched terms	Examples of enriched terms	padj (q)
Additive model				
PFG (down-regulated DEGs)	GO	11	cell division	2.13×10^{-6}
			cell cycle	0.0010
			microtubule-based process	0.0018
Dominant model				
PFG (up-regulated DEGs)	GO	119	protein localisation to endoplasmic reticulum	6.49×10^{-24}
			protein targeting to ER	6.41×10^{-23}
			protein folding	3.87×10^{-4}
	REACTOME	49	translation	7.92×10^{-32}
			rRNA processing	1.28×10^{-24}
	KEGG	9	ribosome	6.82×10^{-32}
			protein processing in endoplasmic reticulum	0.0061
PFG (down-regulated DEGs)	GO	83	regulation of cellular macromolecule biosynthetic process	2.77×10^{-31}
			regulation of gene expression	1.22×10^{-22}
			cilium organisation	0.0078
	REACTOME	7	cilium assembly	2.53×10^{-5}
			RNA polymerase II transcription	8.46×10^{-18}
EP (up-regulated DEGs)	GO	7	tissue development	3.38×10^{-5}
			developmental process	7.96×10^{-4}
EP (down-regulated DEGs)	GO	15	chemical synaptic transmission	1.36×10^{-7}
			synaptic signalling	3.17×10^{-7}
			synaptic vesicle cycle	0.0080
	REACTOME	2	potassium channels	1.53×10^{-4}
			neuronal system	0.0015

15-1,420 DEGs were tested for enriched biological terms and pathways using g:profiler (see Section 2.5.3). All human genes annotated in the Ensembl database were used as background. Significance threshold was set to q<0.01 using the tailor-made g:SCS algorithm for multiple testing. DEGs, differentially expressed genes; PFG, posterior foregut; EP, endocrine precursors.

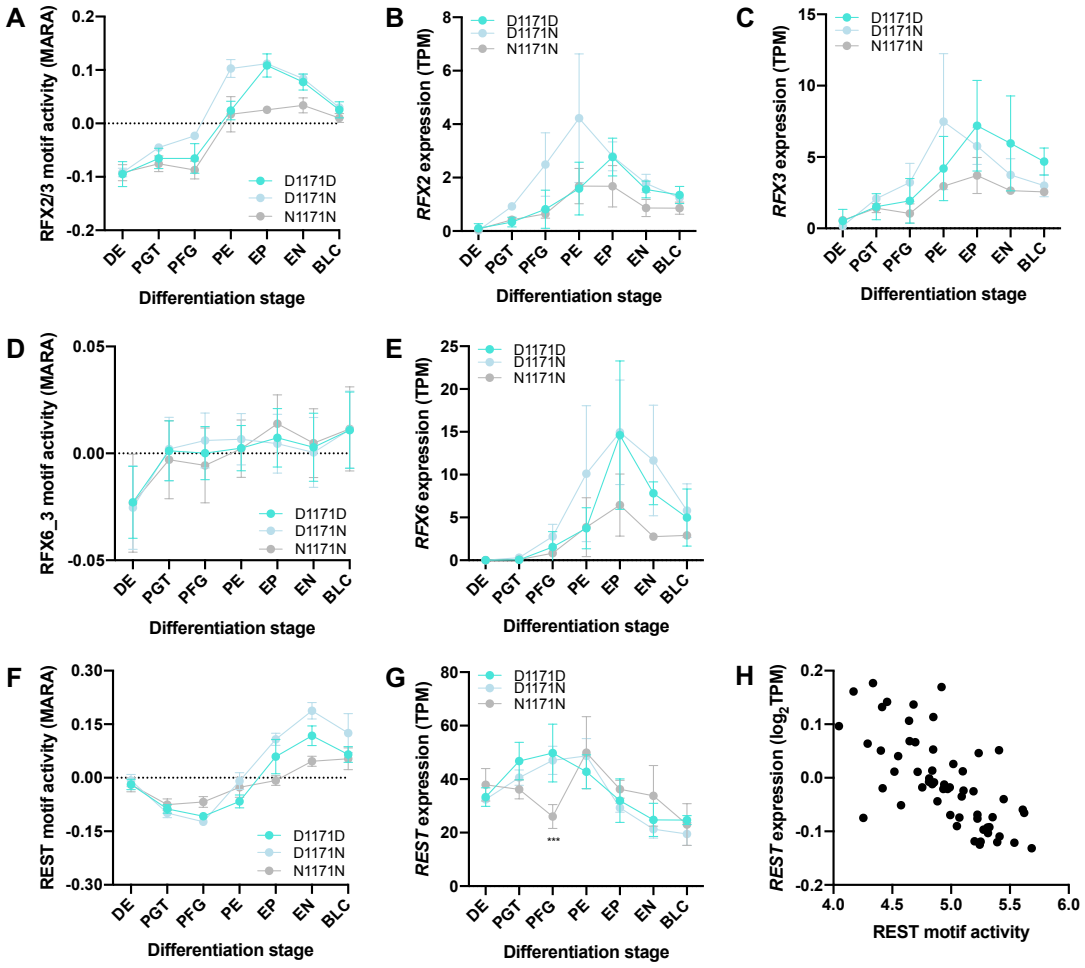


Figure 5.9: RFX2/3, RFX6 and REST motif activity profiles in *RREB1* D1171D, D1171N and N1171N lines across beta cell development (A) RFX2/3 motif activity profiles for *RREB1* D1171D, D1171N and N1171N lines during beta cell development calculated using MARA. *RFX2* (B) and *RFX3* (C) expression profiles. (D) RFX6 transcriptional activity did not differ between *RREB1* D1171D, D1171N and N1171N lines. (E) *RFX6* expression profiles during beta cell development. (F) REST motif activity patterns for *RREB1* lines during beta cell development. (G) *REST* expression was significantly decreased in *RREB1* N1171N clones at stage 3 (PFG). (H) *REST* expression negatively correlated with REST motif activity. D1171D n=3-4, D1171N n=3, N1171N n=2; values are displayed as mean±SD; *p*-values ***<0.001 for two-way ANOVA followed by Tukey's multiple comparisons test (A, D and F) or Wald test followed by Benjamini-Hochberg adjustment (dominant model) (B, C, E and G); Pearson correlation (H); TPM, transcripts per million; D1171D, turquoise; D1171N, lightblue; N1171N, darkgrey.

of RFX6 was similar in *RREB1* D1171D, D1171N and N1171N lines and *RFX6* transcript levels tended to be slightly lower in *RREB1* N1171N clones for stages 5 to 7 compared to the other *RREB1* lines (Figure 5.9 D and E).

REST motif activity was significantly lower in *RREB1* N1171N than D1171N clones at stages 5, 6 and 7 (p<0.05) as well as compared to D1171D clones at stages 5 and 6 (p<0.05), suggesting REST was more repressive in these clones (Figure 5.9 F). *REST* expression was significantly down-regulated in *RREB1* N1171N compared to D1171D and D1171N clones at stage 3 (dominant model, q=3.16x10^{-4}) (Figure 5.9 G), and negatively correlated with REST motif activity (r=-0.7525, p=4.15x10^{-12}), confirming REST as transcriptional repressor (Figure 5.9 H).

In Chapter 3, MARA and differential gene expression analysis had revealed higher motif activity and expression for RFX2 and RFX3 across all seven stages of beta cell development as well as increased REST motif activity for stages 4 to 7 in *RREB1* KO clones (LOF) (see Sections 3.3.4.8 and 3.3.4.10). *RREB1* N1171N lines had shown inverse profiles, i.e. RFX2/3 and REST motif activities were markedly lower at stages 5 and 6 compared to *RREB1* D1171D and D1171N lines. Given this reciprocal behaviour compared to *RREB1* KO lines, the analysis suggests that p.N1171 (c.3511A) was acting as a GOF mutation.

5.3.4.3 The *RREB1* T2D protective allele at rs9379084 likely represents a gain-of-function allele

Assessment of gene expression profiles between *RREB1* D1171D, D1171N and N1171N lines further supported that the T2D protective allele acted as a GOF allele. Genes that were up-regulated in *RREB1* KO compared to D1171D lines, tended to show lower expression in *RREB1* N1171N lines across beta cell development when compared to expression profiles of *RREB1* D1171D and D1171N lines (Figure 5.10). These included genes encoding TFs such as *NKX6.1*, *NEUROG3*, *NKX2.2* and *NEUROD1* (Figure 5.10 A-D) as well as direct RREB1 target genes involved in cell-cell communication and adhesion such as *GJD2* and *NCAM1* (Figure 5.10 E and F), RREB1 *cis*-regulated genes participating in insulin secretion *SNAP25*

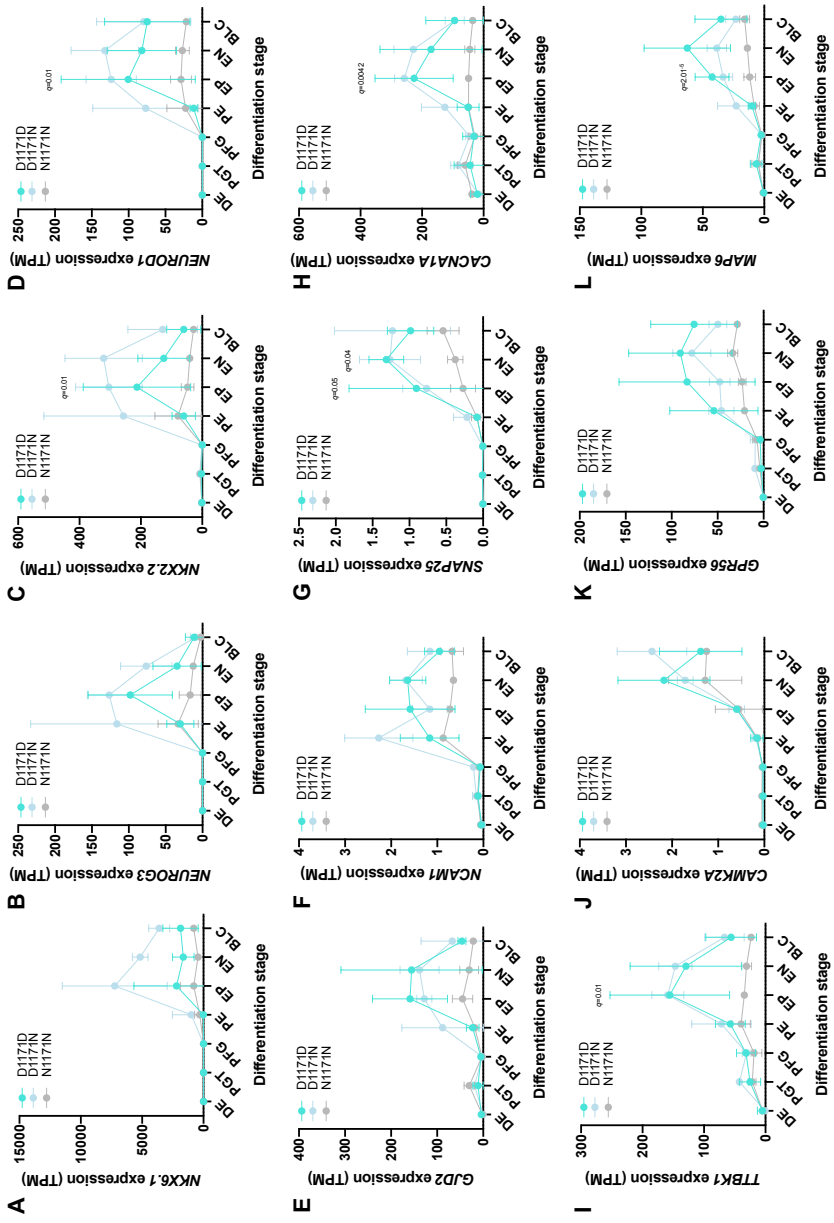


Figure 5.10: Allelic effects of the *RREB1* T2D protective allele at rs9379084 on expression profiles of selected genes
Expression profiles of genes encoding TFs *NKX6.1* (A), *NEUROG3* (B), *NKX2.2* (C) and *NEUROD1* (D) in *RREB1* D1171D, D1171N and N1171N lines. Expression of direct RREB1 target genes *GJD2* (E), *NCAM1* (F), *SNAP25* (G), *CACNA1A* (H) and *TTBK1* (I) in *RREB1* D1171D, D1171N and N1171N lines. Expression of RFX2/3 target genes *CAMK2A* (J), *GPR56* (K) and *MAP6* (L). D1171D n=3-4, D1171N n=3, N1171N n=2; values are displayed as mean±SD; *p*-values for Wald test followed by Benjamini-Hochberg adjustment (dominant model); DE, definitive endoderm; PGT, primitive gut tube; PFG, posterior foregut; PE, pancreatic endoderm; EP, endocrine precursors; EN, endocrine-like cells; BLC, beta-like cells; TPM, transcripts per million; D1171D, turquoise; D1171N, lightblue; N1171N, darkgrey.

and *CACNA1A* (Figure 5.10 G and H) and target genes of yet unknown function in beta cell development like *TTBK1* (Figure 5.10 I). Moreover, RFX2/3 target genes *CAMK2A*, *GPR56* and *MAP6*, significantly up-regulated in *RREB1* KO lines, tended to be lower expressed in *RREB1* N1171N lines (Figure 5.10 J-L).

5.3.4.4 Weighted gene co-expression network analysis revealed that DEGs were enriched in modules of co-expressed genes

To assess whether the *RREB1* T2D protective allele affected transcriptional networks associated with beta cell development, WGCNA was performed on 17,189 protein-coding genes and lincRNAs. In total, 18 co-expressed modules were identified between differentiated *RREB1* D1171D, D1171N and N1171N lines, containing between 73 and 4,639 genes. Evaluation of module eigengene (ME) expression revealed that stage 3 (PFG) contained five modules, which were significantly different between *RREB1* N1171N and D1171N and four modules which were markedly different between *RREB1* N1171N and D1171D clones (Table 5.4), in line with differential gene expression analysis which had identified a notable number of DEGs at stage 3.

Three of the differentially expressed ME modules were enriched for DEGs up-regulated in *RREB1* N1171N compared to D1171N and D1171D PFG cells (dominant model). These included the tan (∼65% DEGs, q=2.89x10^{-76}) (Figure 5.11 A), the greenyellow (∼17% DEGs, q=3.23x10^{-8}) (Figure 5.11 B) and the salmon (∼14% DEGs, q=2.05x10^{-4}) (Figure 5.11 C) modules. Genes in the tan module were associated with gene expression processes, reflected in an enrichment of biological terms including 'RNA processing' (p=2.81x10^{-10}) and 'translation' (p=0.0013), in line with findings from previous DEG enrichment analysis (see Table 5.3). Down-regulated DEGs were enriched in the yellow model (∼9% DEGs, q=5.49x10^{-10}) (Figure 5.11 D), which contained genes associated with biological terms related to the nervous system ('generation of neurons' (p=6.68x10^{-4}), 'neurogenesis' (p=0.0021), 'regulation of nervous system development' (p=0.0027)).

Table 5.4: Differentially expressed module eigengenes identified in WGCNA for *RREB1* D1171D, D1171N and N1171N lines at stage 3 of beta cell development

MEs	# genes	PFG (padj)
black	207	
blue	4,085	
brown	875	
cyan	98	D1171D: q=0.0024
green	511	
greenyellow	154	D1171N: q=4.45x10^{-9}, D1171D: q=9.53x10^{-7}
grey	4,481	
grey60	73	
lightcyan	82	D1171N: q=0.0237
magenta	186	
midnightblue	90	
pink	205	
purple	185	
red	252	
salmon	113	D1171N: q=6.23x10^{-4}, D1171D: q=0.0050
tan	123	D1171N: q=1.48x10^{-7}, D1171D: q=1.78x10^{-6}
turquoise	4,639	
yellow	830	D1171N: q=0.0041

17,189 protein-coding genes and lincRNAs tested; two-way ANOVA followed by Tukey's multiple comparisons test (p-values shown for N1171N vs D1171N or D1171D); MEs, module eigengenes; PFG, posterior foregut.

Pancreatic progenitor and endocrine (progenitor) marker genes were enriched in the blue or turquoise modules, respectively. None of these modules were differentially expressed between *RREB1* N1171N and D1171N or D1171D lines.

5.3.5 Effects of the rs9379084 T2D protective variant on primary islet function

Having observed a potential effect of the *RREB1* T2D protective allele at rs9379084 on beta cell development, I next sought to investigate, whether it affects human

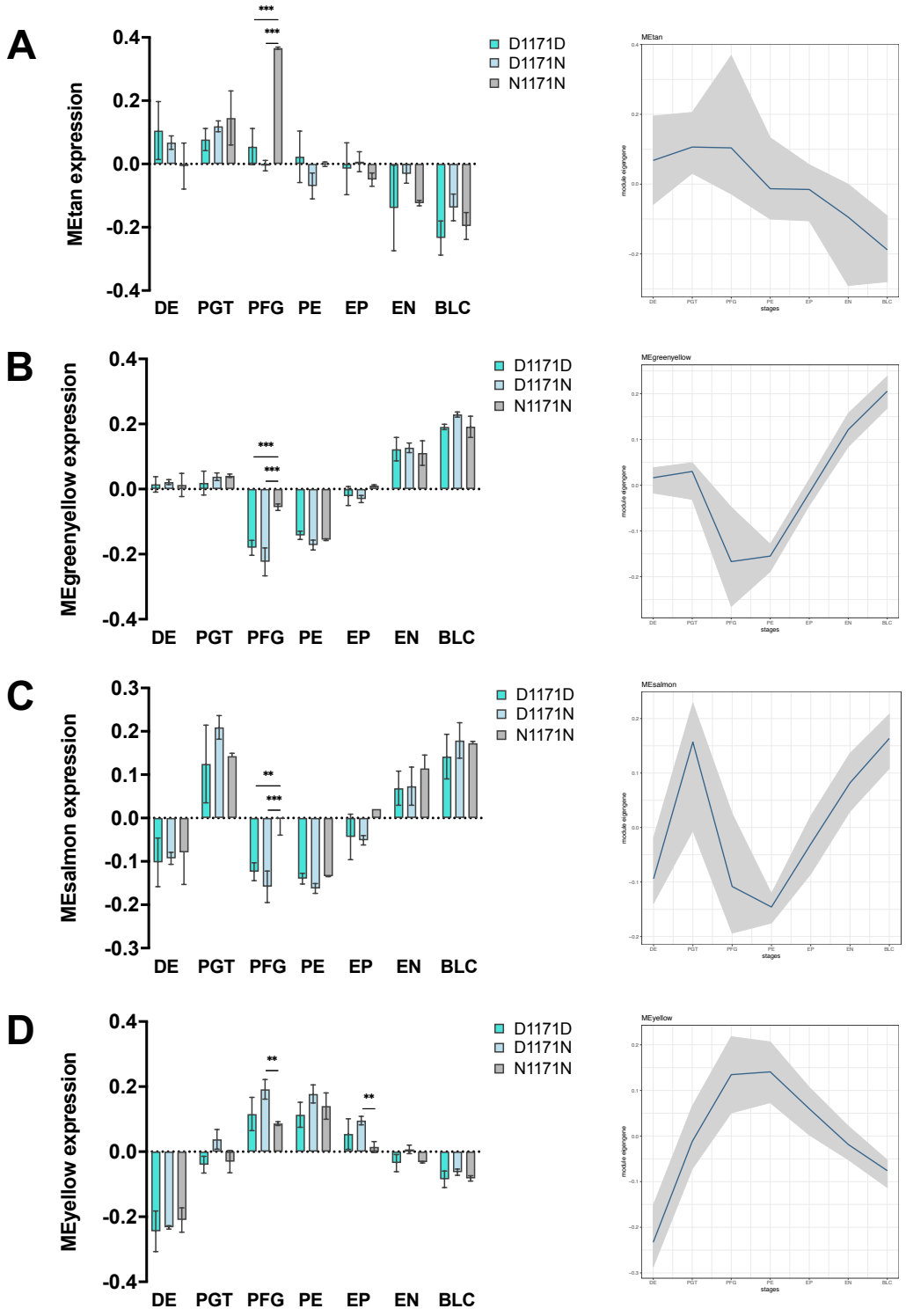


Figure 5.11: Analysis of modules of co-expressed genes using WGCNA Bar plots (left) and ribbon plots (right) showing ME expression of selected modules in *RREB1* D1171D, D1171N and N1171N lines throughout beta cell development. (A) MEtan, (B) MEgreenyellow, (C) MEsalmon, (D) MEyellow. D1171D n=3-4, D1171N n=3, N1171N n=2; values are displayed as mean±SD ; p-values $^{**}<0.01$, $^{***}<0.001$ for two-way ANOVA followed by Tukey's multiple comparisons test; D1171D, turquoise; D1171N, lightblue; N1171N, darkgrey.

beta cell function.

For this purpose, insulin secretion and content measurements of intact human islets isolated from cadaveric donors were analysed. In the Alberta Diabetes Institute IsletCore database 14 individuals heterozygous for the T2D protective allele (D1171N) at rs9379084 and 14 sex-, age- (p=0.7689), BMI- (p=0.9860) and CIT- (p=0.8441) matched donors homozygous for the T2D risk allele (D1171D) were identified (Table 5.5). Islets of donors were analysed for insulin content as well as basal (1mM glucose) and high (16.7mM glucose) insulin secretion (sequential incubations). There were no significant differences in insulin content (p=0.4680) (Figure 5.12 A) or insulin secretion (basal, p=0.6582; high, p=0.7556) (Figure 5.12 B) between islets from individuals heterozygous for the T2D protective allele and non-carriers.

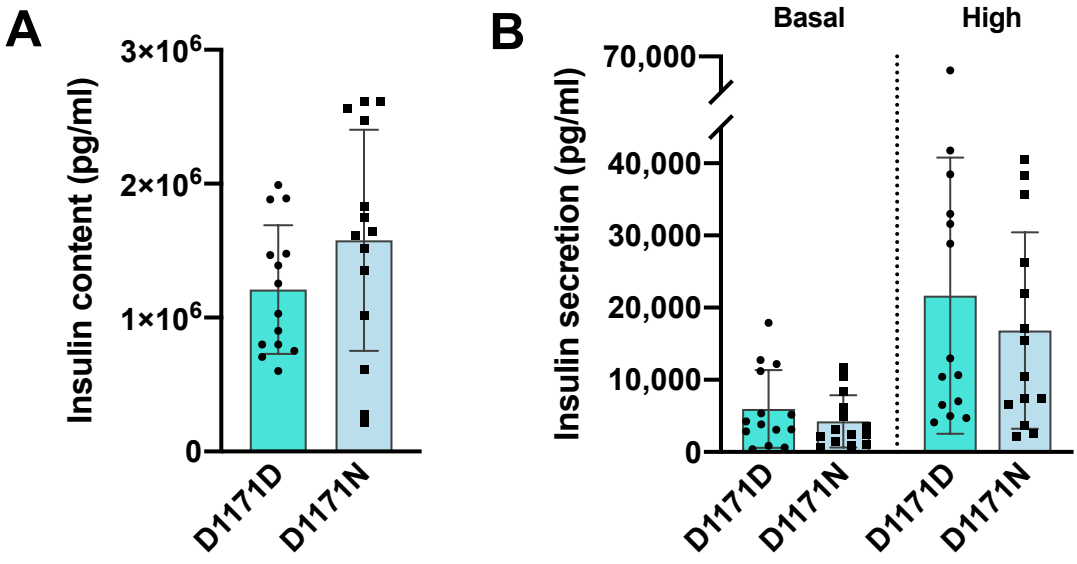


Figure 5.12: Effects of the rs9379084 T2D-associated variant on insulin content and secretion measurements from intact human islets Primary human islets isolated from donors heterozygous for the rs9379084 T2D protective allele (D1171N) and matched individuals homozygous for the T2D risk allele (D1171D) were analysed for total insulin content (A) and insulin secretion (B) under basal (1mM) and high (16.7mM) glucose conditions. n=14; values are displayed as mean±SD; paired t-test; D1171D, turquoise; D1171N, lightblue.

Table 5.5: Islet donor details and isolation characteristics

	Age (years)	BMI (kg/m²)	Sex	CIT (hours)		Age (years)	BMI (kg/m²)	Sex	CIT (hours)
D1171N	38	24.1	Male	10.7	**D1171D**	46	24.5	Male	16.0
	55	23.6	Male	13.5		52	23.8	Male	17.3
	80	19.7	Male	14.5		73	23.7	Male	19.0
	65	27.1	Male	17.5		62	26	Male	17.5
	53	29.6	Male	19.0		51	29.4	Male	16.6
	58	33.0	Male	15.0		60	31.9	Male	12.5
	54	24.7	Female	15.0		55	24.1	Female	13.5
	65	24.9	Female	17.0		59	24.8	Female	17.2
	62	19.0	Female	14.0		66	18.5	Female	13.8
	76	19.3	Female	10.5		71	19.7	Female	11.6
	43	21.3	Female	11.5		43	24.6	Female	2.2
	59	35.1	Female	19.3		66	33.1	Female	20.0
	54	32.9	Female	3.5		54	30.6	Female	10.1
	56	32.0	Female	13.8		50	31.7	Female	16.3

Details of 14 islet donors, carrying one (D1171N) or two (D1171D) T2D risk alleles from the Alberta Diabetes Institute IsletCore database. BMI, body mass index; CIT, cold ischaemia time.

5.4 Discussion

Coding variant fine-mapping has identified the T2D-associated p.D1171N variant in *RREB1* as causal and established *RREB1* as effector transcript. Neither functional effect of the variant nor molecular mechanisms underlying T2D association are known. RREB1 expression levels were not affected by the T2D protective allele, implicating that p.N1171 exerted neither a deleterious nor a beneficial effect on RREB1 protein.

The *RREB1* T2D protective allele p.N1171 likely results in a GOF

To interrogate the impact of individual T2D-associated *RREB1* alleles (p.D1171D, p.D1171N and p.N1171N) on beta cell differentiation, a developmental model system based on CRISPR/Cas9 genome-engineered hiPSCs was generated. Differentiation along the endocrine lineage into BLCs and evaluation of stage-specific key developmental TF expression revealed reduced levels of endocrine progenitor markers *NEUROG3*, *NEUROD1* and *PAX4* in *RREB1* N1171N lines at stage 5 (EP), suggesting a negative impact of the T2D protective allele on beta cell development. This was slightly unexpected as the p.N1171 allele is associated with reduced T2D susceptibility and lower FG levels, suggesting a protective effect on T2D risk for individuals carrying the T2D protective allele.

Comparison of differentiated *RREB1* N1171N to *RREB1* KO (LOF mutants) lines, which were characterised by increased endocrine progenitor (*NEUROG3*, *NKX2.2*, *NEUROD1*) and concomitantly decreased pancreas progenitor (*CPA2*, *HNF1B*, *NOTCH1*) marker expression at stage 5, suggested that the T2D protective allele acted as a GOF allele. This was supported by the observation that a subset of RREB1 *cis*-regulated genes revealed reciprocal expression patterns in *RREB1* N1171N lines when compared to expression profiles observed in differentiated *RREB1* KO clones.

Transcriptome profiling of differentiated *RREB1* lines pointed to an effect of the T2D protective allele on the generation of early pancreatic progenitors. Differential

gene and WGCN analysis detected significant differences in gene and module expression primarily at stage 3 during beta cell development. Although DEGs were detected using either an additive or a dominant model, gene expression profiles and ME expression results were more supportive of the dominant inheritance model, as *RREB1* D1171N and D1171D clones often showed similar behaviour which was distinct from *RREB1* N1171N lines.

Characterisation of T2D-associated variants is hampered by limitations of hiPSC-based beta cell developmental models

In vitro hiPSC-based differentiation models, although having been successfully applied in diabetes disease modelling, investigating the effects of single point mutations on beta cell development (Saarimäki-Vire et al. 2017; Balboa et al. 2018; Ng et al. 2019), are still in a developmental or optimisation phase and there are limitations associated with this model. For example, differentiation efficiencies can vary greatly across different hPSC lines, even if they are isogenic. Undetected off-target effects acquired through genome engineering might modify the differentiation ability of individual clones. Assessment of developmental marker expression in differentiated *RREB1* D1171D, D1171N and N1171N lines had shown high variability in expression levels for some of the TFs between clones of the same genotype, hampering inference on whether the observed effect was real, i.e. attributable to the cell's genotype, or just of technical nature.

A further confounding factor is presented by the observation that endocrine cell populations generated with the currently available *in vitro* differentiation protocols are quite heterogeneous (Rezania et al. 2014; Pagliuca et al. 2014; Russ et al. 2015; Perez-Alcantara et al. 2018). Bulk RNA-Seq might therefore not be able to pick up modest effects of T2D-associated variants on transcriptional differences during beta cell differentiation, as the majority of sequenced cells are not of the desired cell type. Indeed, less than 12% (more like 1-4%) of *RREB1* hiPSC-derived BLCs co-expressed NKX6.1 and C-peptide.

Purification of differentiated cells using flow cytometry with specific cell surface markers prior to bulk RNA-Seq represents a potential approach to reduce confounding effects arising from heterogeneous cell populations (Ramond et al. 2017; Ramond et al. 2018). Furthermore, single-cell RNA-Seq provides a much more sensitive and robust method to identify endocrine cells amongst a heterogeneous cell population and characterise subtle effects exerted by genetic variants, as shown by recent studies analysing single cells from human and mouse islets at different developmental time points and disease states as well as a study investigating NDM-associated *INS* mutations in hiPSC-derived BLCs (Carrano et al. 2017; Baron et al. 2016; Segerstolpe et al. 2016; Balboa et al. 2018). Therefore, to gain further insights into the impact of the *RREB1* T2D protective allele on beta cell development and to shed light on potential disease mechanisms transcriptome profiling using single-cell RNA-Seq represents a promising future approach.

Primary human islet function was not affected by p.D1171N

Despite the prominent insulin content phenotype observed in *RREB1* KO mature beta cells, no significant effect of the T2D protective allele on cellular insulin content was seen in islets from cadaveric donors heterozygous for p.D1171N. Again a subtle effect of the genetic variant on insulin protein might have been obscured by the heterogeneous nature of primary human islets. Despite matching for sex, age, BMI and CIT, donors might have been carriers of other T2D-associated genetic risk variants confounding insulin-associated effects. In addition, assuming *RREB1* rs9379084 is dominantly inherited, individuals carrying either one (p.D1171N) or two (p.D1171D) T2D risk alleles would be expected to be phenotypically similar, hence there would be no difference in insulin content and secretion measures in islets from donors of these two genotypes. As the Alberta Diabetes Institute IsletCore database only contained two islet samples from adult donors homozygous for the T2D protective allele (p.N1171N), experimental power was too low to draw any conclusions on their impact on insulin secretion or content.

6 Discussion

6.1 A novel role for RREB1 in beta cell development and function

The last 10 years has seen unprecedented progress in our understanding of the genetic landscape of T2D (Fuchsberger et al. 2016; Morris 2018; Mahajan et al. 2018a; Spracklen et al. 2019). Ongoing efforts focus on translating these genetic discoveries into T2D risk mechanisms with the aim of identifying novel drugable targets and pathways to refine current and discover novel therapeutic interventions for diabetes. Although, for the majority of T2D risk loci the underlying molecular mechanisms are still unknown, substantial progress has been made since the discovery of the first T2D GWAS hits (Barroso et al. 2019). Advances in high-throughput genome-wide techniques have improved regulatory annotations and enabled the generation of chromatin state maps in human pancreatic islets (Gaulton et al. 2010; Stitzel et al. 2010; Gaulton et al. 2015; Varshney et al. 2017; Varshney et al. 2018; Greenwald et al. 2019). Integration of epigenomic data with ever growing GWAS datasets has provided evidence that most variants affect regulatory sequences, i.e. map to tissue-specific enhancers (Parker et al. 2013; Pasquali et al. 2014; Thurner et al. 2018; Chiou et al. 2019). Identification of T2D-associated coding variants in a handful

of genes has facilitated rapid identification of effector transcripts and pointed to potential mechanisms underlying the T2D associations (Gloyn et al. 2003; Florez et al. 2004; Sladek et al. 2007; Orho-Melander et al. 2008; Fuchsberger et al. 2016; Lek et al. 2016). Disease models, comprising animal and cellular studies, are crucial for target validation. In particular, human-based cellular systems are of great importance for clinical translation, as rodent-based models often do not recapitulate human disease phenotypes due to species differences as discussed in Section 1.4. The work presented in this thesis contributes to the list of studies aiming to provide biological insights from T2D-associated variants through characterising the role of the T2D-associated gene *RREB1* in beta cell development and function.

Loss of RREB1 had a positive effect on endocrine differentiation, in particular on the generation of endocrine precursors, as well as the differentiation state of mature beta cells. Endocrine progenitor markers (*NEUROG3*, *NKX2.2*, *NEUROD1*) were significantly up-regulated while pancreas progenitor markers (*CPA2*, *HNF1B*, *NOTCH1*) were markedly down-regulated in hiPSC-derived *RREB1* KO EP cells during beta cell development. Loss of RREB1 in mature beta cells equally revealed increased expression of genes implicated in beta cell function, connectivity and maturity (*GCK*, *CHGB*, *SNAP25*, *GJD2*, *UCN3*, *CDK5R2*, *PTPRN*) with simultaneous down-regulation of beta cell disallowed genes (*PDGFRA*, *IGFBP4*, *CD302*).

Up-regulation of endocrine progenitor marker and beta cell maturity genes as well as down-regulation of pancreatic progenitor-specific and beta cell forbidden genes might partially be attributable to a direct effect of RREB1, as a subset of these genes were identified as being *cis*-regulated target genes of RREB1 in mature beta cells (Figure 6.1 (I)). This suggests a bivalent role for RREB1 as transcriptional activator and repressor in endocrine cells, in concordance with previous studies in various cellular systems as discussed in Section 1.5.3. Transcript levels of a subset of newly identified RREB1 *cis*-regulated genes, however, were not affected by loss of RREB1. These might be additionally regulated by other transcriptional activators or repressors masking an effect mediated by RREB1.

RREB1 binding sites were significantly enriched in islet open chromatin, in particularly in active promoters, highlighting RREB1 as a novel regulator of genes important for beta cell function and identity and representing important implications for future T2D GWAS loci fine-mapping efforts, which have previously shown that T2D- and FG-associated variants are enriched in islet regulatory elements, where they effect T2D risk through modification of endocrine-specific TFs binding sites (Parker et al. 2013; Pasquali et al. 2014; Thurner et al. 2018).

6.2 RREB1 and the RFX TF family

TF enrichment analysis identified RFX2 and RFX3 as key regulators driving gene expression variation across *RREB1* KO and control cells. Both the hiPSC-based developmental and the mature beta cell model were characterised by differential transcriptional activity of RFX2 and RFX3. RFX2/3 motif activity was significantly higher in *RREB1* KO compared to control cells during all stages of *in vitro* endocrine differentiation as well as in mature beta cells. As motif activity positively correlated with *RFX2* as well as *RFX3* expression, both RFX proteins likely acted as transcriptional activators. In line with this, RFX target genes *GPR56* and *CAMK2A* were significantly up-regulated in *RREB1* KO cell models.

Of the RFX TF family only RFX3 and RFX6 have been associated with beta cell development, formation and function, while a role for RFX2 in glucose homeostasis has not been described so far. Interestingly, *RFX2* revealed a similar expression profile compared to *RFX3* and *RFX6* in *RREB1* clones, i.e. all three genes were expressed in NEUROG3-positive endocrine progenitors and expression was maintained in developing and adult endocrine islet cells. *RFX2*, *RFX3* and *RFX6* have been shown to contain functional domains allowing them to dimerise as either homo- or heterodimers (Reith et al. 1994; Emery et al. 1996a; Horvath et al. 2004; Rual et al. 2005; Aftab et al. 2008) and human proteome-wide mapping of protein-protein interactions revealed that RFX6 physically interacted with RFX2 and RFX3 (Rual et al. 2005). Whether RFX2 and RFX3 form heterodimers in beta

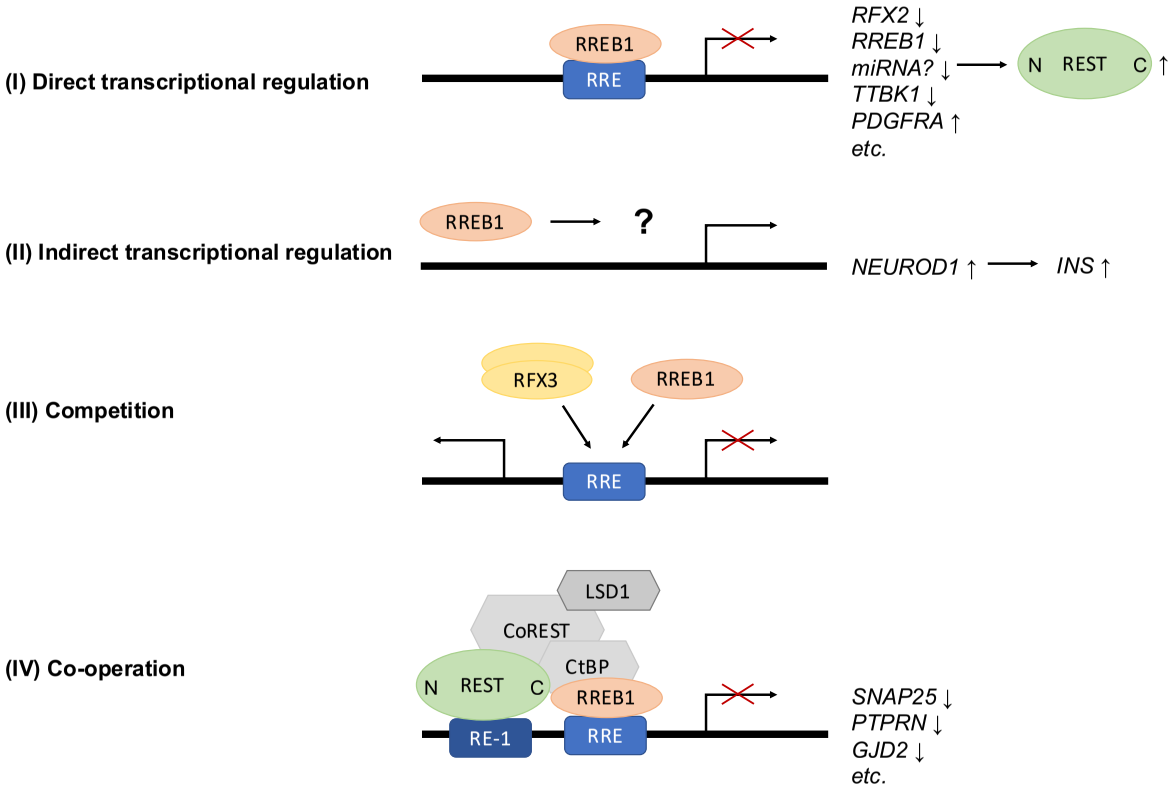


Figure 6.1: Models for RREB1-mediated gene regulation effects in developing endocrine and mature beta cells (I) RREB1 has been shown to both activate and repress transcription in beta cells. RREB1 *cis*-regulated genes comprise genes involved in endocrine differentiation, insulin secretion, cell-to-cell communication and beta cell forbidden genes. A RREB1-mediated effect on miRNAs might impact REST transcriptional activity through modulation of translation efficiency. (II) RREB1 might affect expression of other TFs which in turn modulate gene expression in beta cells. This might underlay the observed reduction in *NEUROD1* transcript levels in *RREB1* KO beta cells, which might be responsible for the down-regulation of *INS* mRNA. (III) RREB1 and RFX3 have been shown to bind to RRE in promoters to regulate gene transcription. Missing competition due to loss of RREB1 could have led to increased RFX3 transcriptional activity. (IV) RREB1 and REST have been shown to be components of multiprotein complexes. They might cooperate in transcriptional regulation through binding to nearby DNA elements and recruitment of co-repressors and chromatin-modifying enzymes including CtBP, CoREST and LSD1 to the promoter. RRE, Ras-responsive element; miRNA, non-coding RNA; RE-1, repressor element 1; CtBP, C-terminal binding protein; CoREST, REST co-repressor; LSD1, lysine-specific histone demethylase 1A.

cells to cooperatively regulate gene expression is currently unknown, but would be interesting to investigate in the future.

Both RFX2 and RFX3 have been shown to be expressed in ciliated tissues, where they modify FOXJ1-mediated cilia gene regulation (Bonnafe et al. 2004; Chung et al. 2012; Didon et al. 2013; Quigley et al. 2017). Endocrine cells have been shown to contain cilia, however their function and whether defective cilia add to compromised beta cell development and function in *Rfx3* KO mice is still to be determined (Ait-Lounis et al. 2007). Interestingly, only *RFX2*, not *RFX3* expression was positively affected by loss of RREB1 in mature beta cells. In line with this, ChIP-Seq analysis identified *RFX2*, but not *RFX3*, as RREB1 *cis*-regulated gene, highlighting RREB1 as a transcriptional repressor of *RFX2* in mature beta cells and likely during endocrine cell differentiation (Figure 6.1 (I)). A study aimed at the prediction of upstream transcriptional regulators of *RFX* genes using TF binding profile analysis did not identify the RREB1 TFBS as being statistically over-represented in *RFX* promoters (Sugiaman-Trapman et al. 2018). The experimentally validated regulator ESR1 shown to inhibit *RFX* transcription (Sugiaman-Trapman et al. 2018), was not found to play a role in *RREB1* KO beta cells, implicating that this is the first report highlighting RREB1 as a potential repressor of *RFX2*. Unexpectedly, only RNA interference mediated KD of *RFX3*, but not of *RFX2*, partially rescued gene expression changes associated with loss of RREB1. *CAMK2A* and *GPR56*, two of the top 20 strongest up-regulated and not identified as RREB1 *cis*-regulated genes were markedly down-regulated in *RFX3*-deficient *RREB1* KO and EV EndoC-βH1 cells, suggesting that increased transcriptional activity of RFX3, caused by RREB1 depletion, had contributed to the transcriptional phenotype associated with loss of RREB1 in mature beta cells. Experimentally proven RFX2 target genes are scarce and not necessarily (*COL1A2*) or only barely (*IL5RA*) expressed in human mature beta cells (Sengupta et al. 2002; Iwama et al. 1999). Therefore, it cannot be ruled out that gene expression changes in *RREB1* KO cells are partly attributable to increased RFX2 transcriptional activity and genes interrogated in siRFX2-KD studies were not reflective of true beta cell RFX2 *cis*-regulated genes.

RFX proteins have been suggested to assemble in multiple combinations with either members of the same family or other proteins, including TFs, to affect transcription regulation (Rual et al. 2005; Rual et al. 2005; Didon et al. 2013; Quigley et al. 2017). In addition to binding to the highly conserved X-box motif, RFX2 and RFX3 have been shown to bind to Ras-responsive elements (RRE) of gene promoters, affecting expression of Ras-responsive targets (Maijgren et al. 2004). RREB1 was first identified as a TF which bound to the upstream RRE of the human calcitonin gene (Thiagalingam et al. 1996) and has since been shown to regulate various other genes by interacting with RREs in their promoter regions (S Zhang et al. 2003; Yamane et al. 2013; Milon et al. 2010; Flajollet et al. 2009; Kent et al. 2013). As RREB1 could not be confirmed as transcriptional regulator of *RFX3*, the exact mechanism underlying up-regulation of RFX3 motif activity remains to be determined. RREB1 *cis*-regulated genes were enriched for computationally predicted RFX3 target genes, raising the question, whether RREB1 and RFX3 might compete for RRE binding. Missing competition caused by KO of RREB1 might explain up-regulation of RFX3 transcriptional activity (Figure 6.1 (III)).

Interestingly, individual depletion of RFX2 or RFX3 in EndoC-βH1 cells led to a significant decrease of RREB1 protein expression, while *RREB1* transcript levels remained unchanged. Whether unaffected mRNA levels are a true observation or confounded by the likely negative feedback loop in which RREB1 controls its own gene expression, is unclear. Therefore, it cannot be ruled out that reduced RREB1 protein levels are a consequence of down-regulated *RREB1* gene expression, in particular as the *RREB1* promoter has been predicted to contain RFX2 and RFX3 binding motifs (Jolma et al. 2010; Jolma et al. 2013; Matys et al. 2006). In particular for RFX2, the question arises, whether it participates in a loop with RREB1 to regulate its own gene expression. For example, high levels of RFX2 might stimulate RREB1 expression to represses further *RFX2* gene transcription. Equally, low levels of RFX2 protein might lead to down-regulation of RREB1 and subsequent derepression and increase of *RFX2* expression until sufficient RFX2 protein levels are present in the cell.

6.3 RREB1 and the transcriptional repressor REST

The TF REST has been suggested to be an important repressor of beta cell differentiation, which either functions upstream of or in parallel to NEUROG3 and blocks endocrine lineage initiation (D Martin et al. 2015). During *in vitro* beta cell development *REST* transcript levels were high in PGT, PFG and PE cells, dropped as cells further differentiated into EP and EN cells and stayed low in BLCs in *RREB1* KO and WT clones. In mature beta cells *REST* expression was very low, consistent with previous studies that have identified *REST* as a disallowed gene in adult beta cells (Atouf et al. 1997; D Martin et al. 2008).

REST motif activity was significantly higher in *RREB1* KO compared to control cells during the last four stages of *in vitro* beta cell differentiation (PE-BLC) as well as in mature beta cells. Increased transcriptional activity implied that REST was less repressive in cells depleted of RREB1, supporting a positive effect of RREB1 depletion on endocrine differentiation and beta cell maturity. Interestingly, REST motif activity negatively correlated with RREB1 transcriptional activity and ChIP-Seq revealed RREB1 binding in intergenic regions distal of the *REST* promoter. *REST* expression was unchanged by loss of RREB1, implicating that *REST* transcription might be regulated by multiple repressors, masking a putative effect of RREB1 on *REST* transcript levels.

The mechanism(s) leading to increased REST transcriptional activity upon RREB1 depletion are unknown. REST protein levels have been suggested of being modulated by non-coding RNAs (miRNAs) (Mortazavi et al. 2006) and RREB1 has been implicated in miRNA promoter repression (Kent et al. 2010). Therefore, loss of RREB1 might have increased expression of REST-targeting miRNAs which in turn affected REST protein levels through RNA interference (Hornstein et al. 2006; Kloosterman et al. 2006) (Figure 6.1 (I)). RREB1 has been identified as a component of a multiprotein complex containing CtBP, CoREST and chromatin-modifying enzymes. In rodents, RREB1 was shown to recruit CtBP and its associated

proteins LSD1 and CoREST to NEUROD1 target promoters including the *Sct*, *Ins1*, *Ins2* and *Gck* promoters, to potentiate NEUROD1-mediated transcription (Ray et al. 2014). Removal of repressive H3K9 methylation marks through LSD1 with subsequent acetylation by the NEUROD1-associated histone acetyltransferase PCAF was suggested as a potential mechanism underlying transcriptional activation (Ray et al. 2014). RREB1 ChIP-Seq analysis revealed a substantial overlap of RREB1 *cis*-regulated and REST target genes. As RREB1 and REST have been shown to recruit similar co-repressor complexes (Andrés et al. 1999; Y Shi et al. 2004; Garriga-Canut et al. 2006), the question arises, whether multiprotein complexes comprising REST, RREB1, co-repressors and associated chromatin-modifying proteins underlay regulation of detected DEGs important for beta cell differentiation and function (Figure 6.1 (IV)). This model in combination with insufficient publicly available data on RREB1 *cis*-regulated genes in beta cells, might explain why TF enrichment analysis of newly identified RREB1 target genes only predicted REST and not RREB1 as putative upstream regulator.

6.4 RREB1 may affect insulin content through NEUROD1

Loss of RREB1 modified insulin content in mature beta cells through a yet unknown mechanism. Significantly reduced cellular insulin content was likely a consequence of down-regulated *INS* gene transcription, as genes encoding the proprotein convertases 1 and 2, which process proinsulin in beta cell secretory granules (Kaufmann et al. 1995), were markedly up-regulated in *RREB1* KO beta cells, arguing against defective insulin processing as the underlying cause leading to reduced insulin content. RREB1 binding was not detected in or close to the *INS* promoter, thus, *INS* gene expression was likely not directly regulated by RREB1 (Figure 6.1 (II)). NEUROD1 and PDX1 are important transcriptional activators of the *INS* gene (Glick et al. 2000). Transcriptional activities of both TFs were reduced and *NEUROD1* transcript levels were significantly lower in *RREB1* KO EndoC-βH1

cells, highlighting these two TFs as most likely responsible for down-regulated *INS* mRNA levels. ChIP-Seq did not identify *NEUROD1* as a direct target gene of RREB1, hence mechanisms leading to down-regulation of *NEUROD1* remain to be determined. Interestingly, during beta cell development hiPSC-derived *RREB1* KO PE cells were characterised by significantly increased *NEUROD1* levels, before expression decreased again to markedly lower levels in BLCs, consistent with observations in RREB1-deficient mature beta cells, suggesting distinct effects of RREB1 on *NEUROD1* during endocrine differentiation and in mature beta cells. As NEUROD1 has been shown to co-occupy the *Ins1* and *Ins2* promoters with the CtBP/RREB1 co-repressor complex in murine beta cells (Ray et al. 2014), this TF should be prioritised in future studies investigating potential mechanisms underlying reduced insulin content in mature beta cells as a consequence of loss of RREB1.

6.5 The *RREB1* T2D protective allele p.N1171 is likely a GOF allele

Differentiation of hiPSC lines homozygous for the p.N1171 variant at *RREB1* rs9379084 into BLCs suggested that the T2D protective allele acted as a GOF allele, negatively affecting endocrine development. Endocrine progenitor markers *NEUROG3*, *NEUROD1* and *PAX4* were significantly down-regulated compared to levels observed in hiPSC-derived *RREB1* D1171N and D1171D EP cells. This finding was slightly unexpected as the p.N1171 allele is associated with a protective effect on T2D risk. As p.N1171 represents a coding variant and *RREB1* is expressed in other diabetes-relevant tissues, effects in these tissues might underlay the T2D association. In particular adipose tissue represents an interesting candidate, as RREB1 has been identified as a novel activator of brown adipocyte differentiation (Brunmeir et al. 2016) and was reported to play a central role in brown adipocyte-specific gene expression regulation in mice (D Pan et al. 2015). RREB1 was shown to recruit the demethylase JMJD3 to *Ucp1* and *Cidea* promoters, where JMJD3 demethylates H3K27me3 marks, which is required for gene expression in brown preadipocytes

(D Pan et al. 2015). Interestingly, removal of repressive chromatin marks in white adipocytes has also been shown to cause browning of white adipose tissue (D Pan et al. 2015) and RREB1 ChIP-Seq analysis identified RREB1 binding in the human *CIDEA* promoter in mature beta cells. The role of RREB1 and a potential effect of the T2D-associated variant p.D1171N in adipocytes will require further investigation.

6.6 Limitations of hiPSC-based studies

Human PSC-based disease modelling represents an approach that is still being developed, thus well-recognised limitations have to be taken into account when interpreting results.

First, differentiation efficiencies are often variable between hPSC lines of different genetic backgrounds, confounding interpretation of the impact of genes or genetic variants on beta cell development (Kyttälä et al. 2016; Rouhani et al. 2014). Generation of isogenic cell lines, i.e. introducing a patient-specific variant in a control hPSC line or correcting patient-derived hiPSC lines using genome editing tools, has partially addressed this limitation. Nonetheless, evaluation of key developmental marker expression in differentiated cells sharing the same genetic background revealed variable differentiation efficiencies amongst *RREB1* WT clones (Section 3.3.3.3) as well as between the unedited *RREB1* D1171N ctrl and the parental SB #1 and #2 lines (Section 5.3.3.1), strongly implicating that the genome editing process affected hiPSCs in an unwanted way. Genome engineering might have caused undetected off-target effects that modified the differentiation ability of individual clones. Effects on differentiation efficiencies in turn might mask subtle and mild effects of genetic variants affecting endocrine cell differentiation capacity or even lead to false conclusions, when technical noise is mistakenly interpreted as a true biological effect.

Over the last couple of years several research groups have developed various *in vitro* and cell-based techniques to identify CRISPR/Cas9-associated genetic off-target effects genome-wide. *In vitro* assays, including Digenome-seq (Kim et al. 2015),

CIRCLE-seq (Tsai et al. 2017) and SITE-Seq (Cameron et al. 2017), are based on *in vitro* Cas9 digestion of gDNA followed by whole-genome sequencing, while cell-based assays comprising GUIDE-seq (genome-wide unbiased identification of DNA DSBs using sequencing) (Tsai et al. 2015), LAM-HTGTS (linear amplification–mediated high-throughput genome-wide translocation sequencing) (Frock et al. 2015) and BLISS (Breaks Labelling In Situ and Sequencing) (Yan et al. 2017) can detect both endogenously occurring and Cas9-created DNA DSBs in a cell type-specific manner.

In vitro beta cell differentiation has been shown to generate heterogeneous endocrine cell populations (Rezania et al. 2014; Pagliuca et al. 2014; Russ et al. 2015; Perez-Alcantara et al. 2018). Data generated via bulk analysis, including transcriptome profiling, might be quite variable, masking effects of small magnitude. Human iPSC-derived *RREB1* BLCs contained between ∼1-12% (Section 5.3.3.1) and ∼16-37% (Section 3.3.3.3) cells that co-expressed NKX6.1 and C-peptide. Thus, it would not be surprising, if a modest effect of the T2D-associated p.N1171 variant on transcriptional differences during beta cell differentiation did not get picked up in bulk RNA-Seq analysis. Isolation of purer differentiated cells via specific cell surface markers and flow cytometry prior to bulk RNA-Seq represents a potential approach to reduce confounding effects arising from heterogeneous cell populations (Ramond et al. 2017; Ramond et al. 2018). Moreover, single-cell gene expression studies comprising single-cell RT-qPCR and single-cell RNA-Seq represent more sensitive and robust approaches to interrogate cellular phenotypes, circumventing limitations associated with heterogeneous differentiation (Petersen et al. 2017; Balboa et al. 2018).

Currently available *in vitro* differentiation protocols use recombinant growth factors and small molecules as exogenous signalling cues to direct hPSCs along the endocrine lineage into BLCs. *In vitro* derived BLCs have been reported to still being functionally immature (Bruin et al. 2014), and their transcriptome resembling fetal rather than adult beta cells (Hrvatin et al. 2014). Consequently, modelling phenotypes associated with metabolic regulation of insulin secretion using hPSC-based systems is limited. For this reason, the consequences of loss

of RREB1 on GSIS were evaluated in the mature beta cell line EndoC-βH1. Interestingly, hPSC-derived BLCs can further be matured through transplantation of cells under the kidney capsule of immunocompromised mice, which has been shown to successfully correct diabetes in these animals (Kroon et al. 2008; Bruin et al. 2013; Rezania et al. 2013; Pagliuca et al. 2014; Rezania et al. 2014). This exemplifies the importance of the cellular environment *in vivo* for proper cell differentiation and highlights that *in vitro* hPSC-based disease models might not accurately recapitulate diabetes progression, which has to be taken into account when extrapolating observations into clinic. Nonetheless, building on the observed *in vivo* maturation capacity, clinical trials involving stem cell-derived islet replacement therapies in humans have started to emerge. In 2014, a Phase I/II clinical trial in patients with type 1 diabetes has been launched by ViaCyte to evaluate the safety and efficacy of a stem cell-derived encapsulated cell replacement therapy for diabetes (https://clinicaltrials.gov/ct2/show/NCT02239354).

Despite these limitations, hPSCs have greatly advanced our understanding of human beta cell development and diabetes pathogenesis through identification of genes and proteins with important roles in human endocrine differentiation as elaborated in Section 1.4. Several research groups are working on improving and optimising *in vitro* differentiation protocols to overcome limitations currently associated with hPSC-derived BLCs which will undoubtedly increase experimental robustness and reproducibility in the future.

6.7 Next steps and future directions

Evaluation of loss of RREB1 in two different cellular systems has pointed to a novel role for RREB1 in glucose homeostasis and left us with several exciting unanswered questions.

RREB1-deficient EndoC-βH1 cells simultaneously displayed characteristics of immature and mature beta cells. For example, beta cell specific signature and functional maturity genes were significantly up-regulated, while *RREB1* KO beta

cells were characterised by markedly reduced cellular insulin content. While up-regulation of genes involved in beta cell function can be attributed to a direct transcriptional effect of RREB1, the exact mechanisms underlying down-regulation of *INS* transcription and reduced insulin content remains to be determined. Future efforts should focus on NEUROD1 as discussed in Section 6.4. As decreased insulin protein did not affect insulin secretion in response to a glucose stimulus under normal conditions, and in light of the finding that reduced insulin production causes less ER stress and beta cell fragility (Szabat et al. 2016), the overall effect of loss of RREB1 seemed to have been positive on mature beta cells. However, FRSK studies showed that reduced insulin content negatively affected beta cells' ability to perform under conditions of high insulin demand, which might be an important implication during pregnancy or obesity.

The question arises whether expression levels of RREB1 have to be tightly controlled in mature beta cells to guarantee proper cellular function. *RREB1* overexpression studies in mature beta cells might shed light on optimal RREB1 levels. In particular, it would be interesting to assess the effect of increased RREB1 expression on insulin content and potentially resulting ER stress, in case of up-regulated *INS* transcription.

The mechanisms underlying RREB1 loss-associated increased transcriptional activity of RFX3 and REST are currently unknown. As RREB1 could not be confirmed as transcriptional regulation of these two genes, interaction at the protein level represents a possible explanation (Figure 6.1 (III) and (IV)). Co-immunoprecipitation experiments represent a promising approach to evaluate whether RREB1 and RFX proteins or RREB1 and REST interact with each other. Unfortunately, commercially available RREB1 antibodies are of insufficient specificity for IP. Although ChIP-Seq was successfully performed in *RREB1* KO EndoC-βH1 cells overexpressing FLAG-tagged RREB1 using a FLAG antibody, RREB1 IP attempts relying on the same cellular system and antibody have been of insufficient quality so far. Stable overexpression of FLAG-tagged RREB1 in *RREB1* KO EndoC-βH1 cells using a lentivirus-based system presents a promising

alternative and IP coupled to mass spectrometry will hopefully shed light on potential RREB1 protein binding partners in mature beta cells genome-wide.

Finally, while efforts to investigate the impact of the *RREB1* T2D-associated variant at rs9379084 on beta cell development have suggested a GOF role for the T2D protective allele, potential disease mechanisms underlying the T2D association remain unknown, likely due to the limitations related to hiPSC-based models discussed in Section 6.6. Better insights into T2D risk mechanisms might be gained by utilising single-cell approaches, i.e. transcriptome profiling at the single cell level or introduction of the T2D-associated variant into GFP-labelled hiPSC reporter lines, followed by differentiation into BLCs and RNA-Seq analysis of fluorescence-activated cell sorted differentiated cells (Gupta et al. 2018).

In addition, overexpression of *RREB1* p.N1171 plasmids in *RREB1* KO EndoC-βH1 cells and assessment of GSIS might reveal, whether the T2D protective allele impacts mature beta cell function. Despite stable genomic integration of Cas9 in *RREB1* KO EndoC-βH1 cells, overexpression studies for ChIP-Seq did show that expected Cas9-mediated cleavage of *RREB1* expression plasmids was not a problem, as RREB1 protein expression was readily detectable by WB analysis. With the ever increasing availability of human islets for research, characterisation of a larger number of islets from donors homozygous for the p.N1171 allele might shed further light on the impact of the T2D protective allele on GSIS and insulin content.

Overall, my results have identified RREB1 as a novel regulator of glucose homeostasis and exemplify the power of integrating genetic analysis with cellular studies using genome edited human iPSC and beta cell lines to drive clinical translation of T2D-associated genes.

Appendices

A

Material Tables

Table A.1: Primer pairs for PCR and Sanger sequencing

Gene	**Genomic region** (*RREB1* NM_001003699.3)	**Forward primer (5' to 3')**	**Reverse primer (5' to 3')**
RREB1	exon 3-4	CAGGCAGTACTGTAGGGG	AGGCCTCTGGGTATGTCC
RREB1	exon 4	GCTTCCATCACTCTGGGC	AGGCCTCTGGGTATGTCC
RREB1	exon 5	GACAGGGCCGTGACTTGC	AATAGGCTGGTCCGAAGG
RREB1	exon 8	TTATTACCAGTGCCTGGC	CCCAAGTTCAATCACTGG
RREB1	exon 10	CCCAGAGGCTGCCTCTC	AGGAAGTAGGACTCACTGCCTG
RREB1	exon 11	TTCTTGAAGGGATGGAGG	GCTGTGGGCTCTGAATCC
RREB1	exon 12	CCACCAAGCTCATGGAC	TCCAGCGCCGCTTCAGG
TP53	ORF	CTCAACAAGATGTTTTGCCA	AGATTCTCTTCCTCTGTGCG
LKO.1 5' (hU6)	plasmid	GACTATCATATGCTTACCGT	NA
M13	plasmid	TGTAAAACGACGGCCAGT	CAGGAAACAGCTATGAC
T7	plasmid	TAATACGACTCACTATAGGG	NA

Table A.2: PCR conditions

RREB1 PCR conditions

Step	Temperature	Duration	Cycles
Initial denaturation	94 °C	10min	1
Denaturation	94 °C	1min	
Annealing	63 °C	1min	32
Elongation	72 °C	1min	
Final extension	72 °C	10min	1

TP53 PCR conditions

Step	Temperature	Duration	Cycles
Initial denaturation	94 °C	10min	1
Denaturation	94 °C	1min	
Annealing	57 °C	1min	32
Elongation	72 °C	30sec	
Final extension	72 °C	10min	1

Table A.3: Primary antibodies used for WB

Antibody	Recommended dilution	Species	Manufacturer	Cat #
FLAG	1:10,000	mouse	Sigma Aldrich	F3165
GAPDH	1:10,000	rabbit	Abcam	ab37168
NEUROD1	1:1000	mouse	Santa Cruz	sc-46684
RFX2	1:1000	rabbit	Human Protein Atlas	HPA048969
RFX3	1:1000	rabbit	Human Protein Atlas	HPA035689
RFX6	1:1000	rabbit	Human Protein Atlas	HPA037696
RREB1	1:500	rabbit	Human Protein Atlas	HPA001756
RREB1	1:500	rabbit	Human Protein Atlas	HPA034843
β-Tubulin (E-10)	1:2000	mouse	Santa Cruz	sc-365791

Table A.4: Secondary antibodies used for WB

Antibody	Recommended dilution	Manufacturer	Cat #
Goat anti-Rabbit IgG, HRP	1:2500	Thermo Fisher Scientific	31460
Rabbit anti-Mouse IgG, HRP	1:2500	Thermo Fisher Scientific	31450

Table A.5: Primary antibodies used for immunostaining

Antibody	Recommended dilution	Species	Manufacturer	Cat #
Belfast anti-insulin	1:500	guinea-pig	in house	NA
RREB1	1:50	rabbit	Human Protein Atlas	HPA001756

Table A.6: Secondary antibodies used for immunostaining

Antibody	Recommended dilution	Manufacturer	Cat #
Donkey-anti-rabbit AF-488	1:100	Invitrogen	A21206
Goat-anti-guinea pig AF-555	1:100	Invitrogen	A21435

Table A.7: FACS antibodies including isotype controls

Antibody	Recommended dilution	Species	Conjugate	Manufacturer	Cat #
NANOG	1:10	mouse	PE	BD Pharmingen™	560589
OCT4	1:10	mouse	PerCP-Cy™ 5.5	BD Pharmingen™	560589

Continued on next page

Antibody	Recommended dilution	Species	Conjugate	Manufacturer	Cat #
SOX2	1:10-1:40	mouse	AF-647	BD Pharmingen™	560589
SSEA-4	1:10	mouse	FITC	BD Pharmingen™	560126
IgG_1, κ isotype	1:10	mouse	PE	BD Pharmingen™	560589
IgG_1, κ isotype	1:10	mouse	PerCP-Cy™ 5.5	BD Pharmingen™	560589
IgG_{2a}, κ isotype	1:10	mouse	AF-647	BD Pharmingen™	560589
IgG_3, κ isotype	1:10	mouse	FITC	BD Pharmingen™	555578
CXCR4	1:40	mouse	PE	R&D Systems	FAB173P
C-peptide	1:200	mouse	AF-647	BD Pharmingen™	565831
NEUROD1	1:40	mouse	PE	BD Pharmingen™	563001
NKX6.1	1:40	mouse	AF-647	BD Pharmingen™	563338
NKX6.1	1:40	mouse	PE	BD Pharmingen™	563023
PDX1	1:40	mouse	PE	BD Pharmingen™	562161
SOX17	1:40	mouse	AF-488	BD Pharmingen™	562205
IgG_{2B} isotype	1:40	mouse	PE	R&D Systems	IC0041P
IgG_1, κ isotype	individual	mouse	PE	BD Pharmingen™	554680
IgG_1, κ isotype	individual	mouse	AF-488	BD Pharmingen™	557721
IgG_1, κ isotype	individual	mouse	AF-647	BD Pharmingen™	557732

Top, pluripotency markers; bottom, differentiation stage-specific expression markers; individual, amount of isotype control was matched to amount of antibody used.

Table A.8: siRNAs for RNA interference studies

Target gene	Manufacturer	Cat #
Non-targeting pool	Dharmacon	D-001810-10-05
RFX2	Dharmacon	L-011129-00-0005
RFX3	Dharmacon	L-011764-00-0005
RREB1	Dharmacon	L-019150-00-0005

All siRNAs were purchased as ON-TARGETplus SMARTpool containing a mixture of four siRNA.

Table A.9: TaqMan® gene expression assays used for real-time PCR and their target amplification regions

Target genes	Assay number	Target region	Manufacturer
ABCC8	Hs00165861_m1	Spans exon boundary	Thermo Fisheer Scientific
ADGRG1	Hs00173754_m1	Spans exon boundary	Thermo Fisheer Scientific
CACNA1A	Hs01579431_m1	Spans exon boundary	Thermo Fisher Scientific
CAMK2A	Hs003924505_m1	Spans exon boundary	Thermo Fisher Scientific
GAPDH	Hs99999905_m1	Spans exon boundary	Thermo Fisher Scientific
GCG	Hs01031536_m1	Spans exon boundary	Thermo Fisher Scientific
GJD2	Hs00706940_s1	Both primers and probe map within a single exon	Thermo Fisher Scientific
HNF4A	Hs00230853_m1	Spans exon boundary	Thermo Fisher Scientific
INS	Hs00355773_m1	Spans exon boundary	Thermo Fisher Scientific
KCNJ11	Hs00265026_s1	Both primers and probe map within a single exon	Thermo Fisher Scientific
NEUROD1	Hs00159598_m1	Spans exon boundary	Thermo Fisher Scientific
NEUROG3	Hs01875204_s1	Both primers and probe map within a single exon	Thermo Fisher Scientific
NKX6.1	Hs00232355_m1	Spans exon boundary	Thermo Fisher Scientific

Continued on next page

Target genes	Assay number	Target region	Manufacturer
PAX4	Hs00173014_m1	Spans exon boundary	Thermo Fisher Scientific
PDX1	Hs00236830_m1	Spans exon boundary	Thermo Fisher Scientific
POU5F1	Hs04260367_gH	Both primers and probe map within a single exon	Thermo Fisher Scientific
PPIA	Hs99999904_m1	Amplicon spans exon junction, probe and one primer sit within one exon	Thermo Fisher Scientific
RFX2	Hs00172177_m1	Spans exon boundary	Thermo Fisher Scientific
RFX3	Hs01060440_m1	Spans exon boundary	Thermo Fisher Scientific
RREB1	Hs00366111_m1	Spans exon boundary	Thermo Fisher Scientific
SOX17	Hs00751752_s1	Both primers and probe map within a single exon	Thermo Fisher Scientific
TBP	Hs00427620_m1	Spans exon boundary	Thermo Fisher Scientific

Table A.10: Real-time TaqMan® quantitative PCR conditions

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

References

Aftab, Syed et al. (Aug. 2008). "Identification and characterization of novel human tissue-specific RFX transcription factors." In: *BMC evolutionary biology* 8.1, p. 226.

Aguayo-Mazzucato, Cristina et al. (Oct. 2015). "MAFA and T3 Drive Maturation of Both Fetal Human Islets and Insulin-Producing Cells Differentiated From hESC." In: *The Journal of Clinical Endocrinology & Metabolism* 100.10, pp. 3651–3659.

Ahlgren, U et al. (Jan. 1997). "Independent requirement for ISL1 in formation of pancreatic mesenchyme and islet cells." In: *Nature* 385.6613, pp. 257–260.

Ahlgren, U et al. (June 1998). "beta-cell-specific inactivation of the mouse Ipf1/Pdx1 gene results in loss of the beta-cell phenotype and maturity onset diabetes." In: *Genes & development* 12.12, pp. 1763–1768.

Ait-Lounis, Aouatef et al. (Apr. 2007). "Novel function of the ciliogenic transcription factor RFX3 in development of the endocrine pancreas." In: *Diabetes* 56.4, pp. 950–959.

Ait-Lounis, Aouatef et al. (July 2010). "The transcription factor Rfx3 regulates beta-cell differentiation, function, and glucokinase expression." In: *Diabetes* 59.7, pp. 1674–1685.

Alberti, K G et al. (July 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." In: *Diabetic medicine : a journal of the British Diabetic Association* 15.7, pp. 539–553.

Allen, Hana Lango et al. (Dec. 2011). "GATA6 haploinsufficiency causes pancreatic agenesis in humans." In: *Nature Genetics* 44.1, pp. 20–22.

Almgren, P et al. (2011). "Heritability and familiality of type 2 diabetes and related quantitative traits in the Botnia Study". In: *Diabetologia* 54.11, pp. 2811–2819.

Ammälä, C et al. (May 1993). "Calcium-independent potentiation of insulin release by cyclic AMP in single beta-cells." In: *Nature* 363.6427, pp. 356–358.

Anders, Carolin et al. (Sept. 2014). "Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease." In: *Nature* 513.7519, pp. 569–573.

Anders, Simon et al. (Oct. 2012). "Detecting differential usage of exons from RNA-seq data." In: *Genome research* 22.10, pp. 2008–2017.

Andersson, Lotta E et al. (2015). "Characterization of stimulus-secretion coupling in the human pancreatic EndoC-βH1 beta cell line." In: *PloS one* 10.3, e0120879.

Andrés, M E et al. (Aug. 1999). "CoREST: a functional corepressor required for regulation of neural-specific gene expression." In: *Proceedings of the National Academy of Sciences* 96.17, pp. 9873–9878.

Apelqvist, Åsa et al. (Aug. 1999). "Notch signalling controls pancreatic cell differentiation". In: *Nature* 400.6747, pp. 877–881.

Ardestani, Amin et al. (Feb. 2018). "The Hippo Signaling Pathway in Pancreatic β-Cells: Functions and Regulations." In: *Endocrine reviews* 39.1, pp. 21–35.

Arnold, Phil et al. (Feb. 2012). "MotEvo: integrated Bayesian probabilistic methods for inferring regulatory sites and motifs on multiple alignments of DNA sequences." In: *Bioinformatics (Oxford, England)* 28.4, pp. 487–494.

Artner, Isabella et al. (Oct. 2010). "MafA and MafB regulate genes critical to beta-cells in a unique temporal manner." In: *Diabetes* 59.10, pp. 2530–2539.

Ashcroft, F M (1988). "Adenosine 5'-triphosphate-sensitive potassium channels." In: *Annual review of neuroscience* 11.1, pp. 97–118.

Ashery-Padan, Ruth et al. (May 2004). "Conditional inactivation of Pax6 in the pancreas causes early onset of diabetes." In: *Developmental biology* 269.2, pp. 479–488.

Atouf, F et al. (Jan. 1997). "Expression of neuronal traits in pancreatic beta cells. Implication of neuron-restrictive silencing factor/repressor element silencing transcription factor, a neuron-restrictive silencer." In: *Journal of Biological Chemistry* 272.3, pp. 1929–1934.

Bae, Byoung-Il et al. (Feb. 2014). "Evolutionarily dynamic alternative splicing of GPR56 regulates regional cerebral cortical patterning." In: *Science* 343.6172, pp. 764–768.

Balboa, Diego et al. (Nov. 2018). "Insulin mutations impair beta-cell development in a patient-derived iPSC model of neonatal diabetes." In: *eLife* 7.

Balboa, Diego et al. (Jan. 2019). "Concise Review: Human Pluripotent Stem Cells for the Modeling of Pancreatic β-Cell Pathology." In: *Stem cells (Dayton, Ohio)* 37.1, pp. 33–41.

Balwierz, Piotr J et al. (May 2014). "ISMARA: automated modeling of genomic signals as a democracy of regulatory motifs." In: *Genome research* 24.5, pp. 869–884.

Barbacci, E et al. (Nov. 1999). "Variant hepatocyte nuclear factor 1 is required for visceral endoderm specification." In: *Development (Cambridge, England)* 126.21, pp. 4795–4805.

Baron, Maayan et al. (Oct. 2016). "A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure." In: *Cell systems* 3.4, 346–360.e4.

Barrangou, Rodolphe et al. (Mar. 2007). "CRISPR provides acquired resistance against viruses in prokaryotes." In: *Science* 315.5819, pp. 1709–1712.

Barroso, Inês et al. (Mar. 2019). "The Genetic Basis of Metabolic Disease." In: *Cell* 177.1, pp. 146–161.

Basford, C L et al. (Feb. 2012). "The functional and molecular characterisation of human embryonic stem cell-derived insulin-positive cells compared with adult pancreatic beta cells." In: *Diabetologia* 55.2, pp. 358–371.

Battaglioli, Elena et al. (Oct. 2002). "REST repression of neuronal genes requires components of the hSWI.SNF complex." In: *Journal of Biological Chemistry* 277.43, pp. 41038–41045.

Beer, Nicola L et al. (Nov. 2009). "The P446L variant in GCKR associated with fasting plasma glucose and triglyceride levels exerts its effect through increased glucokinase activity in liver." In: *Human Molecular Genetics* 18.21, pp. 4081–4088.

Beith, Jennifer L et al. (May 2008). "Insulin stimulates primary beta-cell proliferation via Raf-1 kinase." In: *Endocrinology* 149.5, pp. 2251–2260.

Bernal-Mizrachi, Ernesto et al. (Mar. 2014). "Human β-cell proliferation and intracellular signaling part 2: still driving in the dark without a road map." In: *Diabetes* 63.3, pp. 819–831.

Bhushan, A et al. (Dec. 2001). "Fgf10 is essential for maintaining the proliferative capacity of epithelial progenitor cells during early pancreatic organogenesis." In: *Development (Cambridge, England)* 128.24, pp. 5109–5117.

Blum, Barak et al. (Feb. 2012). "Functional beta-cell maturation is marked by an increased glucose threshold and by expression of urocortin 3." In: *Nature biotechnology* 30.3, pp. 261–264.

Bobis-Wozowicz, Sylwia et al. (Apr. 2011). "Targeted genome editing in pluripotent stem cells using zinc-finger nucleases." In: *Methods (San Diego, Calif.)* 53.4, pp. 339–346.

Bocian-Sobkowska, J et al. (Aug. 1999). "Polyhormonal aspect of the endocrine cells of the human fetal pancreas." In: *Histochemistry and cell biology* 112.2, pp. 147–153.

Bolotin, Alexander et al. (Aug. 2005). "Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin." In: *Microbiology (Reading, England)* 151.Pt 8, pp. 2551–2561.

Bonàs-Guarch, Silvia et al. (Jan. 2018). "Re-analysis of public genetic data reveals a rare X-chromosomal variant associated with type 2 diabetes." In: *Nature communications* 9.1, pp. 321–14.

Bonnafe, E et al. (May 2004). "The transcription factor RFX3 directs nodal cilium development and left-right asymmetry specification." In: *Molecular and Cellular Biology* 24.10, pp. 4417–4427.

Bonny, C et al. (June 2000). "IB1 reduces cytokine-induced apoptosis of insulin-secreting cells." In: *Journal of Biological Chemistry* 275.22, pp. 16466–16472.

Bosco, Domenico et al. (May 2010). "Unique arrangement of alpha- and beta-cells in human islets of Langerhans." In: *Diabetes* 59.5, pp. 1202–1210.

Bottino, Rita et al. (Oct. 2004). "Response of human islets to isolation stress and the effect of antioxidant treatment." In: *Diabetes* 53.10, pp. 2559–2568.

Brinkman, Eva K et al. (Dec. 2014). "Easy quantitative assessment of genome editing by sequence trace decomposition." In: *Nucleic acids research* 42.22, e168–e168.

Briscoe, J et al. (Apr. 1999). "Homeobox gene Nkx2.2 and specification of neuronal identity by graded Sonic hedgehog signalling." In: *Nature* 398.6728, pp. 622–627.

Broglio, F et al. (2006). "Ghrelin: from somatotrope secretion to new perspectives in the regulation of peripheral metabolic functions." In: *Frontiers of hormone research* 35, pp. 102–114.

Brouns, Stan J J et al. (Aug. 2008). "Small CRISPR RNAs guide antiviral defense in prokaryotes." In: *Science* 321.5891, pp. 960–964.

Bruin, Jennifer E et al. (Sept. 2013). "Maturation and function of human embryonic stem cell-derived pancreatic progenitors in macroencapsulation devices following transplant into mice." In: *Diabetologia* 56.9, pp. 1987–1998.

Bruin, Jennifer E et al. (Jan. 2014). "Characterization of polyhormonal insulin-producing cells derived in vitro from human embryonic stem cells." In: *Stem cell research* 12.1, pp. 194–208.

Brunmeir, Reinhard et al. (Dec. 2016). "Comparative Transcriptomic and Epigenomic Analyses Reveal New Regulators of Murine Brown Adipogenesis." In: *PLoS genetics* 12.12, e1006474.

Buecker, Christa et al. (June 2012). "Enhancers as information integration hubs in development: lessons from genomics." In: *Trends in genetics : TIG* 28.6, pp. 276–284.

Buenrostro, Jason D et al. (Dec. 2013). "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position." In: *Nature methods* 10.12, pp. 1213–1218.

Busche, Stephan et al. (Dec. 2015). "Population whole-genome bisulfite sequencing across two tissues highlights the environment as the principal source of human methylome variation." In: *Genome biology* 16.1, pp. 290–18.

Byrne, M M et al. (June 1995). "Altered insulin secretory responses to glucose in subjects with a mutation in the MODY1 gene on chromosome 20." In: *Diabetes* 44.6, pp. 699–704.

Cabrera, Over et al. (Feb. 2006). "The unique cytoarchitecture of human pancreatic islets has implications for islet cell function." In: *Proceedings of the National Academy of Sciences* 103.7, pp. 2334–2339.

Calo, Eliezer et al. (Mar. 2013). "Modification of enhancer chromatin: what, how, and why?" In: *Molecular cell* 49.5, pp. 825–837.

Cameron, Peter et al. (June 2017). "Mapping the genomic landscape of CRISPR-Cas9 cleavage." In: *Nature methods* 14.6, pp. 600–606.

Cardenas-Diaz, Fabian L et al. (Aug. 2019). "Modeling Monogenic Diabetes using Human ESCs Reveals Developmental and Metabolic Deficiencies Caused by Mutations in HNF1A." In: *Cell stem cell* 25.2, 273–289.e5.

Carrano, Andrea C et al. (Sept. 2017). "Interrogating islets in health and disease with single-cell technologies." In: *Molecular metabolism* 6.9, pp. 991–1001.

Carrasco, Manuel et al. (Oct. 2012). "GATA4 and GATA6 control mouse pancreas organogenesis." In: *The Journal of clinical investigation* 122.10, pp. 3504–3515.

Carrat, Gaelle R et al. (Jan. 2017). "Decreased STARD10 Expression Is Associated with Defective Insulin Secretion in Humans and Mice." In: *American journal of human genetics* 0.0.

Casellas, Alba et al. (July 2015). "Insulin-like Growth Factor 2 Overexpression Induces β-Cell Dysfunction and Increases Beta-cell Susceptibility to Damage." In: *Journal of Biological Chemistry* 290.27, pp. 16772–16785.

Cebola, Inês et al. (May 2015). "TEAD and YAP regulate the enhancer network of human embryonic pancreatic progenitors." In: *Nature cell biology* 17.5, pp. 615–626.

Chandra, Vikash et al. (Dec. 2014). "RFX6 regulates insulin secretion by modulating Ca2+ homeostasis in human β cells." In: *Cell reports* 9.6, pp. 2206–2218.

Chang, L et al. (Mar. 2001). "Mammalian MAP kinase signalling cascades." In: *Nature* 410.6824, pp. 37–40.

Chen, Fang et al. (Feb. 2016). "Transcription factor Ets-1 links glucotoxicity to pancreatic beta cell dysfunction through inhibiting PDX-1 expression in rodent models." In: *Diabetologia* 59.2, pp. 316–324.

Chen, Hainan et al. (Oct. 2011). "PDGF signalling controls age-dependent proliferation in pancreatic [bgr]-cells". In: *Nature* 478.7369, pp. 349–355.

Chen, Ruei-Lin et al. (Apr. 2010). "Developmental silencing of human zeta-globin gene expression is mediated by the transcriptional repressor RREB1." In: *Journal of Biological Chemistry* 285.14, pp. 10189–10197.

Chia, Crystal Y et al. (Jan. 2019). "GATA6 Cooperates with EOMES/SMAD2/3 to Deploy the Gene Regulatory Network Governing Human Definitive Endoderm and Pancreas Formation." In: *Stem cell reports* 12.1, pp. 57–70.

Chimienti, Fabrice et al. (Oct. 2006). "In vivo expression and functional characterization of the zinc transporter ZnT8 in glucose-induced insulin secretion." In: *Journal of cell science* 119.Pt 20, pp. 4199–4206.

Chiou, Joshua et al. (July 2019). "Single cell chromatin accessibility reveals pancreatic islet cell type- and state-specific regulatory programs of diabetes risk". In: *bioRxiv* 71, p. 693671.

Choi, Kyung-Min et al. (June 2008). "Effect of ascorbic acid on bone marrow-derived mesenchymal stem cell proliferation and differentiation." In: *Journal of bioscience and bioengineering* 105.6, pp. 586–594.

Choksi, Semil P et al. (Apr. 2014). "Switching on cilia: transcriptional networks regulating ciliogenesis." In: *Development (Cambridge, England)* 141.7, pp. 1427–1441.

Chong, J A et al. (Mar. 1995). "REST: a mammalian silencer protein that restricts sodium channel gene expression to neurons." In: *Cell* 80.6, pp. 949–957.

Chu, Audrey Y et al. (Dec. 2016). "Multiethnic genome-wide meta-analysis of ectopic fat depots identifies loci associated with adipocyte development and differentiation." In: *Nature Genetics*.

Chung, Mei-I et al. (Mar. 2012). "RFX2 is broadly required for ciliogenesis during vertebrate development." In: *Developmental biology* 363.1, pp. 155–165.

Churchill, Angela J et al. (Jan. 2017). "Genetic evidence that Nkx2.2 acts primarily downstream of Neurog3 in pancreatic endocrine lineage development." In: *eLife* 6, R106.

Collombat, Patrick et al. (Oct. 2003). "Opposing actions of Arx and Pax4 in endocrine pancreas development." In: *Genes & development* 17.20, pp. 2591–2603.

Concepcion, Jennifer P et al. (Feb. 2014). "Neonatal diabetes, gallbladder agenesis, duodenal atresia, and intestinal malrotation caused by a novel homozygous mutation in RFX6." In: *Pediatric diabetes* 15.1, pp. 67–72.

Cong, Le et al. (Feb. 2013). "Multiplex genome engineering using CRISPR/Cas systems." In: *Science* 339.6121, pp. 819–823.

Cooper, Tyler T et al. (Feb. 2016). "The IsletCore Program: Improving the Supply of Human Islets to Satisfy the Demand for Research See article in Endocrinology 2016;157:560–569". In: *Endocrinology* 157.3, pp. 997–1002.

Coore, H G et al. (Oct. 1964). "Regulation of insulin secretion studied with pieces of rabbit pancreas incubated in vitro." In: *Biochemical Journal* 93.1, pp. 66–78.

Corradin, Olivia et al. (Jan. 2014). "Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits." In: *Genome research* 24.1, pp. 1–13.

Cortijo, Cedric et al. (Dec. 2012). "Planar cell polarity controls pancreatic beta cell differentiation and glucose homeostasis." In: *Cell reports* 2.6, pp. 1593–1606.

Cortizo, A et al. (Oct. 1990). "Vectorial insulin secretion by pancreatic beta-cells." In: *FEBS letters* 272.1-2, pp. 137–140.

Costello, Leslie C et al. (2012). "Evidence for Changes in RREB-1, ZIP3, and Zinc in the Early Development of Pancreatic Adenocarcinoma". In: *Journal of Gastrointestinal Cancer* 43.4, pp. 570–578.

Crespin, S R et al. (Aug. 1973). "Stimulation of insulin secretion by long-chain free fatty acids. A direct pancreatic effect." In: *The Journal of clinical investigation* 52.8, pp. 1979–1984.

D'Amour, Kevin A et al. (Dec. 2005). "Efficient differentiation of human embryonic stem cells to definitive endoderm." In: *Nature biotechnology* 23.12, pp. 1534–1541.

D'Amour, Kevin A et al. (Nov. 2006). "Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells." In: *Nature biotechnology* 24.11, pp. 1392–1401.

Da Silva Xavier, Gabriela et al. (2013). "Animal models of GWAS-identified type 2 diabetes genes." In: *Journal of diabetes research* 2013.3, pp. 906590–12.

Dadi, Prasanna K et al. (May 2014). "Inhibition of pancreatic β-cell Ca2+/calmodulin-dependent protein kinase II reduces glucose-stimulated calcium influx and insulin secretion, impairing glucose tolerance." In: *Journal of Biological Chemistry* 289.18, pp. 12435–12445.

Date, Shoichi et al. (May 2004). "Finb, a multiple zinc finger protein, represses transcription of the human angiotensinogen gene". In: *International Journal of Molecular Medicine* 13.5, pp. 637–642.

Davidson, E H (2001). *Genomic Regulatory Systems.*

De Franco, E et al. (May 2013). "Biallelic PDX1 (insulin promoter factor 1) mutations causing neonatal diabetes without exocrine pancreatic insufficiency." In: *Diabetic medicine : a journal of the British Diabetic Association* 30.5, e197–200.

De Vas, Matias G et al. (Mar. 2015). "Hnf1b controls pancreas morphogenesis and the generation of Ngn3+ endocrine progenitors." In: *Development (Cambridge, England)* 142.5, pp. 871–882.

De Vos, A et al. (Nov. 1995). "Human and rat beta cells differ in glucose transporter but not in glucokinase gene expression." In: *The Journal of clinical investigation* 96.5, pp. 2489–2495.

DeFronzo, Ralph A et al. (July 2015). "Type 2 diabetes mellitus." In: *Nature reviews. Disease primers* 1.1, p. 15019.

Delmeire, D et al. (Oct. 2003). "Type VIII adenylyl cyclase in rat beta cells: coincidence signal detector/generator for glucose and GLP-1." In: *Diabetologia* 46.10, pp. 1383–1393.

Dembinski, Artur B et al. (Jan. 1980). "Stimulation of Pancreatic Growth by Secretin, Caerulein, and Pentagastrin*". In: *Endocrinology* 106.1, pp. 323–328.

Didon, Lukas et al. (July 2013). "RFX3 modulation of FOXJ1 regulation of cilia genes in the human airway epithelium." In: *Respiratory research* 14.1, pp. 70–13.

Dimas, Antigone S et al. (June 2014). "Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity." In: *Diabetes* 63.6, pp. 2158–2171.

Dimitriadis, George et al. (Aug. 2011). "Insulin effects in muscle and adipose tissue." In: *Diabetes research and clinical practice* 93 Suppl 1, S52–9.

Dobin, Alexander et al. (Jan. 2013). "STAR: ultrafast universal RNA-seq aligner." In: *Bioinformatics (Oxford, England)* 29.1, pp. 15–21.

Dufort, D et al. (Aug. 1998). "The transcription factor HNF3beta is required in visceral endoderm for normal primitive streak morphogenesis." In: *Development (Cambridge, England)* 125.16, pp. 3015–3025.

Dupuis, Josee et al. (Feb. 2010). "New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk." In: *Nature Genetics* 42.2, pp. 105–116.

Durai, Rajaraman et al. (June 2006). "Biology of insulin-like growth factor binding protein-4 and its role in cancer (review)." In: *International journal of oncology* 28.6, pp. 1317–1325.

Dwivedi, Om Prakash et al. (Nov. 2019). "Loss of ZnT8 function protects against diabetes by enhanced insulin secretion." In: *Nature Genetics* 51.11, pp. 1596–1606.

El Zein, Loubna et al. (Sept. 2009). "RFX3 governs growth and beating efficiency of motile cilia in mouse and controls the expression of genes involved in human ciliopathies." In: *Journal of cell science* 122.Pt 17, pp. 3180–3189.

Ellgaard, L et al. (Dec. 1999). "Setting the standards: quality control in the secretory pathway." In: *Science* 286.5446, pp. 1882–1888.

Elliott, Aaron M et al. (Nov. 2010). "High resolution array-CGH characterization of human stem cells using a stem cell focused microarray." In: *Molecular biotechnology* 46.3, pp. 234–242.

Emery, P et al. (Aug. 1996a). "A consensus motif in the RFX DNA binding domain and binding domain mutants with altered specificity." In: *Molecular and Cellular Biology* 16.8, pp. 4486–4494.

Emery, P et al. (Mar. 1996b). "RFX proteins, a novel family of DNA binding proteins conserved in the eukaryotic kingdom." In: *Nucleic acids research* 24.5, pp. 803–807.

ENCODE Project Consortium (Sept. 2012). "An integrated encyclopedia of DNA elements in the human genome." In: *Nature* 489.7414, pp. 57–74.

Ernst, Jason et al. (Aug. 2010). "Discovery and characterization of chromatin states for systematic annotation of the human genome." In: *Nature biotechnology* 28.8, pp. 817–825.

Evans, M J et al. (July 1981). "Establishment in culture of pluripotential cells from mouse embryos." In: *Nature* 292.5819, pp. 154–156.

Farnsworth, Nikki L et al. (Apr. 2014). "New insights into the role of connexins in pancreatic islet function and diabetes." In: *FEBS letters* 588.8, pp. 1278–1287.

Feng, Jianxing et al. (June 2011). "Using MACS to identify peaks from ChIP-Seq data." In: *Current protocols in bioinformatics* Chapter 2.1, Unit 2.14–2.14.14.

Fiaschi-Taesch, Nathalie et al. (Apr. 2009). "Survey of the human pancreatic beta-cell G1/S proteome reveals a potential therapeutic role for cdk-6 and cyclin D1 in enhancing human beta-cell replication and function in vivo." In: *Diabetes* 58.4, pp. 882–893.

Fisher, J B et al. (July 2017). "GATA6 is essential for endoderm formation from human pluripotent stem cells." In: *Biology open* 6.7, pp. 1084–1095.

Flajollet, Sébastien et al. (Dec. 2009). "RREB-1 is a transcriptional repressor of HLA-G." In: *The Journal of Immunology* 183.11, pp. 6948–6959.

Flanagan, Sarah E et al. (Jan. 2014). "Analysis of transcription factors key for mouse pancreatic development establishes NKX2-2 and MNX1 mutations as causes of neonatal diabetes in man." In: *Cell metabolism* 19.1, pp. 146–154.

Flannick, Jason et al. (Apr. 2014). "Loss-of-function mutations in SLC30A8 protect against type 2 diabetes." In: *Nature Genetics* 46.4, pp. 357–363.

Flannick, Jason et al. (June 2019). "Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls." In: *Nature* 570.7759, pp. 71–76.

Fleischer, Tracey C et al. (May 2003). "Identification and characterization of three new components of the mSin3A corepressor complex." In: *Molecular and Cellular Biology* 23.10, pp. 3456–3467.

Florez, J C (July 2008). "Newly identified loci highlight beta cell dysfunction as a key cause of type 2 diabetes: where are the insulin resistance genes?" In: *Diabetologia* 51.7, pp. 1100–1110.

Florez, J C et al. (May 2004). "Haplotype structure and genotype-phenotype correlations of the sulfonylurea receptor and the islet ATP-sensitive potassium channel gene region." In: *Diabetes* 53.5, pp. 1360–1368.

Floyd, J C (Sept. 1980). "Pancreatic polypeptide." In: *Clinics in gastroenterology* 9.3, pp. 657–678.

Fogarty, Marie P et al. (Sept. 2014). "Identification of a regulatory variant that binds FOXA1 and FOXA2 at the CDC123/CAMK1D type 2 diabetes GWAS locus." In: *PLoS genetics* 10.9, e1004633.

Friedrichsen, Birgitte N et al. (Mar. 2006). "Stimulation of pancreatic beta-cell replication by incretins involves transcriptional induction of cyclin D1 via multiple signalling pathways." In: *The Journal of endocrinology* 188.3, pp. 481–492.

Frock, Richard L et al. (Feb. 2015). "Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases." In: *Nature biotechnology* 33.2, pp. 179–186.

Fuchsberger, Christian et al. (July 2016). "The genetic architecture of type 2 diabetes." In: *Nature*.

Fujimoto-Nishiyama, Akiko et al. (Aug. 1997). "A novel zinc finger protein, Finb, is a transcriptional activator and localized in nuclear bodies". In: *Gene* 195.2, pp. 267–275.

Gannon, Maureen et al. (Feb. 2008). "pdx-1 function is specifically required in embryonic beta cells to generate appropriate numbers of endocrine cell types and maintain glucose homeostasis." In: *Developmental biology* 314.2, pp. 406–417.

Gao, Nan et al. (Dec. 2008). "Dynamic regulation of Pdx1 enhancers by Foxa1 and Foxa2 is essential for pancreas development." In: *Genes & development* 22.24, pp. 3435–3448.

Gao, Tao et al. (Feb. 2014). "Pdx1 maintains β cell identity and function by repressing an α cell program." In: *Cell metabolism* 19.2, pp. 259–271.

Garcia-Gonzalez, Miguel A et al. (Sept. 2016). "A suppressor locus for MODY3-diabetes." In: *Scientific reports* 6.1, p. 33087.

Garneau, Josiane E et al. (Nov. 2010). "The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA." In: *Nature* 468.7320, pp. 67–71.

Garriga-Canut, Mireia et al. (Nov. 2006). "2-Deoxy-D-glucose reduces epilepsy progression by NRSF-CtBP-dependent metabolic regulation of chromatin structure." In: *Nature neuroscience* 9.11, pp. 1382–1387.

Gasa, R et al. (Apr. 2008). "Induction of pancreatic islet cell differentiation by the neurogenin-neuroD cascade." In: *Differentiation; research in biological diversity* 76.4, pp. 381–391.

Gasiunas, Giedrius et al. (Sept. 2012). "Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria." In: *Proceedings of the National Academy of Sciences of the United States of America* 109.39, E2579–86.

Gaulton, Kyle J et al. (Mar. 2010). "A map of open chromatin in human pancreatic islets." In: *Nature Genetics* 42.3, pp. 255–259.

Gaulton, Kyle J et al. (Dec. 2015). "Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci." In: *Nature Genetics* 47.12, pp. 1415–1425.

Gavin, J R et al. (Jan. 2003). *Report of the expert committee on the diagnosis and classification of diabetes mellitus.* American Diabetes Association.

George, Nicholas M et al. (Nov. 2015). "Exploiting Expression of Hippo Effector, Yap, for Expansion of Functional Islet Mass." In: *Molecular endocrinology (Baltimore, Md.)* 29.11, pp. 1594–1607.

Gerich, J E (July 1993). "Control of glycaemia." In: *Bailliere's clinical endocrinology and metabolism* 7.3, pp. 551–586.

German-Diaz, Marta et al. (Aug. 2017). "A New Case of Congenital Malabsorptive Diarrhea and Diabetes Secondary to Mutant Neurogenin-3." In: *Pediatrics* 140.2, e20162210.

Gerrish, K et al. (Feb. 2000). "Pancreatic beta cell-specific transcription of the pdx-1 gene. The role of conserved upstream control regions and their hepatic nuclear factor 3beta sites." In: *Journal of Biological Chemistry* 275.5, pp. 3485–3492.

Gheni, Ghupurjan et al. (Oct. 2014). "Glutamate acts as a key signal linking glucose metabolism to incretin/cAMP action to amplify insulin secretion." In: *Cell reports* 9.2, pp. 661–673.

Giresi, Paul G et al. (June 2007). "FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin." In: *Genome research* 17.6, pp. 877–885.

Glaser, Benjamin et al. (June 1988). "Effects of Secretin on the Normal and Pathological β-Cell*". In: *The Journal of Clinical Endocrinology & Metabolism* 66.6, pp. 1138–1143.

Glick, E et al. (Jan. 2000). "Transcription factor BETA2 acts cooperatively with E2A and PDX1 to activate the insulin gene promoter." In: *Journal of Biological Chemistry* 275.3, pp. 2199–2204.

Gloyn, Anna L et al. (Feb. 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." In: *Diabetes* 52.2, pp. 568–572.

Goodge, K A et al. (Aug. 2000). "Translational regulation of proinsulin biosynthesis and proinsulin conversion in the pancreatic beta-cell." In: *Seminars in cell & developmental biology* 11.4, pp. 235–242.

Gosmain, Yvan et al. (Apr. 2012). "Pax6 is crucial for β-cell function, insulin biosynthesis, and glucose-induced insulin secretion." In: *Molecular endocrinology (Baltimore, Md.)* 26.4, pp. 696–709.

Gradwohl, G et al. (Feb. 2000). "neurogenin3 is required for the development of the four endocrine cell lineages of the pancreas." In: *Proceedings of the National Academy of Sciences* 97.4, pp. 1607–1611.

Greenwald, William W et al. (May 2019). "Pancreatic islet chromatin accessibility and conformation reveals distal enhancer networks of type 2 diabetes risk." In: *Nature communications* 10.1, pp. 2078–12.

Grimes, J A et al. (Mar. 2000). "The co-repressor mSin3A is a functional component of the REST-CoREST repressor complex." In: *Journal of Biological Chemistry* 275.13, pp. 9461–9467.

Grobarczyk, Benjamin et al. (Oct. 2015). "Generation of Isogenic Human iPS Cell Line Precisely Corrected by Genome Editing Using the CRISPR/Cas9 System." In: *Stem cell reviews* 11.5, pp. 774–787.

Gromada, Jesper et al. (Feb. 2007). "Alpha-cells of the endocrine pancreas: 35 years of research but the enigma remains." In: *Endocrine reviews* 28.1, pp. 84–116.

Grotz, A K et al. (Oct. 2019). "A CRISPR/Cas9 genome editing pipeline in the EndoC-βH1 cell line to study genes implicated in beta cell function". In: *Wellcome Open Research* 4, p. 150.

GTEx Consortium (May 2015). "Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans." In: *Science* 348.6235, pp. 648–660.

Gu, Chunyan et al. (Apr. 2010). "Pancreatic beta cells require NeuroD to achieve and maintain functional maturity." In: *Cell metabolism* 11.4, pp. 298–310.

Gu, Guoqiang et al. (May 2002). "Direct evidence for the pancreatic lineage: NGN3+ cells are islet progenitors and are distinct from duct progenitors." In: *Development (Cambridge, England)* 129.10, pp. 2447–2457.

Gu, Guoqiang et al. (Jan. 2004). "Global expression analysis of gene regulatory pathways during endocrine pancreatic development." In: *Development (Cambridge, England)* 131.1, pp. 165–179.

Guo, Min et al. (May 2017). "Using hESCs to Probe the Interaction of the Diabetes-Associated Genes CDKAL1 and MT1E." In: *Cell reports* 19.8, pp. 1512–1521.

Guo, Ting et al. (2011). "ISL1 promotes pancreatic islet cell proliferation." In: *PloS one* 6.8, e22387.

Guo, Xin et al. (Aug. 2012). "Glycolysis in the control of blood glucose homeostasis". In: *Acta Pharmaceutica Sinica B* 2.4, pp. 358–367.

Gupta, Shailesh Kumar et al. (Apr. 2018). "NKX6.1 induced pluripotent stem cell reporter lines for isolation and analysis of functionally relevant neuronal and pancreas populations." In: *Stem cell research* 29, pp. 220–231.

Gutierrez-Aguilar, Ruth et al. (Feb. 2014). "The role of the transcription factor ETV5 in insulin exocytosis." In: *Diabetologia* 57.2, pp. 383–391.

Haase, T N et al. (Apr. 2013). "Growth arrest specific protein (GAS) 6: a role in the regulation of proliferation and functional capacity of the perinatal rat beta cell." In: *Diabetologia* 56.4, pp. 763–773.

Haefliger, Jacques-Antoine et al. (Apr. 2003). "The scaffold protein IB1/JIP-1 is a critical mediator of cytokine-induced apoptosis in pancreatic beta cells." In: *Journal of cell science* 116.Pt 8, pp. 1463–1469.

Haeussler, Maximilian et al. (July 2016). "Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR." In: *Genome biology* 17.1, p. 148.

Hald, J et al. (Jan. 2012). "Pancreatic islet and progenitor cell surface markers with cell sorting potential." In: *Diabetologia* 55.1, pp. 154–165.

Haliyur, Rachana et al. (Jan. 2019). "Human islets expressing HNF1A variant have defective β cell transcriptional regulatory networks." In: *The Journal of clinical investigation* 129.1, pp. 246–251.

Hancili, Suna et al. (May 2018). "A novel NEUROG3 mutation in neonatal diabetes associated with a neuro-intestinal syndrome." In: *Pediatric diabetes* 19.3, pp. 381–387.

Hang, Yan et al. (Sept. 2011). "MafA and MafB activity in pancreatic β cells." In: *Trends in endocrinology and metabolism: TEM* 22.9, pp. 364–373.

Hang, Yan et al. (June 2014). "The MafA transcription factor becomes essential to islet β-cells soon after birth." In: *Diabetes* 63.6, pp. 1994–2005.

Hani, E H et al. (Nov. 1999). "Defective mutations in the insulin promoter factor-1 (IPF-1) gene in late-onset type 2 diabetes mellitus." In: *The Journal of clinical investigation* 104.9, R41–8.

Hansson, Mattias et al. (Oct. 2004). "Artifactual insulin release from differentiated embryonic stem cells." In: *Diabetes* 53.10, pp. 2603–2609.

Harmon, Jamie S et al. (Mar. 2005). "Oxidative stress-mediated, post-translational loss of MafA protein as a contributing mechanism to loss of insulin gene expression in glucotoxic beta cells." In: *Journal of Biological Chemistry* 280.12, pp. 11107–11113.

Harmon, Jamie S et al. (Nov. 2009). "beta-Cell-specific overexpression of glutathione peroxidase preserves intranuclear MafA and reverses diabetes in db/db mice." In: *Endocrinology* 150.11, pp. 4855–4862.

Harrow, Jennifer et al. (Sept. 2012). "GENCODE: the reference human genome annotation for The ENCODE Project." In: *Genome research* 22.9, pp. 1760–1774.

Hart, Nathaniel J et al. (Feb. 2019). "Use of human islets to understand islet biology and diabetes: progress, challenges and suggestions." In: *Diabetologia* 62.2, pp. 212–222.

Hart, Traver et al. (Aug. 2017). "Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens". In: *G3 (Bethesda, Md.)* 7.8, pp. 2719–2727.

Hastoy, Benoit et al. (Nov. 2018). "Electrophysiological properties of human beta-cell lines EndoC-βH1 and -βH2 conform with human beta-cells." In: *Scientific reports* 8.1, p. 16994.

Haumaitre, C et al. (Feb. 2005). "Lack of TCF2/vHNF1 in mice leads to pancreas agenesis." In: *Proceedings of the National Academy of Sciences* 102.5, pp. 1490–1495.

Hay, Colin W et al. (Dec. 2006). "Comparative analysis of insulin gene promoters: implications for diabetes research." In: *Diabetes* 55.12, pp. 3201–3213.

Head, W Steven et al. (July 2012). "Connexin-36 gap junctions regulate in vivo first- and second-phase insulin secretion dynamics and glucose tolerance in the conscious mouse." In: *Diabetes* 61.7, pp. 1700–1707.

Heintzman, Nathaniel D et al. (May 2009). "Histone modifications at human enhancers reflect global cell-type-specific gene expression." In: *Nature* 459.7243, pp. 108–112.

Henquin, J C et al. (Mar. 1994). "Multisite control of insulin release by glucose." In: *Diabete & metabolisme* 20.2, pp. 132–137.

Henquin, Jean-Claude et al. (Dec. 2006). "Nutrient control of insulin secretion in isolated normal human islets." In: *Diabetes* 55.12, pp. 3470–3477.

Herman, William H (Apr. 2013). "The economic costs of diabetes: is it time for a new treatment paradigm?" In: *Diabetes care* 36.4, pp. 775–776.

Herrera, P L (June 2000). "Adult insulin- and glucagon-producing cells differentiate from two independent cell lineages." In: *Development (Cambridge, England)* 127.11, pp. 2317–2322.

Hilling, Denise E et al. (2014). "Effects of donor-, pancreas-, and isolation-related variables on human islet isolation outcome: a systematic review." In: *Cell transplantation* 23.8, pp. 921–928.

Hindorff, Lucia A et al. (June 2009). "Potential etiologic and functional implications of genome-wide association loci for human diseases and traits." In: *Proceedings of the National Academy of Sciences of the United States of America* 106.23, pp. 9362–9367.

Hnisz, Denes et al. (Nov. 2013). "Super-enhancers in the control of cell identity and disease." In: *Cell* 155.4, pp. 934–947.

Hockemeyer, Dirk et al. (July 2011). "Genetic engineering of human pluripotent cells using TALE nucleases." In: *Nature biotechnology* 29.8, pp. 731–734.

Hodson, David J et al. (June 2014). "Incretin-modulated beta cell energetics in intact islets of Langerhans." In: *Molecular endocrinology (Baltimore, Md.)* 28.6, pp. 860–871.

Hohl, Mathias et al. (Nov. 2005). "Cell type-specific regulation of RE-1 silencing transcription factor (REST) target genes." In: *The European journal of neuroscience* 22.9, pp. 2216–2230.

Horikoshi, Momoko et al. (May 2016). "Transancestral fine-mapping of four type 2 diabetes susceptibility loci highlights potential causal regulatory mechanisms." In: *Human Molecular Genetics* 25.10, pp. 2070–2081.

Hornstein, Eran et al. (June 2006). "Canalization of development by microRNAs." In: *Nature Genetics* 38 Suppl.S6, S20–4.

Horvath, Gary C et al. (Nov. 2004). "RFX2 is a potential transcriptional regulatory factor for histone H1t and other genes expressed during the meiotic phase of spermatogenesis." In: *Biology of reproduction* 71.5, pp. 1551–1559.

Horvath, Gary C et al. (Dec. 2009). "RFX2 is a candidate downstream amplifier of A-MYB regulation in mouse spermatogenesis." In: *BMC developmental biology* 9.1, p. 63.

Hrvatin, Sinisa et al. (Feb. 2014). "Differentiated human stem cells resemble fetal, not adult, β cells." In: *Proceedings of the National Academy of Sciences of the United States of America* 111.8, pp. 3038–3043.

Hu, Frank B (June 2011). "Globalization of Diabetes". In: *Diabetes care* 34.6, pp. 1249–1257.

Hua, Haiqing et al. (July 2013). "iPSC-derived β cells model diabetes due to glucokinase deficiency." In: *The Journal of clinical investigation* 123.7, pp. 3146–3153.

Huang, H P et al. (May 2000). "Regulation of the pancreatic islet-specific gene BETA2 (neuroD) by neurogenin 3." In: *Molecular and Cellular Biology* 20.9, pp. 3292–3307.

IDF (2017). *IDF Diabetes Atlas, 8th edn.*

Ihm, Sung-Hee et al. (May 2006). "Effect of donor age on function of isolated human islets." In: *Diabetes* 55.5, pp. 1361–1368.

Ihry, Robert J et al. (June 2018). "p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells." In: *Nature medicine* 337, p. 816.

Ingelsson, Erik et al. (May 2010). "Detailed physiologic characterization reveals diverse mechanisms for novel genetic Loci regulating glucose and insulin metabolism in humans." In: *Diabetes* 59.5, pp. 1266–1275.

Insel, Paul A et al. (June 2003). "Forskolin as a tool for examining adenylyl cyclase expression, regulation, and G protein signaling." In: *Cellular and molecular neurobiology* 23.3, pp. 305–314.

International Stem Cell Initiative et al. (Nov. 2011). "Screening ethnically diverse human embryonic stem cells identifies a chromosome 20 minimal amplicon conferring growth advantage." In: *Nature biotechnology* 29.12, pp. 1132–1144.

Iwama, A et al. (June 1999). "Dimeric RFX proteins contribute to the activity and lineage specificity of the interleukin-5 receptor alpha promoter through activation and repression domains." In: *Molecular and Cellular Biology* 19.6, pp. 3940–3950.

Jacquemin, P et al. (June 2000). "Transcription factor hepatocyte nuclear factor 6 regulates pancreatic endocrine cell differentiation and controls expression of the proendocrine gene ngn3." In: *Molecular and Cellular Biology* 20.12, pp. 4445–4454.

Janky, Rekin's et al. (July 2014). "iRegulon: from a gene list to a gene regulatory network using large motif and track collections." In: *PLoS computational biology* 10.7, e1003731.

Javierre, Biola M et al. (Nov. 2016). "Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters." In: *Cell* 167.5, 1369–1384.e19.

Jennings, Rachel E et al. (Oct. 2013). "Development of the human pancreas from foregut to endocrine commitment." In: *Diabetes* 62.10, pp. 3514–3522.

Jennings, Rachel E et al. (Sept. 2015). "Human pancreas development." In: *Development (Cambridge, England)* 142.18, pp. 3126–3137.

Jensen, J et al. (Jan. 2000). "Control of endodermal endocrine development by Hes-1." In: *Nature Genetics* 24.1, pp. 36–44.

Jeon, Jongmin et al. (Sept. 2009). "Endocrine cell clustering during human pancreas development." In: *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* 57.9, pp. 811–824.

Jinek, Martin et al. (Aug. 2012). "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." In: *Science* 337.6096, pp. 816–821.

Johansson, Kerstin A et al. (Mar. 2007). "Temporal control of neurogenin3 activity in pancreas progenitors reveals competence windows for the generation of different endocrine cell types." In: *Developmental cell* 12.3, pp. 457–465.

Jolma, Arttu et al. (June 2010). "Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities." In: *Genome research* 20.6, pp. 861–873.

Jolma, Arttu et al. (Jan. 2013). "DNA-binding specificities of human transcription factors." In: *Cell* 152.1-2, pp. 327–339.

Jonsson, J et al. (Oct. 1994). "Insulin-promoter-factor 1 is required for pancreas development in mice." In: *Nature* 371.6498, pp. 606–609.

Jørgensen, Mette Christine et al. (Oct. 2007). "An illustrated review of early pancreas development in the mouse." In: *Endocrine reviews* 28.6, pp. 685–705.

Kaddis, John S et al. (Apr. 2009). "Human pancreatic islets and diabetes research." In: *JAMA* 301.15, pp. 1580–1587.

Kaihara, Kelly A et al. (May 2013). "β-Cell-specific protein kinase A activation enhances the efficiency of glucose control by increasing acute-phase insulin secretion." In: *Diabetes* 62.5, pp. 1527–1536.

Kambal, Mohammed Abdulmageed et al. (2019). "Mitchell-Riley Syndrome Due to a Novel Mutation in RFX6". In: *Frontiers in Pediatrics* 7, p. 2160.

Kaprio, J et al. (Nov. 1992). "Concordance for type 1 (insulin-dependent) and type 2 (non-insulin-dependent) diabetes mellitus in a population-based cohort of twins in Finland." In: *Diabetologia* 35.11, pp. 1060–1067.

Kato, Masato et al. (Nov. 2007). "The novel endocytic and phagocytic C-Type lectin receptor DCL-1/CD302 on macrophages is colocalized with F-actin, suggesting a role in cell adhesion and migration." In: *The Journal of Immunology* 179.9, pp. 6052–6063.

Kaufman, R J et al. (Oct. 2010). "The unfolded protein response is required to maintain the integrity of the endoplasmic reticulum, prevent oxidative stress and preserve differentiation in β-cells." In: *Diabetes, Obesity and Metabolism* 12 Suppl 2, pp. 99–107.

Kaufmann, J E et al. (Sept. 1995). "Sequence requirements for proinsulin processing at the B-chain/C-peptide junction." In: *Biochemical Journal* 310 (Pt 3).3, pp. 869–874.

Kent, O A et al. (Dec. 2010). "Repression of the miR-143/145 cluster by oncogenic Ras initiates a tumor-promoting feed-forward pathway." In: *Genes & development* 24.24, pp. 2754–2759.

Kent, O A et al. (May 2013). "RREB1 repressed miR-143/145 modulates KRAS signaling through downregulation of multiple targets." In: *Oncogene* 32.20, pp. 2576–2585.

Kent, O A et al. (Nov. 2016). "An oncogenic KRAS transcription program activates the RHOGEF ARHGEF2 to mediate transformed phenotypes in pancreatic cancer." In: *Oncotarget* 5.0.

Kim, Daesik et al. (Mar. 2015). "Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells." In: *Nature methods* 12.3, 237–43–1 p following 243.

Kistler, W Stephen et al. (Oct. 2009). "Differential expression of Rfx1-4 during mouse spermatogenesis." In: *Gene expression patterns : GEP* 9.7, pp. 515–519.

Klee, Philippe et al. (Dec. 2011). "Connexins protect mouse pancreatic β cells against apoptosis." In: *The Journal of clinical investigation* 121.12, pp. 4870–4879.

Kloosterman, Wigard P et al. (Oct. 2006). "The diverse functions of microRNAs in animal development and disease." In: *Developmental cell* 11.4, pp. 441–450.

Koenigsberger, C et al. (Feb. 2000). "Differential regulation by multiple promoters of the gene encoding the neuron-restrictive silencer factor." In: *Proceedings of the National Academy of Sciences* 97.5, pp. 2291–2296.

Kolch, W (Oct. 2000). "Meaningful relationships: the regulation of the Ras/Raf/MEK/ERK pathway by protein interactions." In: *Biochemical Journal* 351 Pt 2.Pt 2, pp. 289–305.

Kraegen, E W et al. (Mar. 1970). "The gastrointestinal stimulus to insulin release. II. A dual action of secretin." In: *The Journal of clinical investigation* 49.3, pp. 524–529.

Kroon, Evert et al. (Apr. 2008). "Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells in vivo." In: *Nature biotechnology* 26.4, pp. 443–452.

Kulkarni, Rohit N et al. (Sept. 2012). "Human β-cell proliferation and intracellular signaling: driving in the dark without a road map." In: *Diabetes* 61.9, pp. 2205–2213.

Kulkarni, Rohit N et al. (Dec. 2014). "Summary of the Keystone islet workshop (April 2014): the increasing demand for human islet availability in diabetes research." In: *Diabetes* 63.12, pp. 3979–3981.

Kulzer, Jennifer R et al. (Feb. 2014). "A common functional regulatory variant at a type 2 diabetes locus upregulates ARAP1 expression in the pancreatic beta cell." In: *American journal of human genetics* 94.2, pp. 186–197.

Kutlu, Burak et al. (Jan. 2009). "Detailed transcriptome atlas of the pancreatic beta cell." In: *BMC medical genomics* 2.1, p. 3.

Kycia, Ina et al. (Apr. 2018). "A Common Type 2 Diabetes Risk Variant Potentiates Activity of an Evolutionarily Conserved Islet Stretch Enhancer and Increases C2CD4A and C2CD4B Expression". In: *The American Journal of Human Genetics* 102.4, pp. 620–635.

Kyttälä, Aija et al. (Feb. 2016). "Genetic Variability Overrides the Impact of Parental Cell Type and Determines iPSC Differentiation Potential." In: *Stem cell reports* 6.2, pp. 200–212.

Landt, Stephen G et al. (Sept. 2012). "ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia." In: *Genome research* 22.9, pp. 1813–1831.

Langfelder, Peter et al. (Dec. 2008). "WGCNA: an R package for weighted correlation network analysis." In: *BMC bioinformatics* 9.1, p. 559.

Larsen, Hjalte List et al. (June 2017). "The molecular and morphogenetic basis of pancreas organogenesis." In: *Seminars in cell & developmental biology* 66, pp. 51–68.

Latchman, D (1995). *Gene regulation: a eukaryotic perspective.*

Lau, Hwee Hui et al. (May 2018). "The molecular functions of hepatocyte nuclear factors - In and beyond the liver." In: *Journal of hepatology* 68.5, pp. 1033–1048.

Laurent, Louise C et al. (Jan. 2011). "Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture." In: *Cell stem cell* 8.1, pp. 106–118.

Lawlor, Nathan et al. (Feb. 2017). "Single-cell transcriptomes identify human islet cell signatures and reveal cell-type-specific expression changes in type 2 diabetes." In: *Genome research* 27.2, pp. 208–222.

Lawlor, Nathan et al. (Jan. 2019). "Multiomic Profiling Identifies cis-Regulatory Networks Underlying Human Pancreatic β Cell Identity and Function." In: *Cell reports* 26.3, 788–801.e6.

Le Gurun, Sabine et al. (Sept. 2003). "Connexin-36 contributes to control function of insulin-producing cells." In: *Journal of Biological Chemistry* 278.39, pp. 37690–37697.

Lee, Catherine S et al. (Feb. 2005). "Foxa2 is required for the differentiation of pancreatic alpha-cells." In: *Developmental biology* 278.2, pp. 484–495.

Lee, Dong Hyeon et al. (Jan. 2012). "Proteomic Identification of RREB1, PDE6B, and CD209 Up-Regulated in Primitive Gut Tube Differentiated From Human Embryonic Stem Cells". In: *Pancreas* 41.1, pp. 65–73.

Lee, J C et al. (May 2001). "Regulation of the pancreatic pro-endocrine gene neurogenin3." In: *Diabetes* 50.5, pp. 928–936.

Lee, Kihyun et al. (July 2019). "FOXA2 Is Required for Enhancer Priming during Pancreatic Differentiation." In: *Cell reports* 28.2, 382–393.e7.

Lek, Monkol et al. (Aug. 2016). "Analysis of protein-coding genetic variation in 60,706 humans." In: *Nature* 536.7616, pp. 285–291.

Lemaire, K et al. (Sept. 2009). "Insulin crystallization depends on zinc transporter ZnT8 expression, but is not required for normal glucose homeostasis in mice." In: *Proceedings of the National Academy of Sciences of the United States of America* 106.35, pp. 14872–14877.

Levine, Ariel J et al. (Jan. 2006). "GDF3, a BMP inhibitor, regulates cell fate in stem cells and early embryos." In: *Development (Cambridge, England)* 133.2, pp. 209–216.

Li, Chien et al. (Mar. 2007). "Urocortin 3 regulates glucose-stimulated insulin secretion and energy homeostasis." In: *Proceedings of the National Academy of Sciences* 104.10, pp. 4206–4211.

Li, En (Sept. 2002). "Chromatin modification and epigenetic reprogramming in mammalian development." In: *Nature reviews. Genetics* 3.9, pp. 662–673.

Li, Heng et al. (Aug. 2009). "The Sequence Alignment/Map format and SAMtools." In: *Bioinformatics (Oxford, England)* 25.16, pp. 2078–2079.

Liao, Yang et al. (Apr. 2014). "featureCounts: an efficient general purpose program for assigning sequence reads to genomic features." In: *Bioinformatics (Oxford, England)* 30.7, pp. 923–930.

Lindner, T H et al. (Oct. 1999). "A novel syndrome of diabetes mellitus, renal dysfunction and genital malformation associated with a partial deletion of the pseudo-POU domain of hepatocyte nuclear factor-1beta." In: *Human Molecular Genetics* 8.11, pp. 2001–2008.

Liu, Hanshao et al. (Aug. 2009). "DNA damage signalling recruits RREB-1 to the p53 tumour suppressor promoter." In: *The Biochemical journal* 422.3, pp. 543–551.

Liu, Jennifer S E et al. (Feb. 2017). "All mixed up: defining roles for β-cell subtypes in mature islets." In: *Genes & development* 31.3, pp. 228–240.

Liu, Ming et al. (2014). "Proinsulin entry and transit through the endoplasmic reticulum in pancreatic beta cells." In: *Vitamins and hormones* 95, pp. 35–62.

Lonovics, J et al. (Oct. 1981). "Pancreatic polypeptide. A review." In: *Archives of surgery (Chicago, Ill. : 1960)* 116.10, pp. 1256–1264.

Love, Michael I et al. (Dec. 2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2". In: *Genome biology* 15.12, p. 550.

Love, Michael I et al. (2016). *Differential analysis of count data – the DESeq2 package.*

Luo, Yan et al. (2014). "Transcription factor Ets1 regulates expression of thioredoxin-interacting protein and inhibits insulin secretion in pancreatic β-cells." In: *PloS one* 9.6, e99049.

Lynn, F C et al. (June 2007). "Sox9 coordinates a transcriptional network in pancreatic progenitor cells." In: *Proceedings of the National Academy of Sciences* 104.25, pp. 10500–10505.

Lyon, James et al. (Dec. 2015). "Research-Focused Isolation of Human Islets From Donors With and Without Diabetes at the Alberta Diabetes Institute IsletCore". In: *Endocrinology* 157.2, pp. 560–569.

Lyttle, B M et al. (July 2008). "Transcription factor expression in the developing human fetal endocrine pancreas." In: *Diabetologia* 51.7, pp. 1169–1180.

Ma, Shisong et al. (July 2017). "Discovery of Novel Human Gene Regulatory Modules from Gene Co-expression and Promoter Motif Analysis." In: *Scientific reports* 7.1, p. 5557.

Macfarlane, W M et al. (Nov. 1999). "Missense mutations in the insulin promoter factor-1 gene predispose to type 2 diabetes." In: *The Journal of clinical investigation* 104.9, R33–9.

Magnani, Dario et al. (May 2015). "The ciliogenic transcription factor Rfx3 is required for the formation of the thalamocortical tract by regulating the patterning of prethalamus and ventral telencephalon." In: *Human Molecular Genetics* 24.9, pp. 2578–2593.

Mahajan, Anubha et al. (Mar. 2014). "Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility". In: *Nature Genetics* 46.3, pp. 234–244.

Mahajan, Anubha et al. (Jan. 2015). "Identification and functional characterization of G6PC2 coding variants influencing glycemic traits define an effector transcript at the G6PC2-ABCB11 locus." In: 11.1, e1004876.

Mahajan, Anubha et al. (Oct. 2018a). "Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps." In: *Nature Genetics* 66, p. 2888.

Mahajan, Anubha et al. (Apr. 2018b). "Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes." In: *Nature Genetics* 50.4, pp. 559–571.

Maijgren, Susann et al. (Apr. 2004). "Involvement of RFX proteins in transcriptional activation from a Ras-responsive enhancer element." In: *Archives of dermatological research* 295.11, pp. 482–489.

Maitra, Anirban et al. (Oct. 2005). "Genomic alterations in cultured human embryonic stem cells." In: *Nature Genetics* 37.10, pp. 1099–1103.

Malaisse, W J et al. (July 1983). "The stimulus-secretion coupling of glucose-induced insulin release: fuel metabolism in islets deprived of exogenous nutrient." In: *Archives of biochemistry and biophysics* 224.1, pp. 102–110.

Mali, Prashant et al. (Feb. 2013). "RNA-guided human genome engineering via Cas9." In: *Science* 339.6121, pp. 823–826.

Maller, Julian B et al. (Dec. 2012). "Bayesian refinement of association signals for 14 loci in 3 common diseases." In: *Nature Genetics* 44.12, pp. 1294–1301.

Manduchi, Elisabetta et al. (July 2018). “A High Resolution Capture-C Promoter "Interactome" Implicates Causal Genes at Type 2 Diabetes GWAS Loci”. In: *Diabetes* 67.Supplement 1, 1705–P.

Marbach, Daniel et al. (Apr. 2016). “Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases.” In: *Nature methods* 13.4, pp. 366–370.

Martin, David et al. (Dec. 2003). “Critical role of the transcriptional repressor neuron-restrictive silencer factor in the specific control of connexin36 in insulin-producing cell lines.” In: *Journal of Biological Chemistry* 278.52, pp. 53082–53089.

Martin, David et al. (Aug. 2008). “Functional significance of repressor element 1 silencing transcription factor (REST) target genes in pancreatic beta cells.” In: *Diabetologia* 51.8, pp. 1429–1439.

Martin, David et al. (2012). “Specific silencing of the REST target genes in insulin-secreting cells uncovers their participation in beta cell survival.” In: *PloS one* 7.9, e45844.

Martin, David et al. (Sept. 2015). “REST represses a subset of the pancreatic endocrine differentiation program.” In: *Developmental biology* 405.2, pp. 316–327.

Martin, David et al. (2017). “The Importance of REST for Development and Function of Beta Cells.” In: *Frontiers in cell and developmental biology* 5.Pt 8, p. 12.

Martin, G R (Dec. 1981). “Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells.” In: *Proceedings of the National Academy of Sciences* 78.12, pp. 7634–7638.

Martins-Taylor, Kristen et al. (June 2011). “Recurrent copy number variations in human induced pluripotent stem cells.” In: *Nature biotechnology* 29.6, pp. 488–491.

Mastracci, Teresa L et al. (2013). “Regulation of Neurod1 contributes to the lineage potential of Neurogenin3+ endocrine precursor cells in the pancreas.” In: *PLoS genetics* 9.2, e1003278.

Mathelier, Anthony et al. (Jan. 2014). “JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles.” In: *Nucleic acids research* 42.Database issue, pp. D142–7.

Matys, V et al. (Jan. 2006). “TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes.” In: *Nucleic acids research* 34.Database issue, pp. D108–10.

Maurano, Matthew T et al. (Sept. 2012). “Systematic localization of common disease-associated variation in regulatory DNA.” In: *Science* 337.6099, pp. 1190–1195.

McCulloch, Laura J et al. (Dec. 2011). “GLUT2 (SLC2A2) is not the principal glucose transporter in human pancreatic beta cells: implications for understanding genetic association signals at this locus.” In: *Molecular genetics and metabolism* 104.4, pp. 648–653.

McGrath, K E et al. (Sept. 1999). “Embryonic expression and function of the chemokine SDF-1 and its receptor, CXCR4.” In: *Developmental biology* 213.2, pp. 442–456.

McGrath, Patrick S et al. (July 2015). “The Basic Helix-Loop-Helix Transcription Factor NEUROG3 Is Required for Development of the Human Endocrine Pancreas.” In: *Diabetes* 64.7, pp. 2497–2505.

McLean, Amanda B et al. (Jan. 2007). “Activin a efficiently specifies definitive endoderm from human embryonic stem cells only when phosphatidylinositol 3-kinase signaling is suppressed.” In: *Stem cells (Dayton, Ohio)* 25.1, pp. 29–38.

Merkle, Florian T et al. (Apr. 2017). "Human pluripotent stem cells recurrently acquire and expand dominant negative P53 mutations." In: *Nature* 29.7653, pp. 1132–233.

Miettinen, P J et al. (June 2000). "Impaired migration and delayed differentiation of pancreatic islet cells in mice lacking EGF-receptors." In: *Development (Cambridge, England)* 127.12, pp. 2617–2627.

Miguel-Escalada, Irene et al. (June 2019). "Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes." In: *Nature Genetics* 51.7, pp. 1137–1148.

Milon, Beatrice C et al. (Feb. 2010). "Ras responsive element binding protein-1 (RREB-1) down-regulates hZIP1 expression in prostate cancer cells". In: *The Prostate* 70.3, pp. 288–296.

Mitchell, J et al. (Dec. 2004). "Neonatal diabetes, with hypoplastic pancreas, intestinal atresia and gall bladder hypoplasia: search for the aetiology of a new autosomal recessive syndrome." In: *Diabetologia* 47.12, pp. 2160–2167.

Mojica, Francisco J M et al. (Feb. 2005). "Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements." In: *Journal of molecular evolution* 60.2, pp. 174–182.

Morris, Andrew P (Feb. 2018). "Progress in defining the genetic contribution to type 2 diabetes susceptibility." In: *Current opinion in genetics & development* 50, pp. 41–51.

Morris, Andrew P et al. (Sept. 2012). "Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes." In: *Nature Genetics* 44.9, pp. 981–990.

Mortazavi, Ali et al. (Oct. 2006). "Comparative genomics modeling of the NRSF/REST repressor network: from single conserved sites to genome-wide repertoire." In: *Genome research* 16.10, pp. 1208–1221.

Mosedale, Merrie et al. (Feb. 2012). "Neurexin-1α contributes to insulin-containing secretory granule docking." In: *Journal of Biological Chemistry* 287.9, pp. 6350–6361.

Mukhopadhyay, Nishit K et al. (Sept. 2007). "The Zinc Finger Protein Ras-Responsive Element Binding Protein-1 Is a Coregulator of the Androgen Receptor: Implications for the Role of the Ras Pathway in Enhancing Androgenic Signaling in Prostate Cancer". In: *Molecular endocrinology (Baltimore, Md.)* 21.9, pp. 2056–2070.

Mularoni, Loris et al. (2017). "The Pancreatic Islet Regulome Browser." In: *Frontiers in Genetics* 8, p. 13.

Murai, Kiyohito et al. (2004). "Direct interaction of NRSF with TBP: chromatin reorganization and core promoter repression for neuron-specific gene transcription." In: *Nucleic acids research* 32.10, pp. 3180–3189.

Murao, Naoya et al. (2017). "Essential roles of aspartate aminotransferase 1 and vesicular glutamate transporters in β-cell glutamate signaling for incretin-induced insulin secretion." In: *PloS one* 12.11, e0187213.

Muraro, Mauro J et al. (Oct. 2016). "A Single-Cell Transcriptome Atlas of the Human Pancreas." In: *Cell systems* 3.4, 385–394.e3.

Murtaugh, L Charles (Feb. 2007). "Pancreas and beta-cell development: from the actual to the possible." In: *Development (Cambridge, England)* 134.3, pp. 427–438.

Murtaugh, L Charles et al. (Dec. 2003). "Notch signaling controls multiple steps of pancreatic differentiation." In: *Proceedings of the National Academy of Sciences* 100.25, pp. 14920–14925.

Mziaut, Hassan et al. (Jan. 2008). "ICA512 signaling enhances pancreatic beta-cell proliferation by regulating cyclins D through STATs." In: *Proceedings of the National Academy of Sciences of the United States of America* 105.2, pp. 674–679.

Nair, Gopika et al. (June 2015). "Islet formation in mice and men: lessons for the generation of functional insulin-producing β-cells from human pluripotent stem cells." In: *Current opinion in genetics & development* 32, pp. 171–180.

Naruse, Y et al. (Nov. 1999). "Neural restrictive silencer factor recruits mSin3 and histone deacetylase complex to repress neuron-specific target genes." In: *Proceedings of the National Academy of Sciences* 96.24, pp. 13691–13696.

Naya, F J et al. (Sept. 1997). "Diabetes, defective pancreatic morphogenesis, and abnormal enteroendocrine differentiation in BETA2/neuroD-deficient mice." In: *Genes & development* 11.18, pp. 2323–2334.

Newman, B et al. (Oct. 1987). "Concordance for type 2 (non-insulin-dependent) diabetes mellitus in male twins." In: *Diabetologia* 30.10, pp. 763–768.

Ng, Natasha Hui Jin et al. (June 2019). "HNF4A Haploinsufficiency in MODY1 Abrogates Liver and Pancreas Differentiation from Patient-Derived Induced Pluripotent Stem Cells." In: *iScience* 16, pp. 192–205.

Nica, Alexandra C et al. (Sept. 2013). "Cell-type, allelic, and genetic signatures in the human pancreatic beta cell transcriptome." In: *Genome research* 23.9, pp. 1554–1562.

Niland, Joyce C et al. (2010). "Effectiveness of a web-based automated cell distribution system." In: *Cell transplantation* 19.9, pp. 1133–1142.

Nishigori, H et al. (Aug. 1998). "Frameshift mutation, A263fsinsGG, in the hepatocyte nuclear factor-1beta gene associated with diabetes and renal dysfunction." In: *Diabetes* 47.8, pp. 1354–1355.

Nostro, M Cristina et al. (Mar. 2011). "Stage-specific signaling through TGFβ family members and WNT regulates patterning and pancreatic specification of human pluripotent stem cells." In: *Development (Cambridge, England)* 138.5, pp. 861–871.

Nostro, M Cristina et al. (Apr. 2015). "Efficient generation of NKX6-1+ pancreatic progenitors from multiple human pluripotent stem cell lines." In: *Stem cell reports* 4.4, pp. 591–604.

Orho-Melander, Marju et al. (Nov. 2008). "Common missense variant in the glucokinase regulatory protein gene is associated with increased plasma triglyceride and C-reactive protein but lower fasting glucose concentrations." In: *Diabetes* 57.11, pp. 3112–3121.

Oslowski, Christine M et al. (Apr. 2010). "The binary switch between life and death of endoplasmic reticulum-stressed beta cells." In: *Current opinion in endocrinology, diabetes, and obesity* 17.2, pp. 107–112.

Owerbach, D et al. (July 1980). "The insulin gene is located on chromosome 11 in humans." In: *Nature* 286.5768, pp. 82–84.

Pachkov, Mikhail et al. (Jan. 2013). "SwissRegulon, a database of genome-wide annotations of regulatory sites: recent updates." In: *Nucleic acids research* 41.Database issue, pp. D214–20.

Pagliuca, Felicia W et al. (Oct. 2014). "Generation of functional human pancreatic β cells in vitro." In: *Cell* 159.2, pp. 428–439.

Pal, Aparna et al. (Feb. 2016). "Loss-of-Function Mutations in the Cell-Cycle Control Gene CDKN2A Impact on Glucose Homeostasis in Humans." In: *Diabetes* 65.2, pp. 527–533.

Pan, Dongning et al. (Dec. 2015). "Jmjd3-Mediated H3K27me3 Dynamics Orchestrate Brown Fat Development and Regulate White Fat Plasticity". In: *Developmental cell* 35.5, pp. 568–583.

Pan, Fong Cheng et al. (Mar. 2011). "Pancreas organogenesis: from bud to plexus to gland." In: *Developmental dynamics : an official publication of the American Association of Anatomists* 240.3, pp. 530–565.

Pan, Guangjin et al. (Aug. 2006). "A negative feedback loop of transcription factors that controls stem cell pluripotency and self-renewal." In: *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 20.10, pp. 1730–1732.

Parker, Stephen C J et al. (Oct. 2013). "Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants." In: *Proceedings of the National Academy of Sciences of the United States of America* 110.44, pp. 17921–17926.

Pasquali, Lorenzo et al. (Feb. 2014). "Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants." In: *Nature Genetics* 46.2, pp. 136–143.

Pasquier, Adrien et al. (July 2019). "Lysosomal degradation of newly formed insulin granules contributes to β cell failure in diabetes." In: *Nature communications* 10.1, pp. 3312–14.

Pearl, Esther J et al. (Mar. 2011). "Functional analysis of Rfx6 and mutant variants associated with neonatal diabetes." In: *Developmental biology* 351.1, pp. 135–145.

Perez-Alcantara, Marta et al. (Apr. 2018). "Patterns of differential gene expression in a cellular model of human islet development, and relationship to type 2 diabetes predisposition." In: *Diabetologia* 536, pp. 41–9.

Petersen, Maja Borup Kjær et al. (Oct. 2017). "Single-Cell Gene Expression Analysis of a Human ESC Model of Pancreatic Endocrine Development Reveals Different Paths to β-Cell Differentiation." In: *Stem cell reports* 9.4, pp. 1246–1261.

Piccand, Julie et al. (Dec. 2014). "Rfx6 maintains the functional identity of adult pancreatic β cells." In: *Cell reports* 9.6, pp. 2219–2232.

Piper, K et al. (Apr. 2004). "Beta cell differentiation during early human pancreas development." In: *The Journal of endocrinology* 181.1, pp. 11–23.

Pontoglio, M et al. (May 1998). "Defective insulin secretion in hepatocyte nuclear factor 1alpha-deficient mice." In: *The Journal of clinical investigation* 101.10, pp. 2215–2222.

Pound, Lynley D et al. (July 2009). "Deletion of the mouse Slc30a8 gene encoding zinc transporter-8 results in impaired insulin secretion." In: *The Biochemical journal* 421.3, pp. 371–376.

Pound, Lynley D et al. (2012). "The physiological effects of deleting the mouse SLC30A8 gene encoding zinc transporter-8 are influenced by gender and genetic background." In: *PloS one* 7.7, e40972.

Pradeepa, Madapura M (Jan. 2017). "Causal role of histone acetylations in enhancer function." In: *Transcription* 8.1, pp. 40–47.

Pullen, Timothy J et al. (Mar. 2010). "Identification of genes selectively disallowed in the pancreatic islet." In: *Islets* 2.2, pp. 89–95.

Pullen, Timothy J et al. (July 2012). "Overexpression of monocarboxylate transporter-1 (SLC16A1) in mouse pancreatic β-cells leads to relative hyperinsulinism during exercise." In: *Diabetes* 61.7, pp. 1719–1725.

Pullen, Timothy J et al. (June 2013). "When less is more: the forbidden fruits of gene repression in the adult β-cell." In: *Diabetes, Obesity and Metabolism* 15.6, pp. 503–512.

Pullen, Timothy J et al. (2017). "Analysis of Purified Pancreatic Islet Beta and Alpha Cell Transcriptomes Reveals 11β-Hydroxysteroid Dehydrogenase (Hsd11b1) as a Novel Disallowed Gene." In: *Frontiers in Genetics* 8.620, p. 41.

Qu, Xue-Bin et al. (Sept. 2008). "Sox17 facilitates the differentiation of mouse embryonic stem cells into primitive and definitive endoderm in vitro." In: *Development, growth & differentiation* 50.7, pp. 585–593.

Quesada, Iván et al. (Oct. 2008). "Physiology of the pancreatic alpha-cell and glucagon secretion: role in glucose homeostasis and diabetes." In: *The Journal of endocrinology* 199.1, pp. 5–19.

Quigley, Ian K et al. (Jan. 2017). "Rfx2 Stabilizes Foxj1 Binding at Chromatin Loops to Enable Multiciliated Cell Gene Expression." In: *PLoS genetics* 13.1, e1006538.

Quintens, Roel et al. (June 2008). "Why expression of some genes is disallowed in beta-cells." In: *Biochemical Society transactions* 36.Pt 3, pp. 300–305.

Rajagopal, Jayaraj et al. (Jan. 2003). "Insulin staining of ES cell progeny from insulin uptake." In: *Science* 299.5605, pp. 363–1032.

Ramond, Cyrille et al. (July 2017). "Reconstructing human pancreatic differentiation by mapping specific cell populations during development." In: *eLife* 6, p. 36.

Ramond, Cyrille et al. (July 2018). "Understanding human fetal pancreas development using subpopulation sorting, RNA sequencing and single-cell profiling." In: *Development (Cambridge, England)*, dev.165480.

Ran, F Ann et al. (Nov. 2013). "Genome engineering using the CRISPR-Cas9 system." In: *Nature protocols* 8.11, pp. 2281–2308.

Rankin, Scott A et al. (Mar. 2011). "A gene regulatory network controlling hhex transcription in the anterior endoderm of the organizer." In: *Developmental biology* 351.2, pp. 297–310.

Raphael, Benjamin et al. (Aug. 2017). "Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma." In: *Cancer cell* 32.2, 185–203.e13.

Raudvere, Uku et al. (July 2019). "g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update)." In: *Nucleic acids research* 47.W1, W191–W198.

Ravassard, Philippe et al. (Sept. 2011). "A genetically engineered human pancreatic β cell line exhibiting glucose-inducible insulin secretion." In: *The Journal of clinical investigation* 121.9, pp. 3589–3597.

Ray, Subir K et al. (Jan. 2003). "Novel transcriptional potentiation of BETA2/NeuroD on the secretin gene promoter by the DNA-binding protein Finb/RREB-1." In: *Molecular and Cellular Biology* 23.1, pp. 259–271.

Ray, Subir K et al. (June 2014). "CtBP and associated LSD1 are required for transcriptional activation by NeuroD1 in gastrointestinal endocrine cells." In: *Molecular and Cellular Biology* 34.12, pp. 2308–2317.

Reith, W et al. (Feb. 1994). "RFX1, a transactivator of hepatitis B virus enhancer I, belongs to a novel family of homodimeric and heterodimeric DNA-binding proteins." In: *Molecular and Cellular Biology* 14.2, pp. 1230–1244.

Reyes, Alejandro et al. (2016). *Inferring differential exon usage in RNA-Seq data with the DEXSeq package.*

Rezania, Alireza et al. (Aug. 2012). "Maturation of human embryonic stem cell-derived pancreatic progenitors into functional islets capable of treating pre-existing diabetes in mice." In: *Diabetes* 61.8, pp. 2016–2029.

Rezania, Alireza et al. (Nov. 2013). "Enrichment of human embryonic stem cell-derived NKX6.1-expressing pancreatic progenitor cells accelerates the maturation of insulin-secreting cells in vivo." In: *Stem cells (Dayton, Ohio)* 31.11, pp. 2432–2442.

Rezania, Alireza et al. (Nov. 2014). "Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells." In: *Nature biotechnology* 32.11, pp. 1121–1133.

Riedel, M J et al. (Feb. 2012). "Immunohistochemical characterisation of cells co-producing insulin and glucagon in the developing human pancreas." In: *Diabetologia* 55.2, pp. 372–381.

Risso, Davide (2019). "RUVSeq: Remove Unwanted Variation from RNA-Seq Data". In: *academia.edu*.

Risso, Davide et al. (Sept. 2014). "Normalization of RNA-seq data using factor analysis of control genes or samples." In: *Nature biotechnology* 32.9, pp. 896–902.

Rivlin, Noa et al. (Apr. 2011). "Mutations in the p53 Tumor Suppressor Gene: Important Milestones at the Various Steps of Tumorigenesis." In: *Genes & cancer* 2.4, pp. 466–474.

Röder, Pia V et al. (Mar. 2016). "Pancreatic regulation of glucose homeostasis." In: *Experimental & molecular medicine* 48.3, e219–e219.

Roman, Tamara S et al. (Sept. 2017). "A Type 2 Diabetes-Associated Functional Regulatory Variant in a Pancreatic Islet Enhancer at the ADCY5 Locus." In: *Diabetes* 66.9, pp. 2521–2530.

Ron, David et al. (July 2007). "Signal integration in the endoplasmic reticulum unfolded protein response." In: *Nature reviews. Molecular cell biology* 8.7, pp. 519–529.

Roopra, Avtar et al. (June 2004). "Localized domains of G9a-mediated histone methylation are required for silencing of neuronal genes." In: *Molecular cell* 14.6, pp. 727–738.

Roopra, A et al. (Mar. 2000). "Transcriptional repression by neuron-restrictive silencer factor is mediated via the Sin3-histone deacetylase complex." In: *Molecular and Cellular Biology* 20.6, pp. 2147–2157.

Rouhani, Foad et al. (June 2014). "Genetic background drives transcriptional variation in human induced pluripotent stem cells." In: *PLoS genetics* 10.6, e1004432.

Rual, Jean-François et al. (Oct. 2005). "Towards a proteome-scale map of the human protein|[ndash]|protein interaction network". In: *Nature* 437.7062, pp. 1173–1178.

Rubio-Cabezas, Oscar et al. (Sept. 2010). "Homozygous mutations in NEUROD1 are responsible for a novel syndrome of permanent neonatal diabetes and neurological abnormalities." In: *Diabetes* 59.9, pp. 2326–2331.

Rubio-Cabezas, Oscar et al. (Apr. 2011). "Permanent Neonatal Diabetes and Enteric Anendocrinosis Associated With Biallelic Mutations in NEUROG3." In: *Diabetes* 60.4, pp. 1349–1353.

Rubio-Cabezas, Oscar et al. (Sept. 2014). *ISPAD Clinical Practice Consensus Guidelines 2014. The diagnosis and management of monogenic diabetes in children and adolescents.* John Wiley & Sons, Ltd (10.1111).

Rukstalis, J Michael et al. (Nov. 2009). "Neurogenin3: a master regulator of pancreatic islet differentiation and regeneration." In: *Islets* 1.3, pp. 177–184.

Russ, Holger A et al. (July 2015). "Controlled induction of human pancreatic progenitors produces functional beta-like cells in vitro". In: *The EMBO journal* 34.13, pp. 1759–1772.

Rutter, Guy A (Nov. 2004). "Visualising insulin secretion. The Minkowski Lecture 2004." In: *Diabetologia*. Henry Wellcome Laboratories for Integrated Cell Signalling, School of Medical Sciences, University of Bristol, Bristol, UK. g.a.rutter@bris.ac.uk. Springer-Verlag, pp. 1861–1872.

Rutter, Guy A et al. (Sept. 2017). "Local and regional control of calcium dynamics in the pancreatic islet." In: *Diabetes, Obesity and Metabolism* 19 Suppl 1.9, pp. 30–41.

Ryffel, G U (Aug. 2001). "Mutations in the human genes encoding the transcription factors of the hepatocyte nuclear factor (HNF)1 and HNF4 families: functional and pathological consequences." In: *Journal of molecular endocrinology* 27.1, pp. 11–29.

Saad, Reda S et al. (Nov. 2011). "CDX2 as a marker for intestinal differentiation: Its utility and limitations." In: *World journal of gastrointestinal surgery* 3.11, pp. 159–166.

Saarimäki-Vire, Jonna et al. (Apr. 2017). "An Activating STAT3 Mutation Causes Neonatal Diabetes through Premature Induction of Pancreatic Differentiation." In: *Cell reports* 19.2, pp. 281–294.

Salanti, Georgia et al. (Sept. 2009). "Underlying genetic models of inheritance in established type 2 diabetes associations." In: *American journal of epidemiology* 170.5, pp. 537–545.

Salisbury, Rachel J et al. (2014). "The window period of NEUROGENIN3 during human gestation." In: *Islets* 6.3, e954436.

Sander, M et al. (July 1997). "Genetic analysis reveals that PAX6 is required for normal transcription of pancreatic hormone genes and islet development." In: *Genes & development* 11.13, pp. 1662–1673.

Sander, M et al. (Dec. 2000). "Homeobox gene Nkx6.1 lies downstream of Nkx2.2 in the major pathway of beta-cell formation in the pancreas." In: *Development (Cambridge, England)* 127.24, pp. 5533–5540.

Sanjana, Neville E et al. (Aug. 2014). "Improved vectors and genome-wide libraries for CRISPR screening." In: *Nature methods* 11.8, pp. 783–784.

Sansbury, Francis H et al. (Dec. 2015). "Biallelic RFX6 mutations can cause childhood as well as neonatal onset diabetes mellitus." In: *European journal of human genetics : EJHG* 23.12, pp. 1744–1748.

Satin, L S et al. (June 1998). "Neurotransmitters and their receptors in the islets of Langerhans of the pancreas: what messages do acetylcholine, glutamate, and GABA transmit?" In: *Endocrine* 8.3, pp. 213–223.

Schaffer, Ashleigh E et al. (2013). "Nkx6.1 controls a gene regulatory network required for establishing and maintaining pancreatic Beta cell identity." In: *PLoS genetics* 9.1, e1003274.

Scharfmann, Raphael et al. (Sept. 2019). "The supply chain of human pancreatic β cell lines." In: *The Journal of clinical investigation* 129.9, pp. 3511–3520.

Scheuner, Donalyn et al. (July 2005). "Control of mRNA translation preserves endoplasmic reticulum function in beta cells and maintains glucose homeostasis." In: *Nature medicine* 11.7, pp. 757–764.

Schoenherr, C J et al. (Mar. 1995). "The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes." In: *Science* 267.5202, pp. 1360–1363.

Schwitzgebel, V M (Mar. 2014). "Many faces of monogenic diabetes." In: *Journal of diabetes investigation* 5.2, pp. 121–133.

Schwitzgebel, V M et al. (Aug. 2000). "Expression of neurogenin3 reveals an islet cell precursor population in the pancreas." In: *Development (Cambridge, England)* 127.16, pp. 3533–3542.

Scott, Robert A et al. (Sept. 2012). "Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways". In: *Nature Genetics* 44.9, pp. 991–1005.

Scott, Robert A et al. (May 2017). "An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans." In: *Diabetes*, db161253.

Segerstolpe, Åsa et al. (Sept. 2016). "Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes." In: *Cell metabolism*.

Séguin, Cheryle A et al. (Aug. 2008). "Establishment of endoderm progenitors by SOX transcription factor expression in human embryonic stem cells." In: *Cell stem cell* 3.2, pp. 182–195.

Sekine, N et al. (Feb. 1994). "Low lactate dehydrogenase and high mitochondrial glycerol phosphate dehydrogenase in pancreatic beta-cells. Potential role in nutrient sensing." In: *Journal of Biological Chemistry* 269.7, pp. 4895–4902.

Selbach, Matthias et al. (Sept. 2008). "Widespread changes in protein synthesis induced by microRNAs." In: *Nature* 455.7209, pp. 58–63.

Sellick, Gabrielle S et al. (Dec. 2004). "Mutations in PTF1A cause pancreatic and cerebellar agenesis." In: *Nature Genetics* 36.12, pp. 1301–1305.

Senée, Valérie et al. (June 2006). "Mutations in GLIS3 are responsible for a rare syndrome with neonatal diabetes mellitus and congenital hypothyroidism." In: *Nature Genetics* 38.6, pp. 682–687.

Sengupta, Pritam K et al. (July 2002). "The RFX family interacts at the collagen (COL1A2) start site and represses transcription." In: *Journal of Biological Chemistry* 277.28, pp. 24926–24937.

Serre-Beinier, Véronique et al. (Feb. 2009). "Cx36 makes channels coupling human pancreatic beta-cells, and correlates with insulin expression." In: *Human Molecular Genetics* 18.3, pp. 428–439.

Seymour, Philip A et al. (Feb. 2007). "SOX9 is required for maintenance of the pancreatic progenitor cell pool." In: *Proceedings of the National Academy of Sciences* 104.6, pp. 1865–1870.

Shaffer, Lisa G et al. (May 2004). "A cytogeneticist's perspective on genomic microarrays." In: *Human reproduction update* 10.3, pp. 221–226.

Shalem, Ophir et al. (Jan. 2014). "Genome-scale CRISPR-Cas9 knockout screening in human cells." In: *Science* 343.6166, pp. 84–87.

Shi, Yujiang et al. (Apr. 2003). "Coordinated histone modifications mediated by a CtBP co-repressor complex". In: *Nature* 422.6933, pp. 735–738.

Shi, Yujiang et al. (Dec. 2004). "Histone demethylation mediated by the nuclear amine oxidase homolog LSD1." In: *Cell* 119.7, pp. 941–953.

Shi, Zhong-Dong et al. (May 2017). "Genome Editing in hPSCs Reveals GATA6 Haploinsufficiency and a Genetic Interaction with GATA4 in Human Pancreatic Development." In: *Cell stem cell* 20.5, 675–688.e6.

Shiao, Meng-Shin et al. (Mar. 2008). "Adaptive evolution of the insulin two-gene system in mouse." In: *Genetics* 178.3, pp. 1683–1691.

Singh, Rajendra Kumar et al. (May 2011). "The IL-8-regulated chemokine receptor CXCR7 stimulates EGFR signaling to promote prostate cancer growth." In: *Cancer research* 71.9, pp. 3268–3277.

Sladek, Robert et al. (Feb. 2007). "A genome-wide association study identifies novel risk loci for type 2 diabetes." In: *Nature* 445.7130, pp. 881–885.

Smith, S B et al. (Dec. 1999). "Paired-homeodomain transcription factor PAX4 acts as a transcriptional repressor in early pancreatic development." In: *Molecular and Cellular Biology* 19.12, pp. 8272–8280.

Smith, S B et al. (Oct. 2003). "Neurogenin3 and hepatic nuclear factor 1 cooperate in activating pancreatic expression of Pax4." In: *Journal of Biological Chemistry* 278.40, pp. 38254–38259.

Smith, S B et al. (Feb. 2010). "Rfx6 directs islet formation and insulin production in mice and humans." In: *Nature* 463.7282, pp. 775–780.

Solomon, Benjamin D et al. (Nov. 2009). "Compound heterozygosity for mutations in PAX6 in a patient with complex brain anomaly, neonatal diabetes mellitus, and microophthalmia." In: *American journal of medical genetics. Part A* 149A.11, pp. 2543–2546.

Son, Mi-Young et al. (Nov. 2013). "Unveiling the critical role of REX1 in the regulation of human stem cell pluripotency." In: *Stem cells (Dayton, Ohio)* 31.11, pp. 2374–2387.

Sosa-Pineda, B et al. (Mar. 1997). "The Pax4 gene is essential for differentiation of insulin-producing beta cells in the mammalian pancreas." In: *Nature* 386.6623, pp. 399–402.

Soyer, Josselin et al. (Jan. 2010). "Rfx6 is an Ngn3-dependent winged helix transcription factor required for pancreatic islet cell development." In: *Development (Cambridge, England)* 137.2, pp. 203–212.

Spiegel, Ronen et al. (Nov. 2011). "Clinical characterization of a newly described neonatal diabetes syndrome caused by RFX6 mutations." In: *American journal of medical genetics. Part A* 155A.11, pp. 2821–2825.

Spits, Claudia et al. (Dec. 2008). "Recurrent chromosomal abnormalities in human embryonic stem cells." In: *Nature biotechnology* 26.12, pp. 1361–1363.

Spracklen, Cassandra N et al. (June 2019). "Identification of type 2 diabetes loci in 433,540 East Asian individuals". In: *bioRxiv*, p. 685172.

Starr, Timothy K et al. (Mar. 2009). "A transposon-based genetic screen in mice identifies genes altered in colorectal cancer." In: *Science* 323.5922, pp. 1747–1750.

Steiner, Donald J et al. (May 2010). "Pancreatic islet plasticity: interspecies comparison of islet architecture and composition." In: *Islets* 2.3, pp. 135–145.

Steinthorsdottir, Valgerdur et al. (Mar. 2014). "Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes." In: *Nature Genetics* 46.3, pp. 294–298.

Stepniewski, J et al. (Feb. 2015). "Induced pluripotent stem cells as a model for diabetes investigation." In: *Scientific reports* 5, p. 8597.

Stevens, Tim J et al. (Mar. 2017). "3D structures of individual mammalian genomes studied by single-cell Hi-C." In: *Nature* 589, p. 2931.

Stewart, Andrew F et al. (June 2015). "Human β-Cell Proliferation and Intracellular Signaling: Part 3". In: *Diabetes* 64.6, pp. 1872–1885.

Stitzel, Michael L et al. (Nov. 2010). "Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci." In: *Cell metabolism* 12.5, pp. 443–455.

Stoffers, D A et al. (Jan. 1997). "Pancreatic agenesis attributable to a single nucleotide deletion in the human IPF1 gene coding sequence." In: *Nature Genetics* 15.1, pp. 106–110.

Sugiaman-Trapman, Debora et al. (Mar. 2018). "Characterization of the human RFX transcription factor family by regulatory and target gene analysis." In: *BMC genomics* 19.1, p. 181.

Sui, Lina et al. (Sept. 2012). "FGF signaling via MAPK is required early and improves Activin A-induced definitive endoderm formation from human embryonic stem cells." In: *Biochemical and biophysical research communications* 426.3, pp. 380–385.

Sussel, L et al. (June 1998). "Mice lacking the homeodomain transcription factor Nkx2.2 have diabetes due to arrested differentiation of pancreatic beta cells." In: *Development (Cambridge, England)* 125.12, pp. 2213–2221.

Svensson, Per et al. (Apr. 2009). "MFng is dispensable for mouse pancreas development and function." In: *Molecular and Cellular Biology* 29.8, pp. 2129–2138.

Swisa, Avital et al. (Jan. 2017). "PAX6 maintains β cell identity by repressing genes of alternative islet cell types." In: *The Journal of clinical investigation* 127.1, pp. 230–243.

Szabat, Marta et al. (June 2012). "Maintenance of β-cell maturity and plasticity in the adult pancreas: developmental biology concepts in adult physiology." In: *Diabetes* 61.6, pp. 1365–1371.

Szabat, Marta et al. (Jan. 2016). "Reduced Insulin Production Relieves Endoplasmic Reticulum Stress and Induces β Cell Proliferation." In: *Cell metabolism* 23.1, pp. 179–193.

Tabuchi, H et al. (July 2000). "Regulation of insulin secretion by overexpression of Ca2+/calmodulin-dependent protein kinase II in insulinoma MIN6 cells." In: *Endocrinology* 141.7, pp. 2350–2360.

Takahashi, Kazutoshi et al. (Aug. 2006). "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors." In: *Cell* 126.4, pp. 663–676.

Takahashi, Kazutoshi et al. (Nov. 2007). "Induction of pluripotent stem cells from adult human fibroblasts by defined factors." In: *Cell* 131.5, pp. 861–872.

Taleb, Nadine et al. (Sept. 2011). "RFX6 is needed for the development and maintenance of the β-cell phenotype". In: *Islets* 3.5, pp. 291–293.

Tang, D et al. (Nov. 1995). "An isoform of the neuronal cyclin-dependent kinase 5 (Cdk5) activator." In: *Journal of Biological Chemistry* 270.45, pp. 26897–26903.

Tarasov, Andrei I et al. (Dec. 2004). "Metabolic regulation of the pancreatic beta-cell ATP-sensitive K+ channel: a pas de deux." In: *Diabetes* 53 Suppl 3.Supplement 3, S113–22.

Tarasov, Andrei I et al. (2012). "The mitochondrial Ca2+ uniporter MCU is essential for glucose-induced ATP increases in pancreatic β-cells." In: *PloS one* 7.7, e39722.

Taylor, Brandon L et al. (Sept. 2013). "Nkx6.1 is essential for maintaining the functional state of pancreatic beta cells." In: *Cell reports* 4.6, pp. 1262–1275.

Teo, Adrian K K et al. (Feb. 2013). "Derivation of human induced pluripotent stem cells from patients with maturity onset diabetes of the young." In: *Journal of Biological Chemistry* 288.8, pp. 5353–5356.

Thiagalingam, Arunthathi et al. (Oct. 1996). "RREB-1, a novel zinc finger protein, is involved in the differentiation response to Ras in human medullary thyroid carcinomas." In: *Molecular and Cellular Biology* 16.10, pp. 5335–5345.

Thiagalingam, Arunthathi et al. (Nov. 1997). "RREB1, a Ras Responsive Element Binding Protein, Maps to Human Chromosome 6p25". In: *Genomics* 45.3, pp. 630–632.

Thomsen, Soren K et al. (Aug. 2014). "The pancreatic β cell: recent insights from human genetics." In: *Trends in endocrinology and metabolism: TEM* 25.8, pp. 425–434.

Thomsen, Soren K et al. (Dec. 2016). "Systematic Functional Characterization of Candidate Causal Genes for Type 2 Diabetes Risk Variants." In: *Diabetes* 65.12, pp. 3805–3811.

Thomsen, Soren K et al. (July 2018). "Type 2 diabetes risk alleles in PAM impact insulin release from human pancreatic β-cells." In: *Nature Genetics* 63, p. 2158.

Thorrez, Lieven et al. (Jan. 2011). "Tissue-specific disallowance of housekeeping genes: the other face of cell differentiation." In: *Genome research* 21.1, pp. 95–105.

Thurman, Robert E et al. (Sept. 2012). "The accessible chromatin landscape of the human genome." In: *Nature* 489.7414, pp. 75–82.

Thurmond, Debbie C (2007). "Regulation of Insulin Action and Insulin Secretion by SNARE-Mediated Vesicle Exocytosis". In: *Mechanisms of Insulin Action*. New York, NY: Springer, New York, NY, pp. 52–70.

Thurner, Matthias et al. (Feb. 2018). "Integration of human pancreatic islet genomic data refines regulatory mechanisms at Type 2 Diabetes susceptibility loci." In: *eLife* 7, p. 1363.

Tiyaboonchai, Amita et al. (Feb. 2017). "GATA6 Plays an Important Role in the Induction of Human Definitive Endoderm, Development of the Pancreas, and Functionality of Pancreatic β Cells." In: *Stem cell reports* 0.0.

Trask, Barbara J (Oct. 2002). "Human cytogenetics: 46 chromosomes, 46 years and counting." In: *Nature reviews. Genetics* 3.10, pp. 769–778.

Travers, Mary E et al. (Mar. 2013). "Insights into the molecular mechanism for type 2 diabetes susceptibility at the KCNQ1 locus from temporal changes in imprinting status in human islets." In: *Diabetes* 62.3, pp. 987–992.

Trynka, Gosia et al. (Feb. 2013). "Chromatin marks identify critical cell types for fine mapping complex trait variants." In: *Nature Genetics* 45.2, pp. 124–130.

Tsai, Shengdar Q et al. (Feb. 2015). "GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases." In: *Nature biotechnology* 33.2, pp. 187–197.

Tsai, Shengdar Q et al. (June 2017). "CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets." In: *Nature methods* 14.6, pp. 607–614.

Tsonkova, Violeta Georgieva et al. (Dec. 2017). "The EndoC-βH1 cell line is a valid model of human beta cells and applicable for screenings to identify novel drug target candidates." In: *Molecular metabolism*.

Tuomi, Tiinamaija et al. (June 2016). "Increased Melatonin Signaling Is a Risk Factor for Type 2 Diabetes." In: *Cell metabolism* 23.6, pp. 1067–1077.

Uren, Anthony G et al. (May 2008). "Large-scale mutagenesis in p19(ARF)- and p53-deficient mice identifies cancer genes and their collaborative networks." In: *Cell* 133.4, pp. 727–741.

Van de Bunt, Martijn et al. (Dec. 2015). "Transcript Expression Data from Human Islets Links Regulatory Signals from Genome-Wide Association Studies for Type 2 Diabetes and Glycemic Traits to Their Downstream Effectors." In: *PLoS genetics* 11.12, e1005694.

Van de Bunt, Martijn et al. (Apr. 2016). "Insights into islet development and biology through characterization of a human iPSC-derived endocrine pancreas model." In: *Islets* 8.3, pp. 83–95.

Van Arensbergen, Joris et al. (June 2010). "Derepression of Polycomb targets during pancreatic organogenesis allows insulin-producing beta-cells to adopt a neural gene activity program." In: *Genome research* 20.6, pp. 722–732.

Varshney, Arushi et al. (Feb. 2017). "Genetic regulatory signatures underlying islet gene expression and type 2 diabetes." In: *Proceedings of the National Academy of Sciences of the United States of America* 114.9, pp. 2301–2306.

Varshney, Arushi et al. (July 2018). "Genetic Effects on Enhancer Activity in Human Pancreatic Islets". In: *Diabetes* 67.Supplement 1, 1706–P.

Vaxillaire, Martine et al. (Apr. 2012). "The lessons of early-onset monogenic diabetes for the understanding of diabetes pathogenesis." In: *Best practice & research. Clinical endocrinology & metabolism* 26.2, pp. 171–187.

Veres, Adrian et al. (May 2019). "Charting cellular identity during human in vitro β-cell differentiation." In: *Nature* 569.7756, pp. 368–373.

Villasenor, Alethia et al. (Nov. 2008). "Biphasic Ngn3 expression in the developing pancreas." In: *Developmental dynamics : an official publication of the American Association of Anatomists* 237.11, pp. 3270–3279.

Visel, Axel et al. (Feb. 2009). "ChIP-seq accurately predicts tissue-specific activity of enhancers." In: *Nature* 457.7231, pp. 854–858.

Voight, Benjamin F et al. (July 2010). "Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis." In: *Nature Genetics* 42.7, pp. 579–589.

Volkov, Petr et al. (Jan. 2017). "Whole-genome Bisulfite Sequencing of Human Pancreatic Islets Reveals Novel Differentially Methylated Regions in Type 2 Diabetes Pathogenesis." In: *Diabetes* 66.4, pp. 1074–1085.

Wang, Allen et al. (Apr. 2015). "Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates." In: *Cell stem cell* 16.4, pp. 386–399.

Wang, Jiafang et al. (July 2006). "Mutant neurogenin-3 in congenital malabsorptive diarrhea." In: *New England Journal of Medicine* 355.3, pp. 270–280.

Wang, Junfeng et al. (Feb. 2004). "The concerted activities of Pax4 and Nkx2.2 are essential to initiate pancreatic beta-cell differentiation." In: *Developmental biology* 266.1, pp. 178–189.

Wang, Sui et al. (Nov. 2007). "Loss of Myt1 function partially compromises endocrine islet cell differentiation and pancreatic physiological function in the mouse." In: *Mechanisms of development* 124.11-12, pp. 898–910.

Wang, Sui et al. (May 2008). "Myt1 and Ngn3 form a feed-forward expression loop to promote endocrine islet cell differentiation." In: *Developmental biology* 317.2, pp. 531–540.

Wang, Weiping et al. (June 2016). "PDX1 and ISL1 differentially coordinate with epigenetic modifications to regulate insulin gene expression in varied glucose concentrations." In: *Molecular and Cellular Endocrinology* 428, pp. 38–48.

Wang, Xianming et al. (Mar. 2018). "Genome-wide analysis of PDX1 target genes in human pancreatic progenitors." In: *Molecular metabolism* 9, pp. 57–68.

Wang, Xianming et al. (Mar. 2019). "Point mutations in the PDX1 transactivation domain impair human β-cell development and function". In: *Molecular metabolism*.

Wapner, Ronald J et al. (Dec. 2012). "Chromosomal microarray versus karyotyping for prenatal diagnosis." In: *New England Journal of Medicine* 367.23, pp. 2175–2184.

Watada, H et al. (Nov. 2000). "Transcriptional and translational regulation of beta-cell differentiation factor Nkx6.1." In: *Journal of Biological Chemistry* 275.44, pp. 34224–34230.

Watanabe, Naoki et al. (June 2009). "A murine model of neonatal diabetes mellitus in Glis3-deficient mice." In: *FEBS letters* 583.12, pp. 2108–2113.

Weedon, Michael N et al. (Jan. 2014). "Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis." In: *Nature Genetics* 46.1, pp. 61–64.

Weir, G C et al. (1985). "Pancreatic somatostatin." In: *Advances in experimental medicine and biology* 188.1, pp. 403–423.

Wijesekara, N et al. (Aug. 2010). "Beta cell-specific Znt8 deletion in mice causes marked defects in insulin processing, crystallisation and secretion." In: *Diabetologia* 53.8, pp. 1656–1668.

Willemsen, Gonneke et al. (Dec. 2015). "The Concordance and Heritability of Type 2 Diabetes in 34,166 Twin Pairs From International Twin Registers: The Discordant Twin (DISCOTWIN) Consortium." In: *Twin research and human genetics : the official journal of the International Society for Twin Studies* 18.6, pp. 762–771.

Wilson, Maria E et al. (Jan. 2003). "Gene expression cascades in pancreatic development." In: *Mechanisms of development* 120.1, pp. 65–80.

Wolfe, Steven A et al. (Oct. 2006). "Transcription factor RFX2 is abundant in rat testis and enriched in nuclei of primary spermatocytes where it appears to be required for transcription of the testis-specific histone H1t gene." In: *Journal of cellular biochemistry* 99.3, pp. 735–746.

Wood, Andrew R et al. (May 2017). "A Genome-Wide Association Study of IVGTT-Based Measures of First Phase Insulin Secretion Refines the Underlying Physiology of Type 2 Diabetes Variants." In: *Diabetes*, db161452.

Wu, Hao et al. (June 2008). "Copy number variant analysis of human embryonic stem cells." In: *Stem cells (Dayton, Ohio)* 26.6, pp. 1484–1489.

Wu, Yujian et al. (Feb. 2016). "Transcription Factor RFX2 Is a Key Regulator of Mouse Spermiogenesis." In: *Scientific reports* 6.1, p. 20435.

Wu, Zhen-Yong et al. (Aug. 2012). "AMPA receptors regulate exocytosis and insulin release in pancreatic β cells." In: *Traffic (Copenhagen, Denmark)* 13.8, pp. 1124–1139.

Xu, Xiaofang et al. (Sept. 2011). "Activin, BMP and FGF pathways cooperate to promote endoderm and pancreatic lineage cell differentiation from human embryonic stem cells." In: *Mechanisms of development* 128.7-10, pp. 412–427.

Xu, Yanwen et al. (Sept. 2006). "The fringe molecules induce endocrine differentiation in embryonic endoderm by activating cMyt1/cMyt3." In: *Developmental biology* 297.2, pp. 340–349.

Xuan, Shouhong et al. (Oct. 2012). "Pancreas-specific deletion of mouse Gata4 and Gata6 causes pancreatic agenesis." In: *The Journal of clinical investigation* 122.10, pp. 3516–3528.

Xuan, Shouhong et al. (Mar. 2016). "GATA4 and GATA6 regulate pancreatic endoderm identity through inhibition of hedgehog signaling." In: *Development (Cambridge, England)* 143.5, pp. 780–786.

Yamagata, K et al. (Dec. 1996a). "Mutations in the hepatocyte nuclear factor-1alpha gene in maturity-onset diabetes of the young (MODY3)". In: *Nature* 384.6608, pp. 455–458.

Yamagata, K et al. (Dec. 1996b). "Mutations in the hepatocyte nuclear factor-4alpha gene in maturity-onset diabetes of the young (MODY1)". In: *Nature* 384.6608, pp. 458–460.

Yamane, Takuya et al. (May 2013). "Transcriptional Activation of the Cholecystokinin Gene by DJ-1 through Interaction of DJ-1 with RREB1 and the Effect of DJ-1 on the Cholecystokinin Level in Mice". In: *PloS one* 8.11, e78374.

Yan, Winston X et al. (May 2017). "BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks." In: *Nature communications* 8.1, pp. 15058–9.

Yang, Yu-Ping et al. (Aug. 2011). "Context-specific α- to-β-cell reprogramming by forced Pdx1 expression." In: *Genes & development* 25.16, pp. 1680–1685.

Yasunaga, Masahiro et al. (Dec. 2005). "Induction and monitoring of definitive and visceral endoderm differentiation of mouse ES cells." In: *Nature biotechnology* 23.12, pp. 1542–1550.

Yeo, Michele et al. (Jan. 2005). "Small CTD phosphatases function in silencing neuronal gene expression." In: *Science* 307.5709, pp. 596–600.

Yokoi, Norihide et al. (Apr. 2016). "β-Cell glutamate signaling: Its role in incretin-induced insulin secretion." In: *Journal of diabetes investigation* 7 Suppl 1.Suppl. 2, pp. 38–43.

Yorifuji, Tohru et al. (Oct. 2012). "Dominantly inherited diabetes mellitus caused by GATA6 haploinsufficiency: variable intrafamilial presentation." In: *Journal of medical genetics* 49.10, pp. 642–643.

You, A et al. (Feb. 2001). "CoREST is an integral component of the CoREST- human histone deacetylase complex." In: *Proceedings of the National Academy of Sciences* 98.4, pp. 1454–1458.

Yu, Guangchuang et al. (July 2015). "ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization." In: *Bioinformatics (Oxford, England)* 31.14, pp. 2382–2383.

Zeggini, Eleftheria et al. (June 2007). "Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes." In: *Science* 316.5829, pp. 1336–1341.

Zeng, Hui et al. (Sept. 2016). "An Isogenic Human ESC Platform for Functional Evaluation of Genome-wide-Association-Study-Identified Diabetes Genes and Drug Discovery." In: *Cell stem cell* 19.3, pp. 326–340.

Zhang, Bin et al. (2005). *A General Framework for Weighted Gene Co-Expression Network Analysis*.

Zhang, Hui et al. (Sept. 2009). "The LIM-homeodomain protein ISL1 activates insulin gene promoter directly through synergy with BETA2." In: *Journal of molecular biology* 392.3, pp. 566–577.

Zhang, Jia et al. (Apr. 2014). "The diabetes gene Hhex maintains δ-cell differentiation and islet function." In: *Genes & development* 28.8, pp. 829–834.

Zhang, L et al. (July 1999). "A human Raf-responsive zinc-finger protein that binds to divergent sequences." In: *Nucleic acids research* 27.14, pp. 2947–2956.

Zhang, Shuling et al. (Apr. 2003). "p16 INK4a gene promoter variation and differential binding of a repressor, the ras-responsive zinc-finger transcription factor, RREB." In: *Oncogene* 22.15, pp. 2285–2295.

Zhang, Yong et al. (2008). "Model-based analysis of ChIP-Seq (MACS)." In: *Genome biology* 9.9, R137–9.

Zhao, Bin et al. (July 2008). "TEAD mediates YAP-dependent gene induction and growth control." In: *Genes & development* 22.14, pp. 1962–1971.

Zhao, Bin et al. (Aug. 2011). "The Hippo pathway in organ size control, tissue regeneration and stem cell self-renewal." In: *Nature cell biology* 13.8, pp. 877–883.

Zhou, Qiao et al. (July 2007). "A multipotent progenitor domain guides pancreatic organogenesis." In: *Developmental cell* 13.1, pp. 103–114.

Zhou, Vicky W et al. (Jan. 2011). "Charting histone modifications and the functional organization of mammalian genomes." In: *Nature reviews. Genetics* 12.1, pp. 7–18.

Zhu, Zengrong et al. (June 2016). "Genome Editing of Lineage Determinants in Human Pluripotent Stem Cells Reveals Mechanisms of Pancreatic Development and Diabetes." In: *Cell stem cell* 18.6, pp. 755–768.

Zou, Jing et al. (Oct. 2011). "hZIP1 zinc transporter down-regulation in prostate cancer involves the overexpression of ras responsive element binding protein-1 (RREB-1)." In: *The Prostate* 71.14, pp. 1518–1524.